House of Commons
Education Committee

2012 GCSE English results

First Report of Session 2013–14

Report, together with formal minutes, oral and written evidence

Ordered by the House of Commons
to be printed 5 June 2013
The Education Committee

The Education Committee is appointed by the House of Commons to examine the expenditure, administration and policy of the Department for Education and its associated public bodies.

Membership at time Report agreed:

Mr Graham Stuart MP (Conservative, Beverley & Holderness) (Chair)
Neil Carmichael MP (Conservative, Stroud)
Alex Cunningham MP (Labour, Stockton North)
Bill Esterson MP (Labour, Sefton Central)
Pat Glass MP (Labour, North West Durham)
Charlotte Leslie MP (Conservative, Bristol North West)
Siobhain McDonagh MP (Labour, Mitcham and Morden)
Ian Mearns MP (Labour, Gateshead)
Chris Skidmore MP (Conservative, Kingswood)
David Ward MP (Liberal Democrat, Bradford East)
Craig Whittaker MP (Conservative, Calder Valley)

Damian Hinds MP (Conservative, East Hampshire) was also a member of the Committee during the inquiry.

Powers

The Committee is one of the departmental select committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No 152. These are available on the Internet via www.parliament.uk

Publications

The Reports and evidence of the Committee are published by The Stationery Office by Order of the House. All publications of the Committee (including press notices) are on the Internet at www.parliament.uk/education-committee

Committee staff

The current staff of the Committee are Dr Lynn Gardner (Clerk), Geraldine Alexander (Second Clerk), Penny Crouzet (Committee Specialist), Emma Gordon (Committee Specialist), Jake Anders (Committee Specialist), Ameet Chudasama (Senior Committee Assistant), Caroline McElwee (Committee Assistant), and Paul Hampson (Committee Support Assistant)

Contacts

All correspondence should be addressed to the Clerk of the Education Committee, House of Commons, 7 Millbank, London SW1P 3JA. The telephone number for general enquiries is 020 7219 6181; the Committee’s e-mail address is educom@parliament.uk
# Contents

## Report

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>5</td>
</tr>
<tr>
<td>Our inquiry</td>
<td>5</td>
</tr>
<tr>
<td>What happened and when</td>
<td>6</td>
</tr>
<tr>
<td>The judicial review</td>
<td>8</td>
</tr>
<tr>
<td><strong>2 GCSE English and GCSE English Language qualifications</strong></td>
<td>10</td>
</tr>
<tr>
<td>GCSE English syllabuses in 2012</td>
<td>10</td>
</tr>
<tr>
<td>Key factors in the problems experienced last year</td>
<td>10</td>
</tr>
<tr>
<td>Modular/linear structure</td>
<td>11</td>
</tr>
<tr>
<td>Balance of internal and external assessment</td>
<td>12</td>
</tr>
<tr>
<td>Moderation procedures</td>
<td>14</td>
</tr>
<tr>
<td>Moderator feedback</td>
<td>14</td>
</tr>
<tr>
<td>Teacher marking and tolerances</td>
<td>15</td>
</tr>
<tr>
<td>Speaking and listening</td>
<td>18</td>
</tr>
<tr>
<td>Pressures from the school accountability system</td>
<td>19</td>
</tr>
<tr>
<td><strong>3 The role of Ofqual in qualifications design</strong></td>
<td>21</td>
</tr>
<tr>
<td>The impact of organisational change on qualifications development</td>
<td>22</td>
</tr>
<tr>
<td>Warning voices</td>
<td>23</td>
</tr>
<tr>
<td>Forthcoming GCSE and A level reforms</td>
<td>24</td>
</tr>
<tr>
<td><strong>4 Other lessons learned</strong></td>
<td>27</td>
</tr>
<tr>
<td>Exam board communication with schools and colleges</td>
<td>27</td>
</tr>
<tr>
<td>Grade awarding and the comparable outcomes approach</td>
<td>28</td>
</tr>
<tr>
<td><strong>5 The future of three-country qualifications and regulation</strong></td>
<td>32</td>
</tr>
<tr>
<td>Disagreements between regulators</td>
<td>32</td>
</tr>
<tr>
<td>Policy divergence between England, Wales and Northern Ireland</td>
<td>34</td>
</tr>
<tr>
<td>Implications for Ofqual</td>
<td>36</td>
</tr>
<tr>
<td><strong>6 The investigation process</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>7 Conclusion</strong></td>
<td>41</td>
</tr>
<tr>
<td>Conclusions and recommendations</td>
<td>43</td>
</tr>
<tr>
<td>Formal Minutes</td>
<td>50</td>
</tr>
<tr>
<td><strong>Appendix: Memorandum submitted by the Department for Education</strong></td>
<td>48</td>
</tr>
<tr>
<td>Witnesses</td>
<td>51</td>
</tr>
</tbody>
</table>
List of printed written evidence 51
List of Reports from the Committee during the current Parliament 52
Summary

GCSE English is a high stakes qualification for young people, their teachers and their schools and colleges. When things go wrong with such a qualification, the consequences are very serious for all involved, but particularly for the young people sitting the examination.

The 2012 GCSE English results prompted significant concern, which resulted ultimately in an application for judicial review. The legal challenge was unsuccessful, with the judge endorsing the actions taken by Ofqual and the exam boards. However, there are some very sobering lessons to be learned from the events with GCSE English in 2012. Confidence in the exam system has been shaken and a substantial amount of public money has been spent.

Following the problems experienced in 2012, several changes have been put in place or are proposed to strengthen current GCSE English qualifications until they are replaced in summer 2017. We welcome these changes and believe that they will help to make the current English GCSEs more robust and more resistant to pressures from the accountability system.

Several of the problems with GCSE English can be traced to the qualifications development phase. This underlines the vital importance of getting decisions right during qualifications design. Exam board experts raised concerns at the time, but these were not acted upon by the regulator. One of the crucial lessons to be learned from this episode is that Ofqual and Ministers should listen when concerns are raised during qualification development, especially when they come from specialists in the field.

The Government is embarking on the most significant and wide-ranging reforms to GCSEs and A levels since Ofqual was established as an independent regulator. Forthcoming reforms of GCSEs and A levels will put the new arrangements to the test. We recommend that the DfE and Ofqual set out in detail their respective roles and responsibilities in qualification design, particularly regarding how subject content will be developed, and publish this information before their consultations on GCSE reform have ended.

There is still much to be done to restore confidence in English GCSEs, particularly among teachers. Exam boards must continue to take steps to improve their communication with schools and colleges. Ofqual must also improve its communication with schools and colleges about the marking and grading of exams and its role in standards setting, in order to improve confidence in its work as regulator, particularly among teachers and head teachers. Ofqual should also keep the comparable outcomes approach under review.

Current GCSEs and A levels are jointly owned by the three regulators in England, Wales and Northern Ireland. These arrangements have been tested to the limit in recent months and it appears that the era of three-country qualifications and regulation is coming to an
end. We believe that such an outcome would be regrettable and urge Ministers to do everything possible to retain joint ownership of GCSEs and A levels.

Ofqual now has the necessary expertise to investigate problems with the exam system and this forms a key part of its work as regulator. Ofqual is accountable to Parliament, predominantly through this Committee. In the exceptional event that a wide-ranging and in-depth inquiry is required into the decisions and actions of Ofqual, the Government and the House of Commons should ensure that this Committee is adequately resourced to enable it to investigate the technical processes and procedures in question.
# Introduction

1. GCSE English is a high stakes qualification for young people, their teachers and their schools and colleges. Grade C or above in GCSE English is vital for progression to further education and employment. It is a key threshold in school accountability measures, which are used by Ofsted in forming inspection judgements. GCSE English qualifications are also a significant source of income for exam boards, particularly those with large market shares. When things go wrong with such a qualification, the consequences are very serious for all involved, but particularly for the young people sitting the examination.

2. The 2012 GCSE English results prompted significant concern among students, parents and teachers, with many schools receiving unexpected and disappointing results, particularly around the grade C/D borderline. This ultimately resulted in an application for judicial review being made against two exam boards and the regulator. The legal challenge was unsuccessful, but the episode has highlighted several issues with the design features of GCSE English qualifications and their associated procedures, as well as raising questions about the role of Ofqual and the exam boards in qualifications development. It also illustrates starkly the interaction between the exams system and school accountability measures.

## Our inquiry

3. Responding to public concern, we held a series of oral evidence sessions in September 2012 and, following the outcome of the judicial review, in March 2013. We heard evidence from school leaders, exam board representatives, the regulator Ofqual (on two occasions), the Welsh Minister for Education and Skills and one of his officials, and the Secretary of State for Education.

4. In September 2012 we put a substantial list of questions to Ofqual about the 2012 GCSE English results. We published these questions and Ofqual’s responses on 2 November 2012.¹ Although we did not set out terms of reference for our inquiry or request written evidence, we have received a range of correspondence from schools, local authorities, teaching unions and parents, detailing their concerns about the 2012 GCSE English results. This correspondence has helped to inform our inquiry.

5. We have benefited from the expertise of our standing adviser on education, Professor Alan Smithers, and of our specialist advisers for this inquiry, Dr Newman Burdett and the late Kathleen Tattersall. All three shared their knowledge and experience of the examination system with us and this has proved invaluable.²

---


² Professor Smithers, Director of the Centre for Education and Employment Research, University of Buckingham, declared no interests. Dr Newman Burdett, Head of the Centre for International Comparisons, National Foundation for Educational Research, declared the following interests: until August 2012 freelance consultant employed by OCR as Deputy Chair for Science, and NFER is working with AQA on developing new qualifications. Kathleen Tattersall
6. This report does not re-examine the issues investigated as part of the judicial review. We focus instead on the lessons to be learned, highlighting some of the weaknesses in the exam system which have emerged and need to be addressed, and also the ways in which school accountability measures have an impact on schools’ behaviour with regard to exams. Some issues, particularly concerning qualifications design, are especially pertinent given the forthcoming GCSE and A level reforms.

What happened and when

7. GCSE results were published on 23 August 2012. Almost immediately concerns were raised about the results for GCSE English qualifications. Joint Council for Qualifications (JCQ) figures showed that the percentage of candidates achieving A*-C in GCSE English dropped from 65.4 per cent in 2011 to 63.9 per cent in 2012, a decrease of 1.5 per cent.³ Behind this overall slight drop was considerable variability in results at school level, with some schools achieving results that were significantly lower than predictions or in previous years and others doing better than expected.

8. On 25 August 2012 Ofqual announced that it was conducting an investigation into how the grade boundaries for GCSE English had been set. It published a report with the findings of this investigation on 31 August 2012.⁴ The report concluded that the June grade boundaries had been set at an appropriate level, comparable with the standard in previous years, and following correct procedures. Ofqual found that the issue lay with the January 2012 grade boundaries, based on a much smaller cohort of candidates, which were too generous. Ofqual announced in this report that young people would be offered an exceptional, one-off re-sit opportunity in November 2012, instead of having to wait until the following January.

9. On 10 September 2012, the Welsh Government published a regulatory report, GCSE English Language 2012: an investigation into the outcomes for candidates in Wales. The report found that the 2012 summer awards had not delivered comparable outcomes for candidates in Wales compared to 2011, with a 3.9 per cent drop in overall results at grade C and above which was “not secure or supported by any reasonable justification.”⁵ The report recommended that WJEC should re-grade its GCSE English Language exams to achieve outcomes that were as similar as possible to the outcomes achieved by candidates in 2011.⁶

10. On 11 September 2012, Leighton Andrews AM, Minister for Education and Skills in Wales, issued a direction to WJEC, requiring WJEC to re-grade its GCSE English Language candidates in Wales and to re-issue results within seven days. A week later, on 18
September, 2,386 candidates in Wales received revised (increased) grades. This included over 1,200 candidates whose grade increased from a D to a C. Mr Andrews has said that “these revised grades are a fairer and truer indication of candidates’ ability than the original grades”.7

11. On 26 October 2012 an alliance of six professional bodies, 42 councils, 150 schools and 167 pupils launched a legal challenge against AQA, Edexcel and Ofqual regarding the grading of GCSE English in 2012. The judicial review is discussed in paragraphs 16 to 21.

12. Ofqual published the second and final report on its investigations on 2 November 2012. The second report made no change to Ofqual’s initial conclusions. It found that the overall grading of GCSE English in June 2012 was a “fair reflection of the performance of pupils as a whole”.8 Examining and awarding worked as it should, but with “unpalatable consequences”.9 Ofqual found that flaws in the design of the GCSE English qualifications led to some schools experiencing significant variations in results. The report identified the highly flexible modular structure, the high proportion of controlled assessment in GCSE English (60 per cent) and a generous standard marking tolerance (+/-6 per cent) as key factors.10

13. As in its first report, Ofqual pointed to evidence of over-marking by teachers on controlled assessment units, with “unprecedented clustering around perceived grade boundaries”, and suggested that teachers had given marks that anticipated the grade boundaries that exam boards would set.11 Ofqual concluded that the combined effect of the design flaws meant that GCSE English, a very high stakes qualification, was susceptible to pressures from the accountability system, with teachers striving to achieve the best possible outcomes for their students and schools.

14. Ofqual’s findings were disputed by some school leaders, with the Association for School and College Leaders (ASCL) commenting that “for Ofqual to suggest that teachers and schools are to blame is outrageous and flies in the face of the evidence”.12

15. Over 46,000 young people took part in the exceptional, one-off November re-sits, offered free of charge following concerns raised about the summer results. The results of the re-sits were published on 19 December 2012. 36,000 of those re-sitting had achieved a grade D in summer 2013. According to Ofqual, approximately 30 per cent of those students (9,400) gained a grade C in the November re-sit.13 Exam board AQA said that the

---

7 Oral Statement, 25 September 2012
8 Ofqual press notice with second report, 2 November 2012
9 Ofqual second report, paragraph 2.1
10 Controlled assessment is work that is set by exam boards, produced by students under controlled conditions, and marked by teachers. Teachers’ marking is then moderated by exam boards. The marking tolerance is the amount by which teachers’ marks may differ from the moderator’s before the marks are adjusted.
11 Ofqual second report, paragraph 1.13 and 1.32
12 Ofqual report demonstrates that independent investigation is needed into GCSE English, ASCL press release, 2 November 2012
13 Ev 65
results were “very much in line with what we’d expect” based on previous November re- 
sits.14

The judicial review

16. An alliance of six professional bodies, 150 schools, 167 pupils and 42 councils made an 
application for judicial review against exam boards AQA and Edexcel, and the regulator 
Ofqual. The judicial review hearing was held on 11–13 December 2012. The alliance 
argued that the decision to raise grade boundaries “by an unprecedented margin” between 
January 2012 and June 2012 left students victim to “inconsistency within the year group” 
and a “lack of even-handedness.” They called for a re-grading of the June results, in line 
with the January grade boundaries, estimating that 10,000 young people had missed out on 
a grade C as a result of the changes.15 School leaders argued that “these decisions have 
prejudiced the life chances” of the affected children, given that a GCSE English grade C is a 
crucial benchmark used for entry to further education, vocational training and 
employment.16

17. Edexcel told the High Court on 12 December that grade boundaries had to be raised for 
the June exams to stop “an unacceptable lowering of the standard required for a C in 
GCSE, as compared with the standard in previous years”, thus preventing “grade inflation 
without any academic justification.” AQA argued that that while examiners had used 
statistical predictions to inform their judgements, they had “ultimately reached their own 
conclusions and set a higher boundary than that indicated by the predictions to reflect their 
judgement of the standard of the work that merited a C grade”.17 Ofqual argued that its 
actions were “clear, principled and transparent. We had a duty to maintain standards over 
time and between organisations”.18

18. The outcome of the judicial review was announced on 13 February 2013, with the 
judgement in favour of the exam boards and Ofqual. The judge, Lord Justice Elias, 
dismissed the application for judicial review on all counts, although he stated in his 
summing up that “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”.18 He concluded that “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 [...] and not any unlawful action by either 
Ofqual or the AOs [awarding organisations or exam boards]”.19 The findings discuss in 
some detail the setting of grade boundaries in January and June by AQA and Edexcel, as 
well as Ofqual’s role in this. Lord Justice Elias considered how statistical information was

---

14 Thousands receive improved GCSE English grades in re-sit, Daily Telegraph, 19 December 2012
15 Pupils and educational alliance go to court for GCSE re-grading, ASCL press release, 26 October 2012
16 Exam boards defend GCSE English role in High Court, BBC News, 12 December 2012
17 Exams watchdog Ofqual rejects claims GCSE English grades were ‘unfair’, BBC News, 13 December 2012
18 Judgement by Lord Justice Elias in R v AQA, EDEXCEL and Ofqual,[2013] EWHC 211, paragraph 157. Hereafter, 
‘Judgement’
19 Ibid.
used by exam boards to set the boundaries, alongside examiners’ judgement about the quality of candidates’ work, concluding that there was “nothing improper” in predicted outcomes playing a very significant role in grade awarding.20

19. Lord Justice Elias endorsed Ofqual’s view that it was clear, with hindsight, that the January cohort had been treated more leniently than the June cohort, leading to the problem that “Ofqual could not remedy any unfairness between the January and June cohorts without creating further unfairness elsewhere”.21 He explained that, had the same standard been applied in June as in January, there “would have been a significant dilution of standards” and that this was “contrary to the vitally important objective of maintaining the currency of the qualification.” The judge acknowledged that “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”.22

20. Members of the alliance which brought the legal challenge said, understandably, that they were “bitterly disappointed” by the outcome of the judicial review. Sir Steve Bullock, Mayor of Lewisham, commented that “the judge accepted that the case exposed unfairness and that it was right that this was properly investigated in the court room. But that is no consolation for the thousands of students up and down the country who will have to continue to live with the consequences of this unfairness.”23 School leaders have pointed to the wider implications of the case, with Russell Hobby of the National Association of Head Teachers (NAHT) stating that “the fact that a system cannot be managed without unfairness is a strong indictment”.24

21. Ofqual and the exam boards emphasised the importance of learning lessons for the future, particularly in the context of forthcoming GCSE and A level reform. Ofqual stated that “the court agreed with our conclusion that the root of the problems was the poor design of the GCSE English qualification. We want much better qualifications than this, and it is time to look to the future”.25

---

20 Judgement, paragraph 78
21 Judgement, paragraph 123
22 Judgement, paragraphs 122, 153 and 155
23 NAHT alliance response to GCSE verdict, 13 February 2013
24 Ibid.
25 Chief Regulator’s statement on GCSE judicial review decision, 13 February 2013
2 GCSE English and GCSE English Language qualifications

GCSE English syllabuses in 2012

22. Three new GCSE English qualifications were introduced for first teaching in England in September 2010, namely GCSE English Language, English and English Literature. These three courses replaced two GCSEs in English and English Literature.

23. Students in England can take either GCSE English Language and GCSE English Literature or a single GCSE English (which covers both language and literature). The controversy over last summer’s results centred on GCSE English and GCSE English Language (both of which count towards the 5A*-C including English and mathematics school performance measure). Summer 2012 was the first time that these qualifications had been fully awarded. The qualifications will run until new English GCSEs are introduced for first teaching in September 2015 and first examination in summer 2017.

Key factors in the problems experienced last year

24. Ofqual’s investigation found that many of the problems experienced with GCSE English26 in summer 2012 were linked to design flaws in the current English GCSEs. The judicial review concentrated on the probity of the grade awarding process and the way Ofqual and the exam boards sought to deal problems once they arose, rather than examining the design features of the qualifications in depth. However, in his judgement, Lord Justice Elias endorsed Ofqual’s view, concluding that “it was indeed the structure of the qualification itself which is the source of such unfairness as has been demonstrated in this case,” rather than any unlawful action by Ofqual or the exam boards.27

25. Key features highlighted by Ofqual include: the modular structure with a high degree of flexibility, the high proportion of controlled assessment and generous standard marking tolerances, all combined with significant pressures from the school accountability system. Ofqual has suggested that no single factor was responsible; it was the combination of these factors which came together uniquely in English and which proved problematic. Ofqual’s investigation found that the combination of features and the context in which the qualification would operate were not considered during the qualifications design phase.28

26. Since summer 2012, changes have already been put in place, and others are either pending or proposed, which seek to address some of the problems experienced with GCSE English last year. In this chapter we explore the features of GCSE English qualifications which contributed to the problems experienced in 2012, as well as the interaction of these qualifications with the school accountability system. We consider whether sufficient

---

26 GCSE English is used to refer to GCSE English and GCSE English Language, unless specified.
27 Judgement, paragraph 157
28 Ofqual second report, paragraph 2.3
measures have been put in place, or are proposed, to address the weaknesses identified and to strengthen the current English GCSE qualifications until they are replaced in summer 2017.

**Modular/linear structure**

27. Current GCSEs were designed from 2007–09 under the previous Government. Most of these GCSEs were introduced for first teaching in September 2009, with first full awards in summer 2011. GCSEs in English, ICT and mathematics were introduced a year later. As part of these changes, all GCSEs became modular. Prior to this, exam boards mostly offered linear GCSEs, with modular GCSEs available in some subjects.29

28. Glenys Stacey, Chief Executive of Ofqual, told us that modularity, and the high degree of flexibility it afforded students and teachers, contributed to the problems with GCSE English and led to variability in results and unfairness. Ofqual indicated to us that there were 2550 possible routes through GCSE English/English Language,30 with Glenys Stacey commenting “that is a level of complexity that no teacher ought to have to navigate through.” She added:

> we know from our subsequent analysis that some students fared better than others because of the route they took. That, in our view, cannot be fair.31

29. In his judgement, Lord Justice Elias concluded that “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.”32 The latter meant that teachers used this information to make assumptions about where grade boundaries would be set in future. Ofqual announced changes in its second report, which mean that, from 2013, January units for English GCSEs are marked but not graded, and exam boards grade both January and June assessments at the same time in the summer. Changes announced by the Coalition Government in the 2010 White Paper, *The Importance of Teaching*, mean that GCSEs taught from September 2012 and to be examined in summer 2014 are linear.33 Ofqual has welcomed this change, saying that it will be easier to ensure that exams are fair.34 The new GCSEs, planned from September 2015 in some subjects, including English, and to be examined for the first time in summer 2017, are also intended to be linear.35

30. Under the previous Government, GCSEs changed from mostly linear to modular, which, combined with other changes, brought turbulence to the system and

---

29 In a linear qualification, candidates take all examinations at the end of the course. In a modular structure, the course is split into modules or units and candidates can take units at different intervals throughout the course, with opportunities for re-sits to improve their performance.

30 Ofqual’s response to Committee’s question 7

31 Q344

32 Judgement, paragraph 152

33 2010 White Paper, *The Importance of Teaching*, paragraph 4.49

34 Q340

35 Letter from Michael Gove to Glenys Stacey, 6 February 2013
contributed to the problems experienced with GCSE English in 2012. We recommend that, when considering their reforms of GCSEs and A levels, current Ministers think carefully about the cumulative impact and risks of change.

**Balance of internal and external assessment**

31. In GCSE English and GCSE English Language controlled assessment, which is internally assessed, accounts for 60 per cent of the overall marks. Internally assessed units are marked by teachers and teachers’ marking is checked or moderated by the exam board, to ensure that the marking is consistent and in line with required standards.

32. Current GCSEs have 0 per cent, 25 per cent or 60 per cent of marks awarded by controlled assessment. For example, GCSE Mathematics has no controlled assessment, while science GCSEs have 25 per cent. GCSE English Literature has 25 per cent controlled assessment. The only other English Baccalaureate subject which has 60 per cent controlled assessment is modern languages.

33. Having 60 per cent controlled assessment means that a high proportion of assessment for GCSE English is marked and administered by teachers. Ofqual noted in its November 2012 report that English and English Language are not the only subjects with 60 per cent controlled assessment, “but they are the most high-stakes GCSEs with 60 per cent controlled assessment”. Having such a high proportion of controlled assessment in a high stakes qualification puts teachers in a sensitive position: they are given a high degree of control over the assessment in a qualification for which there is strong pressure to deliver good results, both for their pupils and for their schools.

34. Exam board representatives told us that they raised concerns during the qualifications design phase that the proportion of controlled assessment was too high in GCSE English, but these concerns were not acted upon. For example, Ziggy Liaquat, Managing Director of Edexcel, told us that “when we were designing the qualification we asked for there to be 40 per cent controlled assessment in English”. Andrew Hall of AQA was regretful that “we were not anything like forceful enough, I think, in making our point heard”.

35. Ofqual told us that the decision about the proportion of controlled assessment in GCSEs (0, 25 or 60 per cent) was taken by the executive of its predecessor body the Qualifications and Curriculum Authority (QCA), a non-departmental public body reporting to Ministers, in 2007. According to Ofqual, “beyond that time, all subjects seemed able to do was to decide which of those three weightings they went for”.

36. Ofqual’s November 2012 report notes that discussions about qualifications design “were focussed on the curriculum, underpinned by a belief in a common, unitised
approach across all qualifications at similar levels”.\textsuperscript{41} It appears that curriculum considerations and the desire for qualifications to be the same overrode the concerns of assessment specialists. There was also insufficient focus on the context in which GCSE English qualifications would operate in schools. Ofqual found “little evidence of awareness of the tensions this would create, given their [the qualifications’] central role as a performance measure for schools” during the qualifications design phase.\textsuperscript{42}

37. Ofqual has recently published proposals for changes to the way speaking and listening is assessed in the current English GCSEs (see speaking and listening below). The proposals involve a change in the balance of internal and external assessment for the current English and English Language GCSEs, so that controlled assessment would account for only 40 per cent of the overall marks and the weighting for the externally assessed written exam would increase to 60 per cent. Ofqual states in its consultation document that “written exams are more resilient to pressures on schools from accountability measures and therefore we are proposing to adopt the highest weighting for the written papers”.\textsuperscript{43} Ofqual is proposing that these changes would be introduced from summer 2014, so for pupils who are currently in year 10 of their GCSE course. The Association of School and College Leaders (ASCL) has said that “the changes proposed by Ofqual are understandable and we will consider them carefully.” It has, however, voiced concern about the timescale, saying that “as a matter of principle, changes to assessment should never be introduced after students have started a course”.\textsuperscript{44}

38. We welcome Ofqual’s proposals to increase the weighting of external assessment in the current GCSE English and GCSE English Language qualifications from summer 2014, as we agree that this will help to make the qualifications more robust and more resistant to pressure from the school accountability system. We note the concerns expressed by school leaders about the timescales, but believe that benefits of the proposed action outweigh the downsides, particularly given the assurances from Ofqual that it will take steps to minimise any advantage or disadvantage to students caused by the change.

39. It seems likely that in future the balance between internal and external assessment in GCSEs may be rather different than it is at present for some subjects. The Secretary of State has indicated that internal assessment in revised GCSEs “should be kept to a minimum and used only where is a compelling case to do so”.\textsuperscript{45}

40. It is clearly better to decide on a subject-by-subject basis what an appropriate proportion of internal assessment should be, rather than imposing central requirements or models to which all subjects must conform. We welcome Ofqual’s indication that in future these decisions will also take into account the context in which

\textsuperscript{41} Ofqual second report, paragraph 7.9
\textsuperscript{42} Ibid., paragraph 1.48
\textsuperscript{43} Consultation on the Removal of Speaking and Listening Assessment from GCSE English and GCSE English Language, Ofqual, 25 April 2013, p8
\textsuperscript{44} Comment on proposed changes to GCSE English, ASCL, 25 April 2013
\textsuperscript{45} Letter from Michael Gove to Glenys Stacey, 6 February 2013
the qualifications will operate. This is particularly important for GCSE English, given the current proposals for GCSEs in English and mathematics to remain part of a headline threshold accountability measure as part of the new secondary school accountability measures. We look forward to examining Ofqual’s proposals for internal assessment in revised GCSE qualifications in due course.

Moderation procedures

41. A qualification with a high proportion of controlled assessment relies on having robust moderation procedures in place to help maintain standards. As Brian Lightman of the ASCL commented, in such circumstances “you have to have proper safeguards in place.” Procedures for moderating internally assessed work are laid down in the regulatory Code of Practice. Moderation involves exam board moderators checking and marking a sample of work from each school. A standard marking tolerance is set by the exam boards and applied to all A levels and GCSEs. The tolerance is the number of marks by which the teacher’s marks may vary from the moderator’s marks before they are adjusted by the exam board. The standard tolerance used for all GCSE and A level subjects with coursework/controlled assessment is +/-6 per cent of the maximum mark for the unit. It is set by the exam boards.

Moderator feedback

42. Ofqual’s investigation found that “moderators’ feedback to schools was not always sharp enough” and that “in some cases moderators were not as direct with schools as they should have been when they saw signs of over-marking (within tolerance).” Capgemini research commissioned by Ofqual found that if a school or college did not have their marks adjusted by the exam board, “the majority of centres took this to mean that the marking had been accurate”. On the other hand, headteachers have stated that their moderator’s report commented favourably on the accuracy and quality of their marking, and that they found this difficult to square with disappointing results.

43. Exam board representatives assured us that that they have taken steps to improve their moderator feedback to schools. Gareth Pierce of WJEC told us that:

   one clear distinction we need to make is between saying that their work is perfectly acceptable administratively [...] and the different comment of saying they were absolutely spot-on on the standards.

---

46 Q74
47 GCSE, GCE, Principal Learning and Project Code of Practice, May 2011, Ofqual, CCEA and DfES, section 5.
48 Ofqual second report, paragraphs 1.50 and 7.43
49 Report of findings from Centre Interviews, Ofqual second report, appendix 1, p90
50 Q76 (Kenny Frederick and Mike Griffiths)
51 Q276 (Gareth Pierce)
Ofqual is currently checking whether exam boards’ systems and processes can provide enhanced information to schools that have marked generously or severely but within tolerance. We welcome this. Ofqual will then discuss with school leader representatives “how desirable this additional information would be to schools, bearing in mind the additional costs that exam boards would incur to provide it”.52

44. We welcome the steps taken by exam boards to improve their moderator feedback to schools and to make clearer the distinction between administration and standards/marketing issues. We recommend that it is made clear to schools and colleges in moderator feedback if they have been marking generously or severely but within tolerance, and that Ofqual monitors this aspect of exam board communication with schools and colleges more closely in future, to ensure that teacher assessments are fair and accurate.

Teacher marking and tolerances

45. Relatively low numbers of schools had their marks adjusted through moderation for English GCSEs in 2012. For example, for AQA (which has 62 per cent of the market share for GCSE English/English Language), adjustments were made to the marks of 127 centres (5.8 per cent) in English and 86 centres (3.5 per cent) in English Language on unit 3 (reading and writing controlled assessment).53 Ofqual’s investigation found evidence, however, of schools marking generously but within the marking tolerance.

46. Ofqual’s findings are based on analysis by AQA, the largest provider, which compared the marks given by teachers and moderators in a sample of its centres. AQA keyed in the teacher and moderator marks for “approximately every fifth centre.”54 Figures supplied to us by AQA indicate that the sample covered about 20 per cent of AQA’s candidates for English and English Language. Ofqual states that “the results from this sample show clear evidence of over-marking in both English and English Language”.55 The regulator’s report contains charts showing that most schools in this sample tended to mark slightly more generously than the moderator.56

47. Exam boards retain only paper copies of moderation records of schools whose marks are found to be within tolerance. This means that it is labour-intensive for exam boards to enter data electronically, which would enable them to run wider analyses of trends in teacher and moderator marking. Ofqual has recently asked the exam boards how they will review evidence of marking trends within tolerance, ahead of the summer 2013 grade awarding.57 We welcome this. We recommend that Ofqual and the exam boards consider whether changes to moderation systems and processes are needed to ensure

52 Ev 69
53 AQA annex to Ofqual’s second report, p30
54 Ofqual second report, paragraph 6.52 and footnote 59
55 Ibid.,paragraph 6.52
56 Ofqual second report, p58-60
57 Ev 69
that it is easier to analyse and track patterns in schools’ marking of internally assessed work.

48. According to Ofqual, the standard marking tolerance was well established and teachers were aware of it. Ofqual observed that the marking tolerance “meant that schools could be over- or under-marking by up to 6 marks in some units and still have their marks accepted”.\(^{58}\) We are surprised that, given all the assessment expertise residing in exam boards and now in Ofqual, no-one questioned whether a standard tolerance of 6 per cent was appropriate for a high stakes qualification with such a high proportion of controlled assessment. One senior exam board official acknowledged that this was “a fair challenge” when we put this question to him.\(^{59}\) Andrew Hall of AQA suggested that it was because “we did not understand sufficiently[...] the impact of accountability measures.”\(^{60}\) We return to the accountability system later in this chapter.

49. Moving forward, Ofqual recommended in its second report that the tolerance for GCSE English needed to be tightened. The exam boards collectively agreed to reduce the tolerance for GCSE English (to +/-3 per cent of the maximum mark for the unit), with effect from November 2012. We welcome this change. Ofqual is commissioning research to provide an evidence-based method for determining appropriate moderation tolerances for summer 2014 and beyond.\(^{61}\)

50. We recommend that Ofqual and the exam boards consider on a subject by subject basis what an appropriate tolerance might be for new GCSEs and A levels when deciding upon the proportion of internal assessment for each qualification, and that these decisions be informed by the research commissioned by Ofqual.

51. Exam boards have expressed varying views on whether over-marking by teachers contributed to the problems experienced in 2012. In its report for Ofqual, AQA detailed what it described as “compelling evidence” indicating “the high degree of strategic behaviour demonstrated when allocating marks for internally assessed units”.\(^{62}\) By contrast, the other exam boards told us that they had not found evidence of widespread teacher over-marking.\(^{63}\) Ziggy Liaquat of Edexcel stressed, however, that Ofqual’s view was based on the entire cohort, whereas Edexcel’s data accounted for only 10 per cent of the cohort.\(^{64}\)

52. School leaders rejected Ofqual’s findings on teacher over-marking and called into question its evidence base. Kenny Frederick, Principal of George Green’s School in Tower

---

58 Ofqual second report, paragraph 6.54
59 Q205 (Ziggy Liaquat)
60 Ibid. (Andrew Hall)
61 Letter from Glenys Stacey to Pat Glass, 15 April 2013
62 AQA annex to Ofqual’s second report, p33
63 Q264, Q266-7, Q190, Q197 and Q217 (Ziggy Liaquat)
64 Q190 and Q197 (Ziggy Liaquat)
Hamlets (who gave oral evidence to us in September 2012), wrote an open letter to Glenys Stacey, challenging Ofqual to publish its evidence of inflated marking.65

53. Ofqual’s position is that widespread over-marking created inflated marks in the system, which led to exam boards having to make some significant changes to grade boundaries on controlled assessment units between January and June 2012. We note that this is an action which will have been felt collectively by all students, regardless of whether their centre had marked their work accurately or not. Glenys Stacey acknowledged this in November 2012 when she admitted that pupils whose controlled assessment was marked accurately by their teachers may have been penalised and awarded lower grades than they deserved because of over-marking in other schools. Ofqual, however, had no means of identifying these students or recompensing them in any way.66 This, as Ms Stacey acknowledged at the time, is a sobering reflection and a cautionary tale for all concerned. It also relates to the point made in the judicial review judgement that whichever way Ofqual chose to resolve the problems it faced, there was going to be an element of unfairness.67

54. We note the point made by AQA that moderation does not, and cannot, rigorously “police” teacher marking. It was not designed to do so.68 In a subject such as English, absolute marking precision is difficult: there will always be room for debate and a degree of subjectivity. Furthermore, as Mark Dawe of OCR told us, “there also comes a point where the narrower it [the tolerance] gets, the more you might as well just mark the papers yourselves.”69

55. Moderation, as AQA has stated, relies on the professionalism of the teacher community.70 Ofqual’s position on over-marking is an uncomfortable one for teachers, as it calls into question the integrity of some of the profession. We accept Ofqual’s findings about over-marking. However, we can see that its position is not helped by its reliance on a sample from one exam board, by contrasting views among exam board chief executives and by moderator feedback to schools and colleges which has not always been sufficiently clear about marking and standards. Furthermore, we recognise that Ofqual’s action to address over-marking in some schools has led to the unavoidable but highly unsatisfactory situation that students in other schools, whose work was marked accurately by their teachers, may have been penalised. Exam boards and Ofqual must make every effort to ensure that this situation is not repeated in summer 2013.

---

65 Open letter from Kenny Frederick to Glenys Stacey, 7 November 2012
66 Students who were not over marked have suffered, admits Ofqual, Times Educational Supplement, 9 November 2012
67 Judgement, paragraph 153
68 AQA annex to Ofqual’s second report, p31
69 Q274 (Mark Dawe)
70 AQA annex to Ofqual’s second report, p31
Speaking and listening

56. Particular issues emerged with the assessment in GCSE English and GCSE English Language of speaking and listening, which is worth 20 per cent of the overall marks. Controlled assessment in English consists currently of one third speaking and listening and two thirds written controlled assessment. The evidence for speaking and listening is not collected and moderated in the same way as written assessments. Moderators visit schools to observe a sample of students doing their speaking and listening tasks. According to Ofqual, about a third of schools are visited each year, with schools visited on a three-year cycle. Further visits are undertaken to schools where marking is a cause for concern. The judicial review judgement noted Edexcel’s view that the speaking and listening element “had been very lightly moderated, and had led to mark inflation at least across all the AOs [exam boards] as a whole.”

57. Ofqual has told us that it is “not satisfied that we can assure and protect standards in speaking and listening as the qualification is currently designed.” It has recently launched a consultation on proposals which would mean that speaking and listening no longer contributes to a student’s overall mark and grade in GCSE English. Speaking and listening would be assessed in the same way as at present, but achievement would be reported separately and be shown as an endorsement on GCSE certificates. The overall grade would be calculated using only marks from the written papers and reading/writing controlled assessment. Crucially, the change will involve an increased weighting for the exam papers, moving from 40 to 60 per cent (see balance of internal and external assessment above). Ofqual is proposing that the changes will take effect from summer 2014. These are the last planned changes as part of Ofqual’s action to strengthen the current GCSE English qualifications.

58. Ofqual has warned that the change, which is favoured by several exam boards, would have an impact on results. Its consultation document states that:

> overall results in these qualifications will fall if these changes are implemented without any further action on our part, because students generally do better in speaking and listening than in the rest of the qualification. The proportion of candidates attaining grades A*-C would drop noticeably.

Ofqual is therefore proposing to use its comparable outcomes approach to minimise any advantage or disadvantage to students caused by the change.

59. Some have questioned whether it is fair for young people not be credited for performance in what was conceived as an integral part of the course. School leader

---

71 Ofqual second report, p117 and AQA annex, p30
72 Judgement, paragraph 133
73 Q387
74 Consultation on the removal of Speaking and Listening Assessment from GCSE English and GCSE English Language, Ofqual, April 2013, p2
75 Eg : Q241 (Gareth Pierce)
representatives have emphasised the importance of speaking and listening skills. Russell Hobby of the National Union of Head Teachers (NAHT) commented that “it is all very well to say they should still be taught even if they are not in the exam, but, in the current high stakes system, if they are not tested for the league tables, they won’t count.” The Association of School and College Leaders (ASCL) has expressed similar concerns, but acknowledged that “the changes proposed by Ofqual are understandable”.77

60. **The problems experienced with GCSE English in 2012 highlighted serious weaknesses in the moderation of speaking and listening and the consequences for grade awarding. While we agree that speaking and listening are important skills, the current assessment arrangements are not robust enough to ensure that assessment is a reliable, fair and accurate reflection of students’ performance. This risks devaluing the assessment of those skills and also generates further problems in securing standards across the qualification as a whole. On balance, we welcome the action proposed by Ofqual to address the weaknesses in the assessment of speaking and listening.**

**Pressures from the school accountability system**

61. Ofqual concluded that the pressures of the accountability system contributed to the problems experienced with GCSE English in 2012. Its view was that the structural weaknesses of English GCSEs, as highly flexible, modular qualifications, with a high proportion of controlled assessment, meant that the qualifications were susceptible to pressures, as “teachers strove for the best possible outcomes for their students and school”.78 Ofqual stated in its November 2012 report that:

> while no school that we interviewed considered that it was doing anything untoward in teaching and administering these GCSEs, many expressed concerns that other nearby schools were overstepping the boundaries of acceptable practice. It is clearly hard for teachers to maintain their own integrity when they believe that there is widespread loss of integrity elsewhere. No teacher should be forced to choose between their principles on the one hand and their students, school and career on the other.79

62. We reiterate the point we made in our 2012 report on the administration of exams for 15-19 year olds that one should not underestimate the extent to which the accountability system incentivises schools to act in certain ways with regard to exams. Having a significant proportion of internal assessment in a high stakes qualification gave teachers a high degree of control over outcomes against which they and their schools were being judged. It is entirely understandable, and should come as no surprise, that teachers sought to achieve the best outcomes for their pupils and schools. As Andrew Hall of AQA said to us:

---

76 GCSE English plan leaves speaking test out of final grade, 25 April 2012, BBC News
77 Comment on proposed changes to GCSE English, ASCL press release, 25 April 2013
78 Letter from Glenys Stacey to Graham Stuart, 2 November 2012, published as preface to Ofqual’s second report
79 Ofqual second report, paragraph 1.34
I do not think someone sets out to do something fundamentally bad, but we put them in a position where we are asking them to make really quite complex judgements with 60 per cent of the marks in a high stakes exam [...] I think they are being put in a position where their judgements are influenced by the pressures of the accountability system, and that is natural human behaviour. If you set up a system of measurement, do not be surprised that people use it.\textsuperscript{80}

63. We note that an interesting contrast can be found in GCSE Mathematics. This is a similarly high stakes qualification and subject to similar pressures, but with no controlled assessment. It did not experience the same difficulties as GCSE English in summer 2012.

64. Ofqual’s investigation found that “insufficient attention was given to the totality of the changes and incentives on schools” during the qualifications design phase of the current English GCSEs. This has clear implications for qualifications design in the future. Glenys Stacey, Chief Executive of Ofqual, told us that one of the key lessons from the 2012 GCSE English results is that “decisions about the design of the qualification and the detailed design of assessment need to be made in the real world, with a recognition of not simply what is regarded as best assessment practice, but how that might play out in the real world of schools”.\textsuperscript{81}

65. The Government is currently consulting on its proposed changes to secondary school accountability measures. The proposals include a headline threshold measure showing the percentage of pupils achieving a pass in English and mathematics. This measure will be part of the floor standard which all schools must aim to reach. Revised English GCSEs will therefore continue to be high stakes qualifications, not only for individual students, but also for their schools and teachers.

66. We welcome Ofqual’s indication that it will take into account the context in which qualifications operate when regulating and planning qualifications reform. This is especially important for GCSE English, given that it looks likely to remain part of a headline threshold measure, as outlined in the Government’s proposals for secondary school accountability. We recommend that Ofqual indicates publicly and clearly when and how accountability measures are a factor in its decision about how a qualification is designed.

\textsuperscript{80} Q212 and Q214
\textsuperscript{81} Q380
3 The role of Ofqual in qualifications design

67. It is accepted that the regulatory body bears some responsibility for the events of last summer. For example, Lord Justice Elias concluded that the structure of English GCSE qualifications was the source of the difficulties experienced in 2012 and that the structural problems with GCSE English may have been “of Ofqual’s own making (or at least, Ofqual’s predecessor)”.

Ofqual itself criticised decisions made by its predecessor bodies in the design of the current English GCSEs in its November 2012 report, which found that:

> curriculum considerations and a common approach for qualifications appear to have been given more weight than standards when the new qualifications were proposed. Little attention seems to have been given to the difficulties of maintaining standards in a set of new qualifications of such complexity.

The question therefore arises of how these problems occurred and precisely which body was responsible.

68. The qualifications development phase of the current English GCSEs, from late 2007 to late 2009, coincided with a period of organisational change for the regulator, whereby the regulatory function was separated from the Qualifications and Curriculum Authority (QCA), a non-departmental public body reporting to Ministers, and an independent regulator, reporting to Parliament, was established. From April 2008 until April 2010 Ofqual operated in interim form as a regulator, but part of QCA, pending changes in legislation.

69. The interim regulator did not have any statutory powers and was not established as an independent body. A sub-committee of the QCA board was created, chaired by a QCA board member. The framework for the operation of the interim regulator was set by a direction from the Secretary of State issued under the Education Act 1997, and by memoranda of understanding between the then Department for Children, Schools and Families and QCA, and between QCA and the interim regulator. On 1 April 2010 Ofqual was established by the Apprenticeship, Skills, Children and Learning (ASCL) Act 2009, as a non-ministerial government department, reporting directly to Parliament.

---

82 Judgement, paragraph 152
83 Ofqual second report, paragraph 1.48
84 See Ofqual second report, paragraphs 7.2-7.7
85 Introducing Ofqual, 2010/11, Ofqual 2010
The impact of organisational change on qualifications development

70. According to Ofqual, QCA developed the GCSE qualification criteria for the current English GCSEs in late 2007. It was then that the decision was made by the QCA board that all GCSEs would have 0, 25 or 60 per cent controlled assessment. In July 2008, QCA developed and consulted on new GCSE English subject criteria. In 2009 exam boards developed new English GCSEs in accordance with the qualification and subject criteria. The interim Ofqual accredited the current English GCSEs in December 2009, for first teaching in September 2009. Ofqual’s Director of Standards, Dennis Opposs, told us:

The arrangement in 2008 was that it was QCA who was responsible for developing the criteria, for consulting on them and for finalising them, and then passing them to Interim Ofqual. It was not Interim Ofqual’s job at that time to keep a close eye on, or be responsible for, the development of the subject criteria. Interim Ofqual’s job was to pick up those criteria at the end and then sign them off.

71. Evidence from the exam boards suggests that they raised their concerns about the design of the current English GCSEs with the interim regulator, rather than QCA, although they stressed that the interim regulator was smaller and lacked the assessment expertise which Ofqual now has. Mark Dawe of OCR told us that “they say [Ofqual], ‘it is not us’, but it was the original form of the regulator that caused some of these problems with the original design”.

72. Ofqual rejected the suggestion that organisational change was linked to the problems with the design of the current English GCSEs, although Glenys Stacey did acknowledge the commonly held view that substantial change can increase the risk of problems, “because you are imposing an additional demand on your very senior managers.” She told us that “I have seen no evidence to suggest that any turf war or difficulty of creating the new organisation actually has any relevance to those decisions [about qualifications design].” Dennis Opposs stressed that proper care and attention was taken in the original design of the current English GCSEs and that “it would have been very difficult, thinking back to 2008, to imagine the sorts of scenarios that we saw in schools in 2012”.

---

86 GCSE qualification criteria stipulate the structure of GCSEs and their assessment and grading arrangements, as well as their general aims. The regulator produces the GCSE qualification criteria, in line with Government policy steers, and consults upon them.
87 In some subjects (mostly high entry subjects, where more than one exam board offers a GCSE), GCSE subject criteria are produced. These contain more subject specific rules, such as the assessment objectives and the proportion of internal assessment.
88 See Ofqual second report, paragraphs 4.1 to 4.26 and 7.1 to 7.22
89 Q368 (Glenys Stacey)
90 Q180-181 (Ziggy Liaquat), Q182-183
91 Q278
92 Q373
93 Q375
94 Q376
73. Ms Stacey was critical of the way that the interim regulator had been encouraged to brand itself as a different organisation, when it lacked independent powers and authority. She was also clear that having the regulator part of the curriculum authority, as was the case in QCA, had meant that the design rules gave “undue weight to curriculum considerations and did not give sufficient regard to the complexity of awarding or to standards”. This, she argued, underlined the need for an independent regulator, separate from curriculum policy.

74. The fact that the qualifications design phase of the current English GCSEs coincided with a period of organisational change for the regulator has made it challenging for us to understand who did what and when. The evidence we heard would suggest that there is confusion even among the exam boards, who were closely involved at the time. While we accept Ofqual’s argument that organisational change did not directly contribute to the structural problems with English GCSEs, it illustrates the point made by Cambridge Assessment to our previous inquiry that the greatest threat to standards comes at moments of change, even if that change is broadly welcomed.

75. **Contrasting evidence from the exam boards and Ofqual would suggest some confusion about roles and responsibilities with regard to qualifications development while the interim Ofqual was still part of QCA. Encouraging the interim Ofqual to brand itself as a different organisation when it remained a sub-committee of the QCA may have led to confusion over its role, responsibilities and powers at the time.**

76. Ofqual is now responsible for developing, consulting on and finalising the aspects of new qualifications that it regulates, such as assessment structures and grading arrangements, and for accrediting the qualifications. We believe that the decision to separate regulation from curriculum policy and to establish an independent regulator was sound and that the current status of the regulator, independent from Government, and accountable to Parliament, is the right one, but current developments will put this to the test (see Forthcoming A level and GCSE reforms below).

**Warning voices**

77. Exam board representatives told us that they raised concerns about the current English GCSEs during the development and accreditation phase, regarding the high proportion of controlled assessment and technical difficulties in standards and awarding. Ofqual’s November 2012 report supports this, stating that “we understand that English subject experts in exam boards tried unsuccessfully to keep proportions as they were, and influence the balance ultimately struck by QCA”. Glenys Stacey also agreed that “it is true to say that technical experts—in my view, late in the day—raised concerns about the

---

95 Q365  
96 Q343  
97 See Q333, Q343, Q373  
98 Administration of examinations for 15 to 19 year olds in England, HC 141 (2012-13), paragraph 50  
99 Ofqual second report, paragraph 7.16
difficulty of awarding in modular qualifications, modular GCSEs. It was not particularly English, but modular GCSEs”.100 She confirmed that the concerns were considered but no changes were made, as this “really would have been undermining entirely the concept of modularisation”.101

78. Mark Dawe of OCR told us, “if there are lessons learned, it is that when all the awarding bodies [exam boards] are saying, ‘This will not work. This will cause problems.’ They should not be ignored”.102 Ofqual has acknowledged that:

the views of technical assessment experts in exam boards matter, and they need to be consulted as new qualifications are designed. There is some evidence that this did not happen sufficiently or at the right time when these GCSEs were being designed by our predecessor body.103

79. It is clear that warning voices regarding potential problems were raised but not acted upon during the development and accreditation phases of the current English GCSEs. While innovation and change is healthy and essential in any examination system, one of the crucial lessons that must be learned from this episode is that Ofqual and Ministers should listen when concerns are raised, especially when they come from specialists in the field. Balancing innovation and change with sound, specialist advice is the hallmark of a robust and high quality examination system.

Forthcoming GCSE and A level reforms

80. On 6 February 2013 the Secretary of State announced that the Government would like to see new GCSEs in English language, English literature, mathematics, biology, chemistry, physics, combined science (double award), history and geography introduced for first teaching in September 2015, for first examination in summer 2017. These new GCSEs would be available to schools from autumn 2014 to assist with planning. Changes to GCSEs in the remaining subjects to follow as soon as possible after that, with the aim that they are introduced for first teaching in September 2016, for first examination in summer 2018.

81. Significant reforms to A levels are due to take place at the same time. The Government would like A levels in the “facilitating” subjects except languages, namely English (language, literature and language & literature), mathematics and further mathematics, physics, chemistry, biology, history and geography, to be developed for first teaching in September 2015, for first examination in summer 2017. This is a year later than originally planned, a delay to which the Secretary of State agreed following advice from Ofqual.104 The new timetable, as Ofqual noted in a recent letter to the Secretary of State, “aligns with

100 Q329
101 Q329
102 Q323
103 Ev 68
104 See letter from Michael Gove to Glenys Stacey, 22 January 2013
the GCSE subjects being reformed to the same timescale, and promotes coherence between
the subject content for new GCSEs and A levels". 105

82. Ofqual has also agreed that a further six A level subjects with high numbers of entries
will be reformed for first introduction in September 2015 and first examination in summer
2017.106 According to Ofqual, revised A level in subjects covering about 70 per cent of
entries will be introduced to this timetable. Ofqual’s letter of 21 March 2013 to the
Secretary of State does not specify the timetable for the reform of remaining A level
subjects, but its website indicates that the second group of subjects will be introduced for
first teaching in September 2016 and first examination in summer 2018.107

83. Ofqual will clearly have a pivotal role in the development of the new GCSE and A level
qualifications. Its key task, as Glenys Stacey stated in a letter to the Secretary of State, is to
make sure that standards are right.108 This task becomes especially challenging at times of
change, when the risk of volatility in the exam system is increased. As Ofqual stated
recently, “when qualifications change, results vary more than normal—both between
schools and from year to year—and this variability is inevitable for several years as reforms
work their way through the system”. 109 Increased variability is therefore to be expected,
particularly following significant reform of both GCSEs and A levels.

84. Ofqual has described the timetable set out by the Government for qualifications reform
as “challenging”110 and has indicated that it would, if necessary, delay the reforms. Glenys
Stacey has also warned that “the introduction of new GCSEs and A levels at the same time
will place a considerable burden on schools and the effect of this should be monitored.”111

85. Ofqual’s statutory framework requires it to have regard to Government policy steers,
but not, as Glenys Stacey pointed out to us, to follow them slavishly.112 There are signs that
this may be particularly relevant in connection with the forthcoming GCSE reforms and
the issue of tiering. The Secretary of State is concerned that tiered assessments place a cap
on ambition, by limiting foundation tier students’ attainment to a grade C. He has
indicated that he would like reformed GCSEs to avoid tiering.113 Ofqual has indicated that
there may be some difficulties with this approach, with Glenys Stacey describing it to the
Times Educational Supplement as “quite a challenge, really”. 114 She was clear about
Ofqual’s role in evidence to us, stating that:

105 Letter from Glenys Stacey to Michael Gove, 21 March 2013
106 These are: psychology, art and design, sociology, business studies, economics and computing.
updated on 11 April 2013 and accessed on 1 May 2013
108 Letter from Glenys Stacey to Michael Gove, 6 February 2013
109 Ofqual response to the secondary school accountability consultation, 1 May 2013
110 Letter from Glenys Stacey to Michael Gove, 6 February 2013
111 Letter from Glenys Stacey to Michael Gove, 21 March 2013
112 Q379
113 Letter from Michael Gove to Glenys Stacey, 6 February 2013
114 ‘Challenging’ hurdles may trip up Gove’s exam reforms, Times Education Supplement, 15 March 2013
it is Ofqual’s job to determine whether or not new GCSEs will be tiered or whether there is a better arrangement that could ensure that these qualifications remain universal, as the Secretary of State has asked, and remain as accessible to the same proportion of students as they are now [...] we will make the right decision [...] it may be tiering; it may not, but we have a job to do to get to the right solution.115

86. We asked the Secretary of State recently to clarify the respective roles and responsibilities of the Department for Education and Ofqual regarding qualifications development. Mr Gove told us that “we [the DfE] give a broad indication in terms of curriculum content and how we wish syllabuses to look. They take the final decisions on whether any of our suggestions are consistent with standards being maintained over time. By definition, there is dialogue”.116 Following his oral evidence, the Secretary of State sent us a written statement of how the DfE considered that this arrangement with Ofqual will work.117

87. The Government is embarking upon the most significant and wide-ranging reforms to GCSEs and A levels since Ofqual was established as an independent regulator and has set a challenging timetable. We recommend that the DfE and Ofqual set out in detail their respective roles and responsibilities in qualifications development, particularly regarding how subject content will be developed, and publish this information before their respective consultations on the proposed GCSE reforms have ended.

88. We welcome signs that Ofqual is itself prepared to be a warning voice on qualifications reform to Ministers, if it judges this to be necessary, and we were pleased to see that Ministers acted upon Ofqual’s advice in connection with qualifications and market reform earlier in 2013.

89. Ofqual has taken steps to increase its assessment expertise, as recommended in our 2012 exams report. Ofqual must use this expertise to heed warning voices and take on board technical arguments in qualifications design for the forthcoming GCSE and A level reforms. It should be prepared to put a robust case to Government, should those arguments be contrary to Government policy steers.

90. Ofqual’s role in qualification design is likely to come under significant pressure in coming months as the proposed timetable for GCSE reform coincides with significant changes to A levels. Ofqual should ensure that it has in place robust systems and adequate resourcing to undertake this role effectively, and that it gives explicit advice to Ministers about the risks involved in reforming GCSEs and A levels at the same time. It also needs to raise public awareness of the likelihood of increased variability in results during times of significant changes to qualifications.

115 Q379
116 Oral Evidence session, School accountability, qualifications and curriculum reform, 15 May 2013, Q39
117 See Appendix to this Report
4 Other lessons learned

Exam board communication with schools and colleges

91. Ofqual concluded in its November 2012 report that “communications between exam boards and schools could have been better, and more focused, most especially on the dangers of assuming grade boundaries”.\textsuperscript{118} It suggested that exam board communication to schools about the possibility of changes to grade boundaries “may have seemed like pre-flight safety demonstrations, referring to a possibility so remote that it attracted almost no attention”.\textsuperscript{119}

92. Interviews with schools and colleges conducted by Capgemini for Ofqual found that schools “did pay attention to warnings from awarding organisations about the scope for grade boundaries to change, but did not anticipate the size of shift that occurred.” The research also found that advice given orally by exam boards was “less constrained.”\textsuperscript{120} This is echoed by evidence we heard from school leaders. Kenny Frederick, Principal of George Green’s School, told us that “we were working towards the grade boundaries we had with an expectation that [those] might vary between two or three points”.\textsuperscript{121} Brian Lightman of ASCL stated that “we have plenty of schools that were told by boards—and it is not just one board but all boards—that there would be little change from the January boundaries. They were told that in conversations that they had between January and June”.\textsuperscript{122}

93. Capgemini’s research for Ofqual found that teachers were under considerable pressure to provide accurate predictions for grades in GCSE English, with pressure coming from Ofsted, from school leadership teams and from parents. The survey found that schools had developed “highly sophisticated grade prediction and tracking mechanisms” which, prior to summer 2012, had proved highly accurate in predicting grades for GCSE English.\textsuperscript{123} Headteacher representatives confirmed to us in September 2012 that schools are “extremely strategic” in the way they allocate resource and prepare young people for exams, employing sophisticated prediction processes, data analysis and intervention strategies.\textsuperscript{124} As Ofqual has noted, many teachers consider the ability to predict grades accurately to be “a fundamental part of their role as a professional”.\textsuperscript{125}

94. Exam board representatives acknowledged to us that communication with schools is a significant and ongoing challenge. Ziggy Liaquat of Edexcel explained that “we thought we had explained it well, but it clearly had not landed, and those grade boundary movements

\textsuperscript{118} Ofqual second report, paragraph 1.50
\textsuperscript{119} Ibid., 1.33
\textsuperscript{120} Report of findings from Centre Interviews, Ofqual second report, Appendix 1 pp100-101
\textsuperscript{121} Q46
\textsuperscript{122} Q51
\textsuperscript{123} Report of findings from Centre Interviews, Ofqual second report, Appendix 1, p94
\textsuperscript{124} Q30, Q37 and Q59
\textsuperscript{125} Ofqual second report, paragraph 6.10
were a surprise [to schools] [...] we should be relentless in communicating the fact that these grade boundaries can move”. Andrew Hall of AQA reported that although marks rather than grades had been issued for the January 2013 GCSE English exams, some schools had been using last year’s grade boundaries to make assumptions about grades their students would achieve. He welcomed the input from the teacher associations which had “rallied behind this and are telling their membership that that is exactly the wrong thing to do, because grade boundaries have not been set”.

95. Ofqual has acknowledged that it needs to improve its own communication with schools and colleges about how exams are marked and graded and its role in this. Results from Ofqual’s most recent survey of perceptions of GCSEs and A levels suggest that confidence in GCSEs and in Ofqual has been damaged as a result of events with the 2012 GCSE English results, particularly among teachers and head teachers. Glenys Stacey told us that Ofqual is trying to address this through developments on its website and through engagement with teacher representative groups. She admitted that “the experience of GCSE English engendered a great deal of mistrust, and we have a job to do to explain as clearly as we can what we do, how we do it, and why we do it [...] I think there is never going to be enough that we can say to explain what we do”.

96. As part of its monitoring of the January 2013 GCSE English exam series, Ofqual commissioned an independent review of exam boards’ activities, which included their communication with schools and colleges. The review found that exam boards’ communication is improving, but Ofqual has asked the exam boards to take some specific actions to make further improvements.

97. Exam boards must take very seriously the need to communicate better with schools and colleges, in order to improve trust and confidence in the exam system. We welcome signs that this communication is improving, but, as exam boards have acknowledged, this is a significant and ongoing challenge, and there is more to be done. We also welcome Ofqual’s greater scrutiny of this area and recommend that Ofqual continues to be more vigilant in its monitoring of communication between exam boards and schools, as this area is closely linked to public confidence. We also welcome Ofqual’s plans for improving its own communication with schools and colleges about marking and grading in the summer 2013 exams, in order to help restore confidence in Ofqual’s regulation of exams, particularly among teachers and head teachers.

Grade awarding and the comparable outcomes approach

98. The difficulties experienced with GCSE English in 2012 have drawn attention to how the exam boards and Ofqual use the comparable outcomes approach in GCSE and A level...
grade awarding. One of the key issues in the judicial review was how the approach was used to maintain standards and how it influenced grade boundary setting for English GCSEs.

99. Ofqual has a statutory objective to ensure that standards in the qualifications it regulates are consistent over time and across exam boards, as well as consistent with attainment required for comparable qualifications in other countries. It has also a statutory objective to promote public confidence in qualifications such as GCSEs and A levels. As Lord Justice Elias noted in his judgement the two objectives are closely related, as maintaining the currency of the standard is an important element in maintaining public confidence in the system.

100. Grade awarding is a technical process, involving a blend of the professional judgement of examiners with a range of statistical data. The comparable outcomes approach is a statistical tool, whereby exam boards use data on Key Stage 2 attainment to predict the outcomes for each GCSE cohort. Ofqual states that it uses the approach for two key reasons: “to make sure that students taking new qualifications are not disadvantaged in the first years of those qualifications (to allow for teachers being unfamiliar with the qualification), and to make sure grade standards between exam boards are in line so that students sitting with one board do not have an advantage over others”. It describes the approach as “a check at qualification level”. Ofqual’s position is that results in 2012 stayed steady overall at a national level, in part because exam boards applied the comparable outcomes approach to awarding, as the approach “helps to maintain standards over time”.

101. Ofqual has told us that exam boards have used the approach for some years in GCSE awarding to improve comparability between exam boards. A change in recent years has been that Ofqual has strengthened its requirement for exam boards to justify increases in results outside tolerance. Ofqual has used the comparable outcomes approach at AS and A level since 2009 and 2010 respectively and for GCSEs since 2011. This coincides with an end to year-on-year increases in GCSE and A level results at national level.

102. The alliance which made the application for judicial review accused the exam boards and Ofqual of a “statistical fix”, contending that the performance of students had not been fairly reflected in their grade because the results had been unjustly moulded to reflect predicted performance and that statistics had overly dominated the grading process.

---

131 Education Act 2011, section 22 and ASCL Act, section 28
132 Judgement, paragraph 21
133 Ev 67
134 GCSE English Awarding 2012, Ofqual Supplementary Memorandum to Education Committee, 11 September 2012, paragraph 33
135 Ofqual second report, paragraph 1.29
136 Tolerance limits for overall outcomes in a qualification are agreed between Ofqual and the exam boards. These are typically +/-1 per cent. Exam boards can justify improvements in grades beyond these tolerance limits but must provide evidence of improvement or deterioration in performance.
137 Judgement, paragraphs 9-11 and 149
103. In his judgement, Lord Justice Elias found that there was “nothing improper” in predicted outcomes playing a very significant role in grade awarding, concluding 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.” He acknowledged that Key Stage 2 predictions were “widely thought to be the most reliable statistical evidence currently available for the purpose of comparing performance year on year” and noted that a common approach was important to ensure consistency across exam boards.139

104. Exam board views on comparable outcomes vary. AQA is generally very supportive of the approach. Its chief executive officer, Andrew Hall, told us that “AQA has used past performance data for a number of years” and that “Key Stage 2 data is the best available for GCSE”. Ziggy Liaquat of Edexcel described comparable outcomes as “not perfect as a system” but “as good as any system I have come across”. By contrast, Gareth Pierce of WJEC criticised Ofqual’s “over-deterministic” use of a single predictor model, and called for “balanced use of a range of relevant statistical information.”

105. Glenys Stacey rejected the suggestion that Ofqual was unduly reliant on Key Stage 2 data, but acknowledged that she takes “a different view” to WJEC. There are some limitations to the comparable outcomes approach, which Ofqual accepts. For example, it does not work as well for subjects with a high performance element, such as drama. There are also issues with regard to Wales, where students do not sit Key Stage 2 tests and a different statistical tool, which does not rely on Key Stage 2 predictions, is favoured. Particular difficulties arose for WJEC GCSE English Language in 2012 when a majority of WJEC candidates were in schools in England.

106. Ofqual also acknowledges that the comparable outcomes approach “is not well understood by schools and colleges, and not generally trusted.” Andrew Hall of AQA referred in oral evidence to misconceptions about the approach, saying that “one of the things potentially misunderstood at some stages was that the Key Stage 2 data was generating a quota. That absolutely was not the case”.

107. The comparable outcomes approach has also attracted criticism from the teaching profession on the grounds that the approach makes it difficult for schools to demonstrate

138 Judgement, paragraphs 78 and 149
139 Ibid., paragraph 37
140 Q173
141 Q174
142 Q243
143 Q345-6
144 Annex D to Ofqual’s second report, Report from OCR, p11
145 Ev 76. Due to the majority of candidates being in England, it was agreed that in 2012 WJEC would report their projected outcomes using Key Stage 2 predictions for GCSE English Language and GCSE English (which is not taken by candidates in Wales) only.
146 Ev 68
147 Q173
genuine improvements in performance. Ofqual itself voiced concerns to this effect in a letter to the Secretary of State, stating that “one consequence of this approach is that it can make it harder for any genuine increases in the performance of students to be fully reflected in the results”. More recently, Ofqual has qualified this view, telling us that the approach does allow genuine improvements in performance to be reflected in results, but exam boards have to “present evidence of achievement above and beyond expectations”.

108. Ofqual has taken advice from its Standards Advisory Group on its use of the comparable outcomes approach. Glenys Stacey told us that “comparable outcomes is regarded by assessment experts as the best possible model at the moment, and it looks good when compared with how these things are done in many other countries [...] there is no known better way internationally”. Ofqual has confirmed that it will continue to use the comparable outcomes approach in GCSE and A level grade awarding in summer 2013, but that it will keep the approach under review.

109. As we noted in our 2012 exams report, there are significant challenges for Ofqual and the exam boards in explaining the technical difficulties surrounding grade awarding. The problems experienced with GCSE English in 2012 have brought many of the issues to the foreground, with the judicial review scrutinising the grade boundary setting process in considerable detail. There are, however, lingering misconceptions and concerns about unfairness and, as Ofqual has acknowledged, the experience engendered a great deal of mistrust. Work is needed to improve understanding of and restore confidence in the system, particularly with regard to grade awarding and Ofqual’s role in maintaining standards.

110. Ofqual must continue to make greater efforts to explain the complexities of awarding and standards setting, and its role in this, to schools and colleges, in order to improve confidence in the comparable outcomes approach and in Ofqual’s work as regulator. Ofqual should also keep the comparable outcomes approach under review and be prepared to adapt it in the light of experience and/or expert advice.

---

148 Letter from Glenys Stacey to Secretary of State, 22 August 2012
149 Q354 (Glenys Stacey)
150 Q350
151 Ev 67 and Q350
152 Q339
5 The future of three-country qualifications and regulation

111. Particular problems were experienced with GCSE English qualifications in Wales in 2012, which have had ramifications for the joint regulatory arrangements in place between England, Wales and Northern Ireland. These issues played no part in the judicial review. They have, however, tested relations between Ofqual and the Welsh regulator in particular, and have, along with the effect of increasing divergence on qualifications policy, called into question the operation of A levels and GCSEs as three-country qualifications and their regulation.

112. At present, GCSEs are jointly owned by the three regulators: Ofqual, the Welsh Government and the Council for Curriculum, Examinations and Assessment (CCEA) in Northern Ireland. According to Ofqual, when joint ownership of GCSEs was agreed, “it was assumed that a common standard could be agreed.” This position, however, was tested “to the limit” by events in 2012 and is looking increasingly uncertain.

Disagreements between regulators

113. Ofqual was openly critical of Leighton Andrews’ decision to order re-grading of WJEC GCSE English Language candidates in Wales. It commented in its November 2012 report that “this unilateral decision by the regulator in Wales has resulted in more favourable treatment for the 2,300 students who were upgraded than for their English counterparts [...] it signals significant problems for the future, if we are to maintain common standards across borders”.

114. In February 2013 the Welsh Government announced that in future exam boards would need to set separate exam papers and hold separate awarding meetings for GCSE English Language candidates in Wales. Ofqual has expressed concern about the implications of this decision for comparable grading standards between England and Wales, saying that “different question papers will necessarily mean different grade boundaries. We do not think this a satisfactory situation. It damages the credibility of, and confidence in, the GCSE brand for all students, but especially for those in Wales”.

153 There is currently no independent regulator in Wales. The Welsh Government is also the regulator. As such, the Minister for Education and Skills has legal powers as set out in the Education Act 1997. This includes the power to issue a direction to WJEC to re-grade candidates. In Northern Ireland, CCEA regulates and awards GCSE and A level qualifications.

154 Ev 67

155 For further details, see chapter 1, paragraphs 9 and 10.

156 Ofqual second report, paragraph 1.47

157 Written Statement by the Minister for Education and Skills, Welsh Government, 21 February 2013. The conditions apply when an exam board has more than 5,000 candidates in Wales, currently only WJEC.

158 Ev 67
115. Behind much of the disagreements over English GCSEs and the actions taken by the Welsh Minister for Education and Skills lies a difference of opinion between Ofqual and the Welsh Government about the use of the comparable outcomes approach in grade awarding. Mr Andrews told us that “when comparable outcomes was designed as an approach to ensuring that, when qualifications changed, a cohort did not suffer by being the first cohort to sit that new qualification, it was really about ensuring fairness to candidates [...] since then, Ofqual, particularly in the last 12 to 15 months, appears to see comparable outcomes as mechanism for containing grade inflation”.  

116. Ofqual has denied that containing grade inflation comes at the expense of reflecting genuine improvements in performance, pointing out that the approach does allow for increases in results, providing that exam boards can supply an explanation of why standards have improved and evidence to support this. This argument was accepted by the judge in the judicial review, who concluded that “it was legitimate for Ofqual to pursue a policy of comparable outcomes, to ensure a consistent standard year on year”.  

117. As mentioned in chapter 4, there are limitations to the comparable outcomes approach. The approach, which uses statistical predictions based upon Key Stage 2 national tests, presents difficulties in Wales, where students do not sit externally set and marked national tests in English and mathematics at the end of Key Stage 2. WJEC usually uses other statistical indicators to inform its GCSE grade awarding process. A particular arrangement was agreed for the 2012 WJEC GCSE English Language award, when, due to an increase in its market share, the majority of WJEC’s candidates (71 per cent) were students in England. In these circumstances, Ofqual required that the outcomes for WJEC’s GCSE English Language award were modelled against the Key Stage 2 predictions. The Welsh Government agreed to this, albeit with some reservations, although its position now is that this was to the detriment of candidates in Wales.  

118. The changes put in place by the Welsh Government for summer 2013 mean that there will be separate awarding procedures for GCSE English Language candidates in Wales. Ofqual has told us that it is currently discussing with the Welsh regulator’s officials how to find an approach to summer 2013 awarding “that would provide us with the assurance that grade standards in WJEC English language in England and Wales will be comparable. To date we have not been able to gain assurance that common grade standards will be

159 Q306  
160 Q355  
161 Judgement, paragraph 149  
162 This is known as the “common centres” approach, which draws on data from centres with entries in the same subjects in the current and previous years (see GCSE English Language 2012 An investigation into the outcomes for candidates in Wales, Welsh Government, 10 September 2013, paragraph 30)  
163 WJEC set out details of its entry figures at Ev 78  
119. We accept Ofqual’s arguments for using the comparable outcomes approach, as discussed in Chapter 4. Glenys Stacey was explicit with us that she has prioritised standards in Ofqual’s work over the last year and we consider that the comparable outcomes approach is a key way in which Ofqual currently fulfils its statutory objective to maintain comparable standards year on year and between exam boards.

120. We can see, however, that the comparable outcomes approach presents difficulties with regard to candidates in Wales, and in particular when an exam board has substantial numbers of candidates sitting the same qualification in both England and Wales. Regulators are struggling to reach agreement on a way forward and relationships are evidently strained. The divergences in qualifications policy and design, coupled with disagreements over the use of comparable outcomes, mean that it is increasingly unclear how common standards across borders can be maintained. There is a risk, as one exam board chief executive has warned, that we will find ourselves in a situation where we have a qualification with the same name but with different standards. To some extent this has already happened, albeit by a small margin, in respect of the different standards set in England and Wales for the WJEC GCSE English Language in 2012. This is clearly unsatisfactory for all concerned, as it risks generating confusion among students, parents and the wider public, and, ultimately, devaluing the qualifications concerned.

Policy divergence between England, Wales and Northern Ireland

121. Events with GCSE English have taken place against a background of increasing policy divergence between England, and Wales and Northern Ireland. Welsh Ministers announced at the end of January 2013 that GCSEs and A levels will be retained in Wales and that “where necessary, we will strengthen and amend these”. Although Michael Gove announced in early February 2013 that GCSEs will be retained in England, differences are likely to emerge between GCSEs in England and Wales concerning modular/linear structures, the proportion of internal assessment and tiering. Northern Ireland is currently conducting its own review of GCSEs and A levels, which is due to report in June 2013. The Education Minister in Northern Ireland, John O’Dowd, has recently announced interim changes to A levels, retaining the current AS/A2 modular structure, but limiting re-sit opportunities. He has also asked the CCEA to consider a change to the weighting of AS/A2 from 50:50 to 40:60.

165 Ev 67
166 Q390
167 Q388
168 Q252 (Mark Dawe)
169 GCSE and A levels retained in Wales– Ministers accept recommendations of Qualifications Review, 29 January 2013, Welsh Government
170 O’Dowd announces decision on A level changes, Department of Education News Release, 20 May 2013
122. The Westminster Government’s announcements on A level reform in January 2013 prompted particular comment from Ministers in Wales and Northern Ireland. Jeff Cuthbert, Deputy Minister for Skills in Wales, described the Secretary of State for Education’s announcement as “unilateral” and one “which was not shared with us in advance, as we would have hoped under our concordat with the DfE”. Leighton Andrews told us that he tended to learn about announcements made by Mr Gove on GCSEs and A levels through the media. He emphasised to us that he would welcome a Ministerial conference to discuss qualifications and standards, as recommended in our 2012 exams report, saying:

what I do think would be helpful, in a situation where we have commonly branded and commonly owned qualifications, would be more ministerial dialogue when major changes are likely to take place [...] I certainly think a Ministerial conference of the kind requested by this Committee would be helpful.\textsuperscript{172}

123. We were pleased to note the recent meeting between the Secretary of State and Education Ministers John O’Dowd MLA and Leighton Andrews AM. Following this meeting, the Secretary of State wrote to Ministers Andrews and O’Dowd stating that increasing differences between GCSEs and A levels across the three countries were a “natural and legitimate consequence of devolution” and that he believed “the time is right for us to acknowledge that three-country regulation of GCSEs and A levels is no longer an objective towards which we should be working”.\textsuperscript{173} Mr Gove warned in the letter that there is a need to address the issue of titles, as it will no longer be appropriate for markedly different qualifications to share the same title. The letter appeared to come as something of a surprise to Ministers Andrews and O’Dowd. Mr Andrews has signalled that he intends to retain the GCSE and A level brands in Wales, commenting “we wish Mr Gove well with his plans to re-name these qualifications in England”.\textsuperscript{174}

124. All ministers have emphasized their concern that qualifications in the three countries should be robust, valued and enable young people to progress to further education and employment. As Mr Andrews told us, the portability of qualifications is very important.\textsuperscript{175} Employer organisations have said that they are used to working with different exam systems, such as already exists between England and Scotland. Neil Carberry, the CBI’s director of employment and skills, warned that Ministers should consider the practical implications of changes and that “businesses will need to be crystal clear about the differences—to eliminate any confusion if people with similar sounding results, from either side of the border, go for the same job”.\textsuperscript{176}

\textsuperscript{171} Wales and England divide on exams, BBC News, 29 January 2013
\textsuperscript{172} Q322 and Q324, see also Q290, Q291, Q296.
\textsuperscript{173} Letter from Michael Gove to Leighton Andrews AM and John O’Dowd MLA, 20 May 2013
\textsuperscript{174} UK shared exam system faces break up, BBC News, 21 May 2013
\textsuperscript{175} Q299
\textsuperscript{176} UK shared exam system faces break up, BBC News, 21 May 2013
125. Relations between Ministers in England and Wales are clearly under strain, as the era of three-country qualifications and regulation appears to be coming to an end. We believe that such an outcome would be regrettable and hope that even at this stage the joint ownership of GCSEs and A levels will continue. We urge Ministers to do everything possible to bring this about.

Implications for Ofqual

126. Ofqual’s position is that where qualifications differ significantly in terms of structure, assessment opportunities and content, it becomes “extremely difficult to make sure standards are consistent across them” [177]. It believes that a qualification with a common title should have a common standard [178]. As Mark Dawe of OCR put it to us, “either, there are different qualifications in each country with different standards set, or [...] there is one qualification with one common process for standards” [179]. Glenys Stacey told us that there comes a point at which the qualification is no longer the same and you can no longer set a common standard, saying “at that point, and we are fast reaching that point, you are then looking at how you differentiate those qualifications at awarding on the certificate and, ultimately, through titling” [180]. Current indications are that increasing policy divergence and differences in regulatory approaches will mean more differentiation between qualifications is likely. Already there is differentiation on exam certificates, as from summer 2013 exam certificates issued to WJEC GCSE English Language candidates will carry only the WJEC and Welsh Government logos.

127. In May 2012 the regulator and exam board in Northern Ireland, CCEA, announced that it will no longer be offering its GCSEs and A levels to candidates in England, a decision taken “as result of emerging policy differences between England and Northern Ireland.” [181] By contrast, as we found during our 2012 exams inquiry, WJEC has increased its market share in recent years in several GCSE and A level subjects, most notably GCSE English [182]. We understand from Glenys Stacey that Qualifications Wales, the body which will award and regulate qualifications in Wales in future, will wish to market its qualifications in England as well as in Wales [183]. Indeed Ms Stacey suggested to us that this is “fairly central to the financial modelling for the organisation” [184]. When qualifications are marketed and sat by students in England, they are regulated by Ofqual.

---

177 Ofqual comment on Secretary of State’s letter to Ministers Andrews and O’Dowd, 21 May 2013
178 Q391
179 Q252 (Mark Dawe)
180 Q391
181 Letter from CCEA to heads of all centres in England, 18 May 2012
182 First Report of Session 2012-13, The administration of examinations for 15-19 year olds in England, HC 141-I, paragraph 133
183 The Welsh Government has announced that regulatory arrangements in Wales will be changing, with regulation due to become independent of Government. A new body, Qualifications Wales, will be set up to regulate and award qualifications in Wales, similar to arrangements in Scotland and Northern Ireland.
184 Q394
128. The Secretary of State’s letter of 20 May 2013 to Ministers Andrews and O’Dowd suggests that in future separate, and rather different, GCSE and A level will run in future in Wales and Northern Ireland and that regulators will no longer be required to regulate jointly to ensure a common standard. If the Welsh provider wishes in future to offer GCSEs and A levels in England, it will need to offer GCSEs and A levels which are accredited and regulated by Ofqual, while offering separate qualifications for candidates in Wales. There will no longer be the need to agree an approach to standards setting with the Welsh regulator. Although this will in some respects simplify the work of Ofqual, it may take time to understand the full implications of separate qualification and regulatory systems across the three countries. It also does not solve the immediate difficulties for GCSEs taken before 2017.

129. The current jointly owned and regulated GCSEs will continue to run until summer 2016 in some subjects and until summer 2017 in others. We recommend that Ofqual does all it can to improve relations with the Welsh regulator and to agree a way forward on standards setting for the remaining lifetime of the current GCSEs.
6 The investigation process

130. Ofqual conducted an investigation into the problems with the 2012 GCSE English results, publishing an initial report on 31 August 2012, followed by a second, fuller report on 2 November 2012. Following the publication of both reports, questions were raised about the extent to which Ofqual was able to conduct an impartial investigation into what had happened, given its close involvement as regulator in the standards setting process.

131. Sir Mike Tomlinson, who led the independent inquiry into A level grading in 2002, commented that “it is always a problem if the body that investigates the issue might have played a part in creating the issue.”185 In an open letter to the Chair of the Education Committee, Brian Lightman, General Secretary of the ASCL, called for an independent investigation into the events surrounding the GCSE English results, saying that “only an independent review will be able to investigate these issues with the degree of expertise, depth and objectivity which will be required”.186

132. Exam board representatives expressed mixed views about the extent to which Ofqual was able to conduct an impartial investigation. Ziggy Liaquat of Edexcel told us that “Ofqual is obviously deeply involved in the setting of grade boundaries for English. It is an odd situation that they are investigating an issue that involved them.”187 He pointed out, however, that “if you look at the final outcomes of their investigation against the findings of the High Court, you have to say that they did a good job in identifying some of the issues”.188 Other exam board chief executives were more positive, stating that Ofqual was probably best placed to take an overview of what exam boards had been doing and it had the necessary assessment expertise to analyse the technical issues involved.189 Mark Dawe of OCR commented that:

they were set up to do this sort of job, and to some extent if they are not given that role, then why bother with the regulator [...]you needed someone with the expertise to understand what we were handing over to then be able to judge what had happened, and I do not think anyone outside the regulator would have been able to do that job appropriately.190

133. Both AQA and OCR representatives stated that the investigation process did not feel very comfortable, with AQA’s Andrew Hall telling us that “it did not feel that the investigation was carried out by someone who was integral in the process. It felt very independent and we were challenged a lot”.191

185 Notes on a similar scandal: the 2002 A level fiasco, Times Educational Supplement, 7 September 2012
186 Letter from Brian Lightman to Graham Stuart, 9 October 2012
187 Q226 (Ziggy Liaquat)
188 Ibid.
189 Q226 (Andrew Hall), Q277 (Gareth Pierce)
190 Q277 (Mark Dawe)
191 Q226 (Andrew Hall) and Q277 (Mark Dawe)
When we asked Ofqual in September 2012 whether it was effectively investigating itself, Amanda Spielman, Ofqual’s Chair told us that “the issues are mainly around what happens in exam boards and in schools, not around what has happened in Ofqual.” On the same day, however, an exchange of letters between Edexcel and Ofqual was leaked to the press concerning Ofqual’s intervention in Edexcel’s summer 2012 GCSE English award. Ofqual’s interim report had not covered its intervention in either Edexcel or WJEC’s awards, and the publication of this information raised suspicions about Ofqual’s role in grade boundary setting, leading to allegations of its fixing grade boundaries. Ofqual has acknowledged that it could have done more to explain its role in standards setting last summer in its interim report, although it does not believe that this would have prevented the judicial review. Glenys Stacey told us that:

we knew we had done that, and we knew it was entirely proper; we were trying to find out what had gone wrong, not what had gone right, but I can understand now that people may think we could have covered it. I think, with the benefit of hindsight we could have expanded and put a little bit in about our role and how we played it, but we did not really understand that to be at issue.

Mark Dawe, Chief Executive of OCR was critical of Ofqual’s interim report, telling us that “it did close down quite a few options” and “it did not really share the issues and the dilemmas that Ofqual were having to deal with”. Mr Dawe suggested to us that the regulator “had the role of getting everyone around the table and stopping this court case from happening.”

Responding to Ofqual’s November 2012 report, Malcolm Trobe, deputy general secretary of ASCL, stated that “this was an investigation carried out by a regulator into its own conduct and that of the awarding bodies. The chances of an impartial and accurate assessment were never great, which is why ASCL believes more strongly than ever that there must be an independent investigation into what happened.”

The Secretary of State for Education resisted calls for an independent inquiry, pointing out that Ofqual was an independent regulator, accountable to Parliament. He told us in September 2012 that “it would be quite wrong for me to appoint over that regulator an outside body” and that it was “for Parliament to ask the regulator, and all those who are regulated by her, the appropriate questions”.

In the absence of an independent inquiry, the alliance of schools, pupils, local councils and professional bodies made an application for judicial review. Ofqual told us that it is not

---

192 Q162
193 Q396
194 Q238
195 Q279 (Mark Dawe)
196 Ofqual report demonstrates that independent investigation is needed into GCSE English, ASCL press release, 2 November 2012
197 Oral evidence taken before the Education Committee on 12 September 2012, The Responsibilities of the Secretary of State, Q6
unusual for a regulator to be judicially reviewed. What was unusual in this case was that the challenge was brought not by those it regulates, but by others. Ofqual indicated that their legal costs for the judicial review, which have been awarded, will run to several hundred thousand pounds. Glenys Stacey suggested that costs for other defendants (AQA and Edexcel) may be higher.

139. Given the strength of feeling among schools, teachers and pupils about last year’s GCSE English results, we acknowledge that a judicial review was a likely outcome.

140. As exam board representatives noted, Ofqual now has the necessary expertise to investigate problems with the exam system and this forms a key part of its work as regulator. The question remains about how best to investigate decisions and actions taken by Ofqual itself. In this case, Ofqual was investigating some decisions which were taken before it was established in its current form. It was also, however, investigating issues in which it was closely involved as regulator, and in future, it will be increasingly likely that it will find itself investigating decisions it has taken in its present form. Judicial review is, of course, one option, although, as this case demonstrates, this can be a rather long and expensive remedy.

141. Ofqual is accountable to Parliament, predominantly through this Committee. We scrutinise the actions of both the regulator and the regulated, taking the advice of independent experts where appropriate. We also make any necessary recommendations for reform. In the exceptional event that a more wide-ranging and in-depth inquiry is required, the Government and House of Commons must ensure that the Committee is adequately resourced to enable it to investigate the technical processes and procedures in question. The Government should also commit to a presumption that any subsequent recommendations made by the Committee will be implemented.
142. Although the judicial review endorsed the actions of Ofqual and the exam boards, we agree with Ofqual’s assessment that there are some very sobering lessons to be learned from the events with GCSE English in 2012 and that the experience has engendered a great deal of mistrust. Confidence in the exam system has been shaken and a substantial amount of public money has been spent. Many teachers and pupils still feel that the action taken by Ofqual and the exam boards was unfair, even though it was found to be lawful.

143. Lord Justice Elias noted in his judgement that whichever way Ofqual chose to resolve the problems with GCSE English last year, there was going to be an element of unfairness. Once it became clear that the January cohort had been assessed more leniently, Ofqual was engaged in an exercise of “damage limitation.” The judge concluded that at some stage, a decision would have to be taken to depart from the January grade boundaries, which were pitched too low, in order to protect standards and the currency of the GCSE English qualification, and at this point, there would be winners and losers. The judge endorsed Ofqual’s decision to prioritise comparable standards between 2011 and 2012. The relatively small number of students from the January exams received slightly generous results, while the majority of students, the June cohort, received results which were line with previous years.

144. The 2013 results will be crucial in helping to restore confidence in the exam system and in the assessment of GCSE English in particular. Overall, we believe that measures put in place since last summer, along with further proposed changes, should help to strengthen assessment and moderation in the current English GCSEs and make them more resistant to pressures from the accountability system. The increased variability in results between schools came about because it was a new and very flexible qualification. We note Ofqual’s warning that GCSE English is still a relatively new qualification and that some variation school by school in this year’s results is to be expected. All will be watching the 2013 results very closely.

145. Lord Justice Elias attributed the source of the unfairness with GCSE English to the structure of the qualification itself. This underlines the vital importance of getting decisions right during the qualifications design phase. We are reassured that changes since 2007 and the establishment of an independent regulator with greater statutory powers and increased assessment expertise should facilitate this.

146. We welcome Ofqual’s indication that in future it will take into account the context in which qualifications will operate during qualifications development. Careful attention needs to be paid on a subject by subject basis to the balance of internal and external assessment, and to setting appropriate marking tolerances. Ofqual must ensure that sufficient attention is paid to any warning voices, particularly from subject and assessment specialists, during qualifications design and accreditation. The lessons to be learned from GCSE English for qualifications design are especially important as the exam system is facing multiple reforms, with fundamental changes to GCSEs due to take place at the same
time as significant reforms to A levels. We need to be prepared for increased variability in results, as this is more likely at moments of substantial change.

147. The Secretary of State has shown that he is prepared to act on the advice of the Chief Regulator regarding changes to the exam system and qualifications. We recommend that the Government continues to have serious regard to Ofqual’s advice on qualifications design in the forthcoming GCSE and A level reforms. We also recommend that Ofqual ensures it has systems and procedures in place for qualifications design which are sufficiently resourced and appropriately robust to enable it to meet the significant challenges ahead and to secure high quality qualifications for young people.

148. Young people and their future life chances are at the heart of this matter. It is important not to underestimate or forget the impact on those who took the GCSE English exam in 2012 and those who are taking it this year. It is essential to stress that, despite the problems highlighted in this report, students, both in 2012 and 2013, will have gained a GCSE English qualification that is a meaningful and valid reflection of their achievements, and that should enable them to progress to further education and employment.
Conclusions and recommendations

Modular/linear structure

1. Under the previous Government, GCSEs changed from mostly linear to modular, which, combined with other changes, brought turbulence to the system and contributed to the problems experienced with GCSE English in 2012. We recommend that, when considering their reforms of GCSEs and A levels, current Ministers think carefully about the cumulative impact and risks of change. (Paragraph 30)

Balance of internal and external assessment

2. We welcome Ofqual’s proposals to increase the weighting of external assessment in the current GCSE English and GCSE English Language qualifications from summer 2014, as we agree that this will help to make the qualifications more robust and more resistant to pressure from the school accountability system. We note the concerns expressed by school leaders about the timescales, but believe that benefits of the proposed action outweigh the downsides, particularly given the assurances from Ofqual that it will take steps to minimise any advantage or disadvantage to students caused by the change. (Paragraph 38)

3. It is clearly better to decide on a subject-by-subject basis what an appropriate proportion of internal assessment should be, rather than imposing central requirements or models to which all subjects must conform. We welcome Ofqual’s indication that in future these decisions will also take into account the context in which the qualifications will operate. This is particularly important for GCSE English, given the current proposals for GCSEs in English and mathematics to remain part of a headline threshold accountability measure as part of the new secondary school accountability measures. We look forward to examining Ofqual’s proposals for internal assessment in revised GCSE qualifications in due course. (Paragraph 40)

Moderation procedures

4. We welcome the steps taken by exam boards to improve their moderator feedback to schools and to make clearer the distinction between administration and standards/marking issues. We recommend that it is made clear to schools and colleges in moderator feedback if they have been marking generously or severely but within tolerance, and that Ofqual monitors this aspect of exam board communication with schools and colleges more closely in future, to ensure that teacher assessments are fair and accurate. (Paragraph 44)

5. We recommend that Ofqual and the exam boards consider whether changes to moderation systems and processes are needed to ensure that it is easier to analyse and track patterns in schools’ marking of internally assessed work. (Paragraph 47)

6. We are surprised that, given all the assessment expertise residing in exam boards and now in Ofqual, no-one questioned whether a standard tolerance of 6 per cent was
appropriate for a high stakes qualification with such a high proportion of controlled assessment. (Paragraph 48)

7. We recommend that Ofqual and the exam boards consider on a subject by subject basis what an appropriate tolerance might be for new GCSEs and A levels when deciding upon the proportion of internal assessment for each qualification, and that these decisions be informed by the research commissioned by Ofqual. (Paragraph 50)

8. Moderation, as AQA has stated, relies on the professionalism of the teacher community. Ofqual’s position on over-marking is an uncomfortable one for teachers, as it calls into question the integrity of some of the profession. We accept Ofqual’s findings about over-marking. However, we can see that its position is not helped by its reliance on a sample from one exam board, by contrasting views among exam board chief executives and by moderator feedback to schools and colleges which has not always been sufficiently clear about marking and standards. Furthermore, we recognise that Ofqual’s action to address over-marking in some schools has led to the unavoidable but highly unsatisfactory situation that students in other schools, whose work was marked accurately by their teachers, may have been penalised. Exam boards and Ofqual must make every effort to ensure that this situation is not repeated in summer 2013. (Paragraph 55)

**Speaking and listening**

9. The problems experienced with GCSE English in 2012 highlighted serious weaknesses in the moderation of speaking and listening and the consequences for grade awarding. While we agree that speaking and listening are important skills, the current assessment arrangements are not robust enough to ensure that assessment is a reliable, fair and accurate reflection of students’ performance. This risks devaluing the assessment of those skills and also generates further problems in securing standards across the qualification as a whole. On balance, we welcome the action proposed by Ofqual to address the weaknesses in the assessment of speaking and listening (Paragraph 60)

**Pressures from the school accountability system**

10. We welcome Ofqual’s indication that it will take into account the context in which qualifications operate when regulating and planning qualifications reform. This is especially important for GCSE English, given that it looks likely to remain part of a headline threshold measure, as outlined in the Government’s proposals for secondary school accountability. We recommend that Ofqual indicates publicly and clearly when and how accountability measures are a factor in its decision about how a qualification is designed. (Paragraph 66)

**The impact of organisational change on qualifications development**

11. Contrasting evidence from the exam boards and Ofqual would suggest some confusion about roles and responsibilities with regard to qualifications development while the interim Ofqual was still part of QCA. Encouraging the interim Ofqual to brand itself as a different organisation when it remained a sub-committee of the
QCA may have led to confusion over its role, responsibilities and powers at the time. (Paragraph 75)

**Warning voices**

12. *It is clear that warning voices regarding potential problems were raised but not acted upon during the development and accreditation phases of the current English GCSEs. While innovation and change is healthy and essential in any examination system, one of the crucial lessons that must be learned from this episode is that Ofqual and Ministers should listen when concerns are raised, especially when they come from specialists in the field. Balancing innovation and change with sound, specialist advice is the hallmark of a robust and high quality examination system* (Paragraph 79)

**Forthcoming GCSE and A level reforms**

13. *The Government is embarking upon the most significant and wide-ranging reforms to GCSEs and A levels since Ofqual was established as an independent regulator and has set a challenging timetable. We recommend that the DfE and Ofqual set out in detail their respective roles and responsibilities in qualifications development, particularly regarding how subject content will be developed, and publish this information before their respective consultations on the proposed GCSE reforms have ended.* (Paragraph 87)

14. We welcome signs that Ofqual is itself prepared to be a warning voice on qualifications reform to Ministers, if it judges this to be necessary, and we were pleased to see that Ministers acted upon Ofqual’s advice in connection with qualifications and market reform earlier in 2013. (Paragraph 88)

15. *Ofqual has taken steps to increase its assessment expertise, as recommended in our 2012 exams report. Ofqual must use this expertise to heed warning voices and take on board technical arguments in qualifications design for the forthcoming GCSE and A level reforms. It should be prepared to put a robust case to Government, should those arguments be contrary to Government policy steers.* (Paragraph 89)

16. *Ofqual’s role in qualification design is likely to come under significant pressure in coming months as the proposed timetable for GCSE reform coincides with significant changes to A levels. Ofqual should ensure that it has in place robust systems and adequate resourcing to undertake this role effectively, and that it gives explicit advice to Ministers about the risks involved in reforming GCSEs and A levels at the same time. It also needs to raise public awareness of the likelihood of increased variability in results during times of significant changes to qualifications.* (Paragraph 90)

**Exam board communication with schools and colleges**

17. *Exam boards must take very seriously the need to communicate better with schools and colleges, in order to improve trust and confidence in the exam system. We welcome signs that this communication is improving, but, as exam boards have acknowledged, this is a significant and ongoing challenge, and there is more to be done. We also welcome Ofqual’s greater scrutiny of this area and recommend that Ofqual continues*
to be more vigilant in its monitoring of communication between exam boards and schools, as this area is closely linked to public confidence. We also welcome Ofqual’s plans for improving its own communication with schools and colleges about marking and grading in the summer 2013 exams, in order to help restore confidence in Ofqual’s regulation of exams, particularly among teachers and head teachers. (Paragraph 97)

**Grade awarding and the comparable outcomes approach**

18. As we noted in our 2012 exams report, there are significant challenges for Ofqual and the exam boards in explaining the technical difficulties surrounding grade awarding. The problems experienced with GCSE English in 2012 have brought many of the issues to the foreground, with the judicial review scrutinising the grade boundary setting process in considerable detail. There are, however, lingering misconceptions and concerns about unfairness and, as Ofqual has acknowledged, the experience engendered a great deal of mistrust. Work is needed to improve understanding of and restore confidence in the system, particularly with regard to grade awarding and Ofqual’s role in maintaining standards. (Paragraph 109)

19. Ofqual must continue to make greater efforts to explain the complexities of awarding and standards setting, and its role in this, to schools and colleges, in order to improve confidence in the comparable outcomes approach and in Ofqual’s work as regulator. Ofqual should also keep the comparable outcomes approach under review and be prepared to adapt it in the light of experience and/or expert advice. (Paragraph 110)

**The future of three-country qualifications and regulation**

20. Relations between Ministers in England and Wales are clearly under strain, as the era of three-country qualifications and regulation appears to be coming to an end. We believe that such an outcome would be regrettable and hope that even at this stage the joint ownership of GCSEs and A levels will continue. We urge Ministers to do everything possible to bring this about. (Paragraph 125)

21. The current jointly owned and regulated GCSEs will continue to run until summer 2016 in some subjects and until summer 2017 in others. We recommend that Ofqual does all it can to improve relations with the Welsh regulator and to agree a way forward on standards setting for the remaining lifetime of the current GCSEs. (Paragraph 129)

22. Given the strength of feeling among schools, teachers and pupils about last year’s GCSE English results, we acknowledge that a judicial review was a likely outcome. (Paragraph 139)

**The investigation process**

23. As exam board representatives noted, Ofqual now has the necessary expertise to investigate problems with the exam system and this forms a key part of its work as regulator. The question remains about how best to investigate decisions and actions taken by Ofqual itself. In this case, Ofqual was investigating some decisions which were taken before it was established in its current form. It was also, however,
investigating issues in which it was closely involved as regulator, and in future, it will be increasingly likely that it will find itself investigating decisions it has taken in its present form. Judicial review is, of course, one option, although, as this case demonstrates, this can be a rather long and expensive remedy. (Paragraph 140)

24. Ofqual is accountable to Parliament, predominantly through this Committee. We scrutinise the actions of both the regulator and the regulated, taking the advice of independent experts where appropriate. We also make any necessary recommendations for reform. In the exceptional event that a more wide-ranging and in-depth inquiry is required, the Government and House of Commons must ensure that the Committee is adequately resourced to enable it to investigate the technical processes and procedures in question. The Government should also commit to a presumption that any subsequent recommendations made by the Committee will be implemented. (Paragraph 141)

Conclusion

25. The Secretary of State has shown that he is prepared to act on the advice of the Chief Regulator regarding changes to the exam system and qualifications. We recommend that the Government continues to have serious regard to Ofqual’s advice on qualifications design in the forthcoming GCSE and A level reforms. We also recommend that Ofqual ensures it has systems and procedures in place for qualifications design which are sufficiently resourced and appropriately robust to enable it to meet the significant challenges ahead and to secure high quality qualifications for young people. (Paragraph 147)

26. Young people and their future life chances are at the heart of this matter. It is important not to underestimate or forget the impact on those who took the GCSE English exam in 2012 and those who are taking it this year. It is essential to stress that, despite the problems highlighted in this report, students, both in 2012 and 2013, will have gained a GCSE English qualification that is a meaningful and valid reflection of their achievements, and that should enable them to progress to further education and employment. (Paragraph 148)
Appendix

Memorandum submitted by the Department for Education

Office of Qualification and Examinations Registration (Ofqual)

The Office of Qualifications and Examinations Regulation (Ofqual) is a non-ministerial government department that is accountable to Parliament and the Northern Ireland Assembly. It is the regulator of qualifications (other than degrees), examinations and assessments in England and the regulator of vocational qualifications in Northern Ireland.

It was established under the Apprenticeships, Skills, Children and Learning Act 2009 and came into being in April 2010.

It is responsible for securing standards of and efficiency in regulated qualifications, and promoting confidence in qualifications. It recognises awarding organisations and sets requirements they must meet, checks qualifications, monitors delivery and can take action when things go wrong. It has to have regard to the reasonable requirements of, among others, learners, higher education institutions and employers. Ofqual also has a role in relation to National Curriculum assessments.

Working with Government

Whilst Ofqual is an independent regulator, reporting directly to Parliament and the Northern Ireland Assembly, the qualifications it regulates, particularly those taken by students under 19, sit within the overall policy framework set by Ministers, and in particular Ministers have responsibility for the curriculum arrangements which many regulated qualifications assess. It is not for Ofqual to decide which qualifications should be available—that is for Government, or for the market. Ofqual’s role is to make sure that regulated qualifications are assessed properly and that standards are maintained, for example.

Therefore, Ofqual has to “have regard to” Government policy, where directed to do so (for example, the recent letter on A level policy). It also has a duty to provide on request information and advice to the Secretary of State or the Northern Ireland Department of Employment and Learning. Ofqual aims to be a trusted, independent and evidence-based source of advice to Ministers on qualifications policy issues.

In particular, where Ministers wish to use qualifications reform as a policy lever, there is a joint responsibility for the reform programme. Ministers have responsibility for the policy framework including funding and accountability requirements, the curriculum arrangements, and any implementation issues in schools and colleges (for example, workforce implications). Ofqual has responsibility for setting assessment arrangements,
setting standards and regulating the awarding organisations as they develop and deliver the qualifications.

Scope of responsibility

Ofqual regulates qualifications and assessments that are offered in England and vocational qualifications that are offered in Northern Ireland.

Ofqual recently published some principles setting out its approach to qualifications which may be subject to more than one set of regulatory requirements (http://www.ofqual.gov.uk/files/2013-03-22-ofquals-principles-for-regulating-qualifications.pdf).
Formal Minutes

Wednesday 5 June 2013

Members present:

Mr Graham Stuart, in the Chair

Alex Cunningham       Ian Mearns
Bill Esterson         Chris Skidmore
Pat Glass             David Ward
Charlotte Leslie      Craig Whittaker
Siobhain McDonagh

Draft Report (2012 GCSE English results), proposed by the Chair, brought up and read.

Ordered, That the draft Report be read a second time, paragraph by paragraph.

Paragraphs 1 to 148 read and agreed to.

Summary agreed to.

A Paper was appended to the Report.

Resolved, That the Report be the First Report of the Committee to the House.

Ordered, That the Chair make the Report to the House.

Ordered, That embargoed copies of the Report be made available, in accordance with the provisions of Standing Order No. 134.

Written evidence was ordered to be reported to the House for printing with the Report (in addition to that ordered to be reported for publishing on 24 October 2012).

[Adjourned till Tuesday 11 June at 9.15 am]
Witnesses

Tuesday 11 September 2012

Brian Lightman, General Secretary, Association of School and College Leaders (ASCL), Mike Griffiths, Headmaster of Northampton School for Boys and ASCL President, Russell Hobby, General Secretary, National Association of Head Teachers (NAHT), Mrs Kenny Frederick, Principal, George Green’s School, Isle of Dogs and member of the NAHT executive

Glenys Stacey, Chief Executive and Chief Regulator of Qualifications and Examinations, Amanda Spielman, Chair, and Cath Jadhav, Acting Director of Standards, Ofqual

Tuesday 12 March 2013

Andrew Hall, Chief Executive Officer, AQA, and Ziggy Liaquat, Managing Director, Edexcel

Mark Dawe, Chief Executive, OCR, and Gareth Pierce, Chief Executive, WJEC

Leighton Andrews, AM, Minister for Education and Skills, for Wales, and Chris Tweedale, Director of Schools and Young People Group, Welsh Government

Tuesday 26 March 2013

Glenys Stacey, Chief Executive and Chief Regulator of Qualifications and Examinations, Ofqual, and Dennis Opposs, Director of Standards, Ofqual

List of printed written evidence

1 Ofqual Ev 60
2 Leighton Andrews, AM, Minister for Education and Skills for Wales Ev 71
3 AQA Ev 71
4 Pearson Ev 74
5 WJEC Ev 77
# List of Reports from the Committee during the current Parliament

The reference number of the Government’s response to each Report is printed in brackets after the HC printing number.

## Session 2010-12

<table>
<thead>
<tr>
<th>Report Type</th>
<th>Title</th>
<th>Reference Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Special Report</td>
<td>Young people not in education, employment or training: Government Response to the Children, Schools and Families Committee’s Eighth Report of Session 2009-10</td>
<td>HC 416</td>
</tr>
<tr>
<td>Second Special Report</td>
<td>The Early Years Single Funding Formula: Government Response to the Seventh Report from the Children, Schools and Families Committee, Session 2009-10</td>
<td>HC 524</td>
</tr>
<tr>
<td>Third Special Report</td>
<td>Transforming Education Outside the Classroom: Responses from the Government and Ofsted to the Sixth Report of the Children, Schools and Families Committee, Session 2009-10</td>
<td>HC 525</td>
</tr>
<tr>
<td>Fourth Special Report</td>
<td>Sure Start Children’s Centres: Government Response to the Fifth Report from the Children, Schools and Families Committee, Session 2009-10</td>
<td>HC 768</td>
</tr>
<tr>
<td>First Report</td>
<td>Behaviour and Discipline in Schools</td>
<td>HC 516-I and -II</td>
</tr>
<tr>
<td>Second Report</td>
<td>The role and performance of Ofsted</td>
<td>HC 570-I and II</td>
</tr>
<tr>
<td>Third Report</td>
<td>Services for young people</td>
<td>HC 744-I and –II</td>
</tr>
<tr>
<td>Fourth Report</td>
<td>Participation by 16-19 year olds in education and training</td>
<td>HC 850-I and –II</td>
</tr>
<tr>
<td>Fifth Report</td>
<td>The English Baccalaureate</td>
<td>HC 851</td>
</tr>
<tr>
<td>Sixth Report</td>
<td>Services for young people: Government Response to the Committee’s Third Report of Session 2010-12</td>
<td>HC 1501</td>
</tr>
<tr>
<td>Seventh Report</td>
<td>Appointment of HM Chief Inspector, Ofsted</td>
<td>HC 1607-I</td>
</tr>
<tr>
<td>Eighth Report</td>
<td>Chief Regulator of Qualifications and Examinations</td>
<td>HC 1764-I and -II</td>
</tr>
<tr>
<td>Ninth Report</td>
<td>Great teachers: attracting, training and retaining the best</td>
<td>HC 1515-I</td>
</tr>
</tbody>
</table>
### Session 2012–13

| First Report | The administration of examinations for 15–19 year olds in England | HC 141-I (HC 679) |
| Second Report | Appointment of Chair, Social Mobility and Child Poverty Commission | HC 461-I |
| Third Report | Governance and leadership of the Department for Education | HC 700 (HC 919) |
| Fourth Report | Children first: the child protection system in England | HC 137-I (HC 993) |
| Fifth Report | Support for Home Education | HC 559-I (HC 1013) |
| Sixth Report | Pre-legislative scrutiny: Special Educational Needs | HC 631-I |
| Seventh Report | Careers guidance for young people: The impact of the new duty on schools | HC 632-I (HC 1078) |
| Eighth Report | From GCSEs to EBCs: the Government’s proposals for reform | HC 808-I (HC 1116) |

### Session 2013–14

| First Report | 2012 GCSE English results | HC 204 |
Oral evidence

Taken before the Education Committee

on Tuesday 11 September 2012

Members present:
Mr Graham Stuart (Chair)
Neil Carmichael
Alec Cunningham
Bill Esterson
Pat Glass
Damian Hinds
Charlotte Leslie
Siobhain McDonagh
Ian Mearns
Mr David Ward
Craig Whittaker

Examination of Witnesses

Witnesses: Brian Lightman, General Secretary, Association of School and College Leaders (ASCL), Mike Griffiths, Headmaster, Northampton School for Boys and ASCL President, Russell Hobby, General Secretary, National Association of Head Teachers (NAHT), and Kenny Frederick, Principal, George Green’s School, Isle of Dogs, and member of the NAHT executive, gave evidence.

Q1 Chair: Good morning and welcome to this meeting of the Education Select Committee. The Education Select Committee exists to scrutinise education policy and hold those in power to account. Central to that task is ensuring the well-being and the best possible outcomes for all those involved in the education system, not least children, whose futures depend on the performance of the education system and the way they are treated within it. I am delighted that you are able to join us today. Can you, as briefly as you can in the time-constrained situation we are in, set out what you think the problems are with what has happened with GCSE English grade boundaries in 2012?

Brian Lightman: Thank you very much, and thank you very much for the invitation. What I would like to say by way of introduction is this is a completely separate issue from any wider debate that is going on about qualifications reform or anything else. This is about a very specific issue to do with GCSE English. We believe that the implementation of that new purpose, and that has led to a massive variability in outcomes between different schools. We would strongly disagree with the claim in the Ofqual report that the standards were comparable between years. That is simply not the evidence that we have seen, and I am sure we will all be able to give lots of examples of that from many, many different schools and the results of many, many thousands of young people who have been affected at all grade boundaries.

What I want to emphasise here is that this variation between the schools has affected all kinds of schools: maintained schools, academies, HMC schools—Headmasters’ Conference—Girls School Association schools, so leading independent schools as well are telling us that it has affected them. It is not only at the C/D borderline but at other grade borderlines as well. I am sure we can go into more detail in a moment, but those problems are to do with grades coming through that are in no way in line with what one might have reasonably expected young people to be getting, having taught them throughout the time of their course. In particular, you will be aware that the grade boundaries were changed during the course of the year and so, depending on when students had put in their work to be banked, the grade would have been different for exactly the same work. So we are seeing all sorts of variability like that. We have more detail that we can go into.

We are seeing major flaws in the way the process of that examination was managed over the two years of its running, so that, for example, early work that was produced as controlled assessment was not moderated in a way that would have highlighted any issues of leniency, harsh marking or anything else. That should have been identified early, and that led to a major problem at the end, which meant that adjustments were made at the end that had a particularly harsh effect on those pupils who took the exam at the end of the course. I have summarised a vast amount of detail there, but I hope that helps by way of introduction.

Q2 Chair: Of the four key points you made there, I think one is accepted by Ofqual; the Chief Regulator said that it was easier in January and they got lucky. So there is no dispute that there was a difference there. You have said it is not comparable. The fundamental statutory duty of the regulator, after the Education Act 2011, is to ensure comparability, including over time, and you said that you believe, from the data you have seen, that it is not comparable, and that is their central duty. They say that is why so much of this has followed—because they have done that. Can you tell us what evidence you have to back that up?

Brian Lightman: We have had a look at a lot of figures here, and there is a lot of data underlying what we have looked at. I should emphasise that the awarding bodies and Ofqual have been very helpful and forthcoming in providing a lot of that information to us. When you look at it, if you take this approach of comparable outcomes and you think of what the definition of “fairness” is in terms of the Ofqual report saying that standards are comparable between the
years, then what that definition says really is that an appropriate percentage would get a certain grade across the whole cohort. The problem is that, if the marking was wrong early on and students therefore were marked too high and too many C grades were awarded, the only way of addressing that is at the end of the course. So the people who did their work at the end of the course would therefore be marked particularly harshly.

Q3 Chair: You are accepting, are you, that broadly the June results were right except that they were skewed by over-generous application of—
Brian Lightman: No, no, we are not. We are saying that the June marking was particularly harsh in order to compensate for what was apparently over-generous early in the year. The result is that, if you have these different standards being applied during the course of the year, those standards cannot be comparable. If you take just some of the statistics in the foundation tier units, in June 2011, 26.7% got a C grade.

Q4 Chair: This is the first smattering of people taking—
Brian Lightman: The year 10s who took it. In January 2012, it was 37%, and then in June 2012, 10.2%. These are enormous differences and there is no evidence that those papers had any difference in the level of challenge in those examinations between those times. You have these massive differences and effects on some pupils within the year. Therefore, it cannot be that the standard was the same across the whole two years.

Q5 Chair: That is accepted: Ofqual has accepted that the same standard was not applied within year. They claim because it is a new qualification it takes time to bed down; they had limited data; they did the best they could. I think they specifically say in a memo to us that the exam boards could not have known that they were being too generous. That appears to be accepted. Just trying to drill down, your point is that, because they got it wrong there and because they look at the whole cohort together to create a comparable line, it has skewed the June results downwards.
Brian Lightman: Yes.

Q6 Chair: Perhaps Russell wants to come in, but could you explain to what extent that has happened, and as the percentage of people who took the June 2011 or January 2012 English GCSE modules was pretty tiny, is that really a significant skewing? I know that is an entirely different topic to whether or not the expectations were mismanaged, but how skewed was it?
Russell Hobby: I think it goes deeper than just the balance between January and June. If you look at the correspondence that has been released today, you can see quite a strong argument between the exam board and Ofqual on how to use the Key Stage 2 data to predict what attainment should be. There was a genuine professional disagreement about those sorts of things, so already they are not able to use the concept of comparable outcomes to predict this, nor did they have the statistical base early on in the year to apply that notion. So it looks like the comparable outcome methodology failed due to not being able to use predictive data about the ability of the cohort and not being able to extrapolate from the initial size of the sample that they had. Given that comparable outcomes is a methodology that we are increasingly heading towards, as well as looking at how this has adversely affected the students this year, it is extremely dangerous for future years, I think, that we are going to be heading in that direction.

Q7 Chair: Just on the level of the skewing, because Brian made that point—and I know it is just one of many points—again could you explain to what extent the huge numbers of people who took the June 2012 exam were affected? How many people would have been affected by the skewing from a relatively small number of people—I forget; it is a single-figure percentage—who took the GCSE prior to June?
Brian Lightman: This is where I think there are a lot of misunderstandings. There are a lot of figures floating around here, and we have had to delve down very deep into them. If you take the AQA foundation tier—and I want to emphasise that this is not just an AQA problem but they have the largest entry—AQA had 54,000 students taking that foundation tier in June 2011, 31,000 in January 2012 and 140,000 in June 2012. These are not insignificant figures and they do have, obviously, a very significant skewing effect, and I do not think we can dismiss that at all. I know that we are moving towards linear qualifications, but certainly with this qualification there was an option to complete it in year 10 and also in January of year 11.

Q8 Damian Hinds: Just so the point does not pass, can we be 100% clear here? I think you might have been saying something new this morning. If the target level index is 100—that is what it “should” be, based on what has happened in prior years—and if in January people were passing at a rate of 110, are you saying that in June they were passing at a rate of (a) 100 or (b) something like 95 to make the weighted average 100? Because I think that would be a new point and quite a serious one.
Brian Lightman: I think you have hit the nail on the head there: if you are talking about comparable outcomes over the whole cohort as opposed to over those individual students, that is exactly what needs to happen.

Q9 Damian Hinds: But are you saying that is what did happen? That is the question, Brian.
Brian Lightman: Yes, we think that is what did happen, and I think the evidence that has just been mentioned here backs that up in terms of some of the correspondence between Ofqual and awarding bodies.

Q10 Damian Hinds: How? Not in what we have seen.
Brian Lightman: Because if they were outside tolerances, they had to then move the grade boundaries in order to compensate for what had happened in the rest of the year.

Q11 Damian Hinds: I am not sure that follows at all.
Russell Hobby: I do not know whether we do have evidence on that basis with the data. The very complexity of the data itself is one of the things that we are struggling with in here, to understand it. What I think we can say is that the students in June were unfairly treated, because they are competing with the cohort that took their exams in January, even if the results that they got would have been justified by the grading methodology.

Q12 Damian Hinds: Sure. I think that point is accepted and backed. We are now asking a different question: were the pupils taking the exam in June particularly harshly treated, even compared with prior years, because of this need to make up a weighted average across the year cohort?

Brian Lightman: Yes.

Q13 Damian Hinds: Russell does not seem to be saying that is the case. Brian, you do seem to be saying that is the case.

Russell Hobby: No. I do not know.

Brian Lightman: Russell is saying he does not know, but we have been into this and that is the 10.2% statistic that I quoted to you a few minutes ago. How can it be that in January 37% were getting a C and then in June 10.2%? That is where it has been adjusted in order to give the comparable outcome over the whole cohort.

Q14 Chair: That is a question for us to ask Ofqual then, because my understanding—and I wanted to explore that nonetheless—was they had taken the big June cohort, applied a comparable-outcomes expectation and triangulated data on that, and had found that the grade boundaries were wrong, just looking at that group as one. They had then looked back and said, “Oh, so we got the grade boundaries wrong before,” rather than there being a skewing from the prior results. Is that something we need to try to hammer out with the regulator this morning?

Brian Lightman: I think it certainly is, and I think it will come back to the issue of 60% controlled assessment, where the grades had already been announced and been stated and moderated at that early stage. Therefore, you have built up this imbalance, as it were, that needs to be balanced out at the end.

Mike Griffiths: I am here as one representing literally hundreds of head teachers who, like me, have seen their students have their hopes and aspirations really shattered because of this. As a head teacher, my belief is that Ofqual has simply failed to maintain standards; we keep hearing this phrase “maintaining standards”. I am not a statistician. I am not a statistician and I just would like for members of the Committee to be absolutely clear on how it affects a school like mine, where our student cohort was comparable this year compared with other years.

For the past five years, the percentage getting Cs and above in my school—it is only one school, but it is typical of literally hundreds—has ranged between 88.6% and 91.6%, a variation of only 3% over five years. This year, it fell 17% to 74.6%, despite being predicted as 86%. Our English literature results, which all of the students also take, similarly have ranged in those five years from 83% to 91%. They declined in that year, but only in line with expectations, by 1%. So my team of English teachers, as many others across the country, are distraught, devastated and confused as to why this decline has happened. It is the same team of teachers that has been teaching for several years and the students feel the same. I have lots of evidence to support, at an individual school level, the fact that Ofqual has simply failed to maintain the standard. I have 17% of students—that equates in just one school to 36 boys—who have not got a C grade who, were previous years’ standards applied, would have got a C grade or above. It also applies through the grade ranges, and I have evidence, which I can leave with you, of boys who have got nothing but As and A*s across the board, but Bs in English language.

Q15 Chair: Is the variability, which is again—a bit like the difference between January and June—not disputed by Ofqual, the increased variability that leads to schools getting results, after years of consistently predicting what was going to happen, and finding that there is a big difference? The increased variability is accepted; it has happened this year. Is that necessarily a failure of the regulator or could that be an inevitable consequence of the introduction of these new syllabuses with the new modular form and with increased percentage of controlled assessment? Could it be the architecture of the exams that made everybody’s job—yours and the regulator’s—impossible rather than necessarily a failure of the regulator?

Mike Griffiths: I think that my department was as well prepared as it possibly could be. They are highly professional people, and I can, as I say, leave the evidence as to exactly what they did. But as a very final point on that thing about it being due to a change, the board’s very final comment to my school was, “The centre is to be congratulated on its approach to the new specification and the accuracy of the marks awarded to students.”

Russell Hobby: It is a failure of the regulator to have the information many years ago about what impact not necessarily a modular exam system but simply a new exam system would have on this, and not to prepare schools and exam boards to know that the change in grade boundaries this year would not be in line with that in other years but that there would be a dramatic shift.

Q16 Chair: They are quite clear. They say in evidence to us that the expectations of schools were raised by the January grade boundaries, but that “schools are told”—are told—not to rely on those grade boundaries and are told, particularly when there is a new examination, not to rely on it.

Kenny Frederick: They are told it is usually between two and three points, not 10 points. You have staff and youngsters working together, and it is not quite as clear as between January and June, because in a lot of schools they might have done the exam in January, but the controlled assessment, which is 60% of the mark, they would have cashed in in June. They have
done that. So you have the same cohort judged under two different exam grades.

Whose responsibility is it? The exam boards set the exam; Ofqual approves it. They set the assessment criteria; Ofqual approves it. We go to training; my staff know what they have to do. It seems to me that there is no understanding of what goes on in school, because we have to set targets based on Key Stage 2 results. If they come in and they are level four, they have to go out with at least a C, and that is what we work through. We measure them, we test them every six weeks, we look at their strengths, their weaknesses; we work with them. That is gone through. All that work is done. We have double checked everything.

We have had our controlled assessments marked externally by people who do not know the youngsters and people who have been trained by the exam board. The exam board feedback was fantastic. Surely if you do your driving test, is there a quota on doing it? I did mine three times. You might do the theory test, you might pass that and you do the other bit. You get there until you pass it. These are young people at the heart of this. Their whole future is destroyed. They cannot go and do the courses. Some of them have gone into Level 2 courses when they should have gone into Level 3 courses.

Q17 Chair: Did you mean to say their whole future is destroyed? I do think a certain moderation of language might be necessary.

Kenny Frederick: Well, it depends on the children, doesn’t it? I know a number of youngsters who are dropping out and will not go forward. It depends on the youngsters and on the backing they get from their home. So for a lot of these youngsters their future is destroyed, and I do not think that is extreme at all and, talking to head teachers across the country, they would certainly back that up.

Brian Lightman: Could I just say something about that C grade, because it is such an important qualification? We have hundreds and hundreds of examples of students who have had difficulty accessing A level courses and so on. Colleges and sixth forms are trying to be as helpful as they can, but also there is the issue of selective universities, and if you have not got an A in certain GCSEs, you will not get into a selective university. So some of our most able students are at risk, because of this, of losing out on that basis.

Q18 Chair: You are gathering information on that.

Brian Lightman: We are.

Q19 Chair: One of the big issues is to identify any children who have been damaged and estimate how many and to what extent precisely in order to have a clear idea.

Brian Lightman: Yes.

Q20 Neil Carmichael: Can I ask a fairly general question, just to guide the Committee? Are we really talking about a fundamental problem with the architecture of comparable outcomes or are we talking about a failure to appreciate and gather the right kind of data and apply it to the system?

Russell Hobby: We are talking about both, but the concept of comparable outcomes seems to me to be utterly incoherent and unable to be applied in the context of a changing exam system as well. I think they did not gather the data they should have done and did not warn people. If you look at the criteria for comparable outcomes, they need to be able to predict performance based on Key Stage 2 data. They disagreed on how that would apply this year. The only way that you can judge whether teaching standards have risen is using exam data, which is a completely circular argument for the comparable outcomes. You have every school in the country racing to improve its results and you have an exam system that will automatically re-factor the results so every school gets the same at the end of it. It just does not work, full stop.

Q21 Chair: If you have a new qualification and it is made modular, and 60% of the marks are given out by the very teachers who are under targets to deliver, and then there is a major inflation in results, is that going to be because there has been a transformation in teaching from last year to this or because there is a systemic problem that it is the regulator’s duty to intervene in?

Brian Lightman: I think all of those things need to be taken into account. I think the issue that we have here is that, when you introduce a new qualification, whether it is modular or whatever the structure of it, you have to put in processes that are fit for purpose. We do not want grade inflation; we want standards to be maintained and improved. The only way to do that is to make sure that you have rigorous and valid and fit-for-purpose assessment processes. That was clearly not in place and the evidence for that is what happened in January: if we had those processes in place, you would not have had that lenient marking.

Q22 Bill Esterson: To come back to what Mike said about other subjects, from what I understand this has only happened in English language. Do people want to give a view on why that is, and is it to do with the assessment in schools, again, where English and maths are so crucial?

Mike Griffiths: With this comparative data, it would have been useful if Ofqual had looked at some of those things, paired examples, because—the corollary of what Graham was saying—I do not believe that the teaching and preparedness of my students was catastrophic this year and has led to the decline. The evidence I have and that I can leave with you is that students maintained their performance in other subjects—in fact, my maths results went up 1% to 92%—including English literature, and that it was this one subject where there has been this catastrophic failure, which has, I believe, led to a great injustice for tens of thousands of students.

Brian Lightman: There is some evidence that there is a bit of a problem in maths, which we have been picking up recently, but it is on a much, much smaller scale. I think one of the possible reasons behind that is English is a subject in which, for obvious reasons,
a lot of attention is paid to improving performance and raising standards of literacy, and a lot of interventions are put in place. That means that there are schools that have made exceptional progress. I think that perhaps throws some light on why it is that some of our high-profile schools, ones that have been identified as making a particular difference, seem to have fallen foul of what has happened here—because they have put interventions in that have pushed their students above that level of expectation. That is probably why, but there are all sorts of other reasons why, with this particular qualification, it just seems to have gone wrong.

Q23 Siobhain McDonagh: I am the newest member of this Committee and I am completely a layperson, so I am sorry if I am about to embarrass myself. The whole comparable outcomes concept is one I have only come across in the last 24 hours, but it is moving against everything that I want to do as a constituency MP. I want to make our schools the best. I am not dogmatic about the form of those schools. I want the kids with the hardest backgrounds to do as well as they possibly can. So I do not want them to do, at 16, what they are predicted to do at 11. I want them to do as well as they possibly can, as any teacher can inspire them to do. We have two academies, and our experience is they have transformed the lives of loads and loads of kids and that this comparable outcomes concept just puts them back into the basket of: “This is what you got when you came in; this is what you are going to get when you come out.” I do not understand how we can all want improving standards but somehow want to stop or cap people from doing that. Have I got it completely wrong?

Kenny Frederick: No, you are exactly right. Russell Hobby: No, not at all. Mike Griffiths: I would agree entirely, but we are not here, I do not think, to debate the entire examination system and the principle upon which it is based. We have to keep in mind the thousands of individual students this year who now are sitting there with grades that are below what they have earned and what they deserve. What I am looking to this Committee, along with hundreds of other head teachers, to do is to help us to put that right for those children. What happens in the future is for another day’s debate, but I think we need to focus on the students. Ofqual, when it came to the conclusion, said in its report that the grades in June are right. They did that based on a statistical analysis yet again. They did not look at any children’s work before coming to that decision that the grades were right. They did not take students who they had awarded Cs and Ds and look at the controlled assessments and papers to see whether the standards were comparable. It was just a desktop statistical exercise to come to the conclusion, “No, we have got it right.” I just do not think that is fair on the students.

Q24 Chair: Exam boards do look at that work. They do exactly the things you have just said and they ensure that within their whole range of marks there is a fair allocation between and among students. That is what they do. It is Ofqual’s duty then to look from a higher level at ensuring fairness and culpability.

Mike Griffiths: All I can say is something has gone wrong in the system somewhere this year that has led to this situation.

Q25 Damian Hinds: It strikes me we might have two separate discussions already going on. One is about what happened at the macro level, on average, to pupils in England in 2012. The other is about differential patterns among different schools, and I think it might be helpful to take those as two separate questions. If we start with the individual, as we have two distinguished heads with us, Mike and Kenny, could you just explain in simple, entry-level, accessible language what happens in your schools: how many pupils we are talking about, what they did in terms of written papers, what they did in terms of controlled assessments, when they did those things and which particular students you are particularly concerned about?

Mike Griffiths: I had 214 students this year who did the GCSE. The teachers have years of experience and years of producing excellent results. They know what a C grade is and what an A grade is.

Q26 Damian Hinds: Sorry, we will get to that, but, just in simple terms, what did they do in terms of papers, exams, controlled assessments and so on?

Mike Griffiths: We had completed the controlled assessments before January, but did not submit them; everything was submitted at the end. They rigorously examined the standardisation materials, examiners’ reports, as a team. They had a clear set of parameters for each grade. They set practice essays for all of the candidates. For the mock exam, using previous mark schemes, they even purchased a complete set of full practice papers from AQA, all 220 copies, so that the boys would have had practice with the new style of paper. They did the re-drafting that was appropriate. We, as a school, hosted moderation visits and we ensured that our students sat every single element of the exam under exam conditions and received targets depending on their performance, but all of them were submitted at the end of the course.

Q27 Damian Hinds: The final exam, if I remember rightly, was worth 40% of the overall and the controlled assessment was worth 60% of the overall. Is that correct?

Mike Griffiths: I believe so.

Q28 Damian Hinds: So the controlled assessments had happened gradually, obviously within school and marked by the teachers, but for some reason you—and I know many other schools did this as well—did not submit them at the time, but amassed them at the end. Why do you do that?

Mike Griffiths: As a school, it is just our policy in all these things. As a school, we believe, as I think the present Secretary of State does, in the notion that you are best prepared at the end of the course when you have completed all of the work and so on, and in all of our subjects they sit the exams at the end of the course.
Q29 Damian Hinds: Sorry, we are talking about controlled assessment. You say you are best prepared at the end, but you were saying that they did the controlled assessment through the year, but you submitted them at the end. So that would not have an effect on the preparedness of the pupils.

Mike Griffiths: We only submit them for grading at the very end.

Q30 Damian Hinds: Okay, thank you. Kenny, is that similar in your school?

Kenny Frederick: Yes. In my school, 160 youngsters did the English and 54 did English language and literature. The 160 who did English were mostly the foundation tier. They sat the exam in June and they did the controlled assessment, speaking and listening, between January and June. The reason you do that is because all the time you are building up the skills. Teachers all the time are working on different skills, so all the time we are looking at them and testing them and seeing which are their strengths, which are their weaknesses. We regroup them to work with particular teachers to build up particular skills, so it leaves different parts to the end. Plus, you have to look at the whole picture of youngsters who are doing 10 GCSEs and you look at the whole year, where the pressure points are, and you try to work that out to give them a fair chance of doing it. The pressure on young people now is enormous. They are constantly being examined and doing it. The pressure on young people now is enormous. They are constantly being examined and our teachers are far more strategic now in the way that they are able to work with them and prepare them.

Q31 Damian Hinds: Is it correct that your concerns are mostly around the foundation level students—students taking papers for which you cannot score above a grade C?

Kenny Frederick: Yes. 10 points were added to that point.

Q32 Damian Hinds: Sure, but specifically on foundation level; that is my question.

Kenny Frederick: Yes. There have been smaller movements in the higher grades, but that is between two and three points, so it is specifically those lower ones. It is schools where they struggle even more—in deprived areas—that are going to suffer more than others.

Q33 Damian Hinds: Okay, thank you. Can I just check with Mike: is that the same with you? Is it foundation level students or is it foundation level and higher level?

Mike Griffiths: It is all the way through. It does affect the foundation students and I have some data on those, but it also affects the top end as well. For instance, the percentage of my students who got As and A*’s dropped from 34% the previous year to 23%, so there was an 11% drop in the highest grades as well.

Q34 Damian Hinds: Can I flip back now to the macro level and ask Brian and Russell? We know that overall the drop in the proportion of students getting A* to C in the English suite was 1.5 percentage points. We also know from the ASCL survey that 87% of schools thought that their results were worse than they had expected and 42% thought they were more than 10% worse than expected. I know we are discussing GCSE English and not maths, but how do you square those two things?

Brian Lightman: 87% of the people who responded to our survey—

Q35 Damian Hinds: What was the sampling technique, by the way?

Brian Lightman: This is just a survey of our school members. We fully understand that the schools that have done better are less likely to respond. Some did respond. So that would explain that difference there. That 1.5% is a global figure across the whole cohort for the whole year. Since we have done that survey we have been getting a lot more information about the effect on schools, and it is a significant number of schools. For example, we have been doing a survey on the number who feel that they have been pushed below the floor standard of 40% as a result of this particular incident and so far—and it is an open survey—143 schools have told us that.

Q36 Damian Hinds: I think perhaps we might come on to that later. Russell, how do you square those two statistics?

Russell Hobby: I think that a very small percentage change can have quite a large impact when the C/D boundary is so prominent. A very small shift there can have a much larger emotional impact and an impact on people’s plans. I think that is why it feels bigger than the percentages would say, and I think it is bigger than the percentages say, because that 1% is quite a lot of students. Whether it is all of the population of students or just a segment of them, those people have been disadvantaged by that, so I think we need to dig below the number.

Q37 Damian Hinds: Does it suggest that schools were particularly bold and optimistic in their expectations this year or most years?

Russell Hobby: I think it shows that they rely very heavily on predictions for targeting their activities within a school. If you have a student with a secure C who is struggling in maths—or what you think is a secure C—you put your weight and effort into helping them become better at maths. Therefore, you were operating on what you thought was a legitimate expectation, because you thought, “They will never shift the grade boundaries by 10 points on this one.” So I think it is that lack of warning and the way that schools are, in fact, extremely strategic, as Kenny said, in allocating very limited resource.

Q38 Chair: Schools were more optimistic this year, were they not? According to Ofqual, typically schools over-predict. This year they over-predicted more than usual: they predicted 15% more students would get Cs this year than last year, whereas normally it is 12% and they normally over-predict. This is more excessive this year than usual. Isn’t that true or is Ofqual wrong?

Kenny Frederick: No.
Mike Griffiths: It is going to vary from school to school. If my head of English was here, she would resent hugely the implication. Their predictions are incredibly accurate. They are rarely more than 1% out either way, and last year they were quite ashamed because they went up 2% on their predictions. They are pretty good at it. I think it is a bit of an insult to the professional integrity of a lot of people to just assume that schools are artificially inflating their predictions.

Q39 Damian Hinds: Perhaps this comes back to the distinction between the individual school and the overall macro picture, and I am sure that is something we will come back to. Before marks were adjusted, we learned that in correspondence between Edexcel and Ofqual there was, at one point, an eight percentage point growth in performance in GCSE English for the Edexcel set compared with the prior year. Notwithstanding everything we have all been saying about obviously everybody wanting all children to be improving all the time, and, of course, in certain schools there will be exceptional situations where that can happen. The cohorts do change and one of the things we heard on the exam results day, in the briefings, was about the differences between the cohorts. Of course we all know individual students, even very large numbers of students, can do statistically, I just think that is unfair.

Q40 Damian Hinds: The starting point is the prior year.

Kenny Frederick: When we had Olympic medals, did we put in a cap and say, “That is going to be the limit,” or did we put all the measures in? This is a bit like Usain Bolt or Mo Farah running around, getting to the end of—I am not very sporty, sorry—the track, winning the race and suddenly being told, “Sorry, there was another 100 metres you should have done.”

Q41 Damian Hinds: With respect, Ms Frederick, that is not my question at all. I am not talking about an individual student.

Kenny Frederick: But this is what it is all about.

Damian Hinds: Of course we all know individual students, even very large numbers of students, can do extraordinarily well, but at the macro level across the entire year cohort, and bear in mind you were also the head teacher in this school the previous year and all your staff were also doing a great job the previous year, can performance grow by eight percentage points in one year?

Brian Lightman: Yes it can; it depends very largely in one year? Can performance grow by eight percentage points in one year?

Q42 Damian Hinds: But do we believe that would have had a net positive effect or a net negative effect?

Brian Lightman: It depends on the cohort.

Q43 Damian Hinds: But from what we know of the changes in the cohort this time around, as I understand it, would a decline in the number of pupils from independent and selective schools, for example—other things being equal—given the mix of results that you have, have a net positive impact or a net negative impact?

Brian Lightman: That is too technical for us to try to make an informed judgment about at this stage.

Q44 Damian Hinds: Ofqual have helped you by saying: “There were about 23,000 fewer candidates from selective and independent schools, about 3.4% of the total, who will probably have migrated to international GCSE or other qualifications. These candidates will typically have relatively high Key Stage 2 point scores.”

Brian Lightman: One of the points there is that Ofqual did not have the data on the whole cohort when they were preparing for these comparable outcomes. They only have data on 75% of the cohort. You have children who have come in from abroad who did not have Key Stage 2 results. You have students in the independent sector who have not done Key Stage 2 and so on, so it is not data on the whole cohort.

Mike Griffiths: We can, if you like, play games with statistics, but the simple fact is that there are hundreds of schools like mine—one in Suffolk I have data for, another one in our county—where there are dozens and dozens of students who have got D grades. What has happened in the rest of the country, in one sense, as a head teacher, is not relevant. What is relevant is, in my school, the 30 boys who have not got a C that one would have expected, if standards were maintained, would have done; in another school, the 55 students in the same position. You would have almost thought that, if there is a removal of students at the top end, a high performing and successful school like mine and the one I am thinking of in Suffolk would have benefited from that, because their students would, apparently, have been even better. The opposite has occurred, which is that in many high-performing schools, if they entered all their grades at the end of the course, their students have, in great numbers, had their grades downgraded. Whatever you do statistically, I just think that is unfair.

Q45 Damian Hinds: I hope very much we will come back to these individual school effects, which are absolutely crucial. The last question I want to put is to Russell. I was going to ask a question to try to help my understanding of how publishing the grade boundaries in January could affect the actual results of children later on. I can understand how it can affect expectation, but not how it affects outcomes compared with if there had been no announcement of any grade boundaries at all. I think you already answered that question when you said that, if you had a banked C, as it were, in English, you might then move on and put more emphasis on maths or other subjects. Can you just expand on that a little bit more for us?

Chair: Perhaps briefly.

Russell Hobby: I could just say “yes”, like I answered it previously—I could contract on it. If you have a secure C, then it would be sensible, in the child’s own best interests, to focus on subjects in which they are not certain to get that, because if you get a C in
English and a D in maths, that does not do you any good at all, so you would not keep working on. So you may have assumed that child was safe and secure in that C for English, you have done the work that needed to be done and, given limited teaching time, you are going to focus on their other, weaker, subjects. Damian Hinds: Thank you.

Q46 Alex Cunningham: What did Edexcel tell schools about the changes after the January exams? What contact has there been about changes?

Kenny Frederick: We went with AQA and they did not give us any feedback on that at all. We were working towards the grade boundaries we had with an expectation that might vary between two or three points.

Q47 Chair: Who told you it would vary between two and three points?

Kenny Frederick: You always know that with exams. That is expected.

Q48 Chair: So no one told you that; you just thought that.

Kenny Frederick: Yes. If we had been told that suddenly there was going to be a 10-point difference, we would have been able to sit down in the English department to see what they had to do to make sure that the children were able to have a go at getting those additional 10 marks. We are really cross that we did not know, and therefore we could do nothing to support them.

Q49 Alex Cunningham: So the schools did not see it coming.

Kenny Frederick: No. There was no indication.

Mike Griffiths: My head of department, on the morning of the exam results, came into my office in tears, basically offering to resign. We had absolutely no idea this was going to happen. We had no idea from the boards that this was going to happen. As I quoted to you from the board’s own comments, they feel that the department had prepared the students well, had marked them accurately, it was consistent and they had applied to the new specification in an exemplary fashion. So why would those experienced teachers, who know what C and A grades are, have been expecting that huge decline?

Q50 Alex Cunningham: Edexcel are not due before us immediately at this stage, but they have to answer some questions here, surely.

Mike Griffiths: Like Ms Frederick, we were AQA as a board, but, again, had no information from the board and at no stage was the department given any sort of indication that this might happen.

Q51 Neil Carmichael: Do you accept that Ofqual was right in saying that the January grades were a bit generous? If so, to what extent do you make that measurement?

Brian Lightman: We have to accept that, but that should have been identified at the time. What we are saying here is that is the purpose of moderation. That is what should be happening. It is very common for schools to have their marking for course work or controlled assessment—those types of things—adjusted to make sure that it is of the right standard, but that was not the case. We have plenty of schools that were told by boards—and it is not just one board but all boards—that there would be little change from the January boundaries. They were told that in conversations that they had between January and June. Russell Hobby: To be honest, the time to act was before January or after this year’s results, not in the middle of this year. It is almost a belated realisation that they are going to get into trouble with grade inflation, and therefore the June cohort is being punished for that. I think Ofqual should have taken that hit and then revised their procedures for the year to come rather than the June cohort.

Q52 Neil Carmichael: So they were too late in noticing what had gone wrong in January.

Russell Hobby: Yes.

Q53 Neil Carmichael: Effectively then correcting—Brian Lightman: Harshly in June.

Q54 Neil Carmichael: At what point do you think Ofqual realised they had made the mistake?

Russell Hobby: There is some evidence to suggest they knew three years ago that there were going to be problems at this level, and there has been a lack of institutional memory within Ofqual as well to do that. They should have known it was coming rather than realising afterwards. Between January and June, who knows what went on; that is something for them to comment on, but it was far too late.

Brian Lightman: They did have the work of 54,000 candidates in June 2011 to look at as well.

Q55 Neil Carmichael: So they could have done a comparison.

Brian Lightman: Yes.

Q56 Neil Carmichael: You obviously feel they did not do that or at least do it properly.

Brian Lightman: Yes.

Q57 Chair: The allegation is a straightforward failure of moderation and grade boundary understanding at an earlier time when there was sufficient evidence to come to that conclusion, and their suggestion that there was not you do not think is fair.

Brian Lightman: Yes.

Q58 Craig Whittaker: Had schools and head teachers known earlier, would that have made a difference?

Kenny Frederick: Yes.

Q59 Craig Whittaker: In what way? Are you saying you would have worked harder?

Kenny Frederick: If you are the English teacher and you are working with your groups and you work as a team, you look at the exam specification, what needs to be done and how you get the points. It is very
sophisticated now. Teaching is really, really good now and I do not think people recognise that. They would then have looked at what youngsters have to achieve to get the grade.

**Q60 Craig Whittaker:** Why didn’t they do that anyway? If there is room to get these kids to do better anyway, why aren’t teachers automatically doing it?

**Russell Hobby:** Teachers are working as hard as they can.

**Q61 Craig Whittaker:** So it would not have made any difference.

**Russell Hobby:** No, because they could have used those hours on the subjects in which they were not secure, as I was saying to Damian. If you have a secure C, you are not going to spend more time when a student is lacking a D in another subject, because they need that D. It is not the school that needs it; it is the child that needs that to get their places. So it is a question of working smarter rather than harder. I do not think you can get anything more out of the teachers themselves in terms of effort.

**Q62 Craig Whittaker:** So it would not have made any difference then.

**Brian Lightman:** I think you are highlighting a very important point here that is for a later discussion, perhaps, about what teaching is about, because we are so focussed on the exam system here; it is a very strong focus. But that is another discussion, I think.

**Kenny Frederick:** At least give the schools and the teachers a chance to do something and to give children a fair chance at this. This is completely unfair and really is not acceptable at all.

**Q63 Neil Carmichael:** Could you talk about the evidence you have along the lines of Ofqual not applying the comparable outcomes approach in a proper and consistent way?

**Brian Lightman:** Yes. We have collected statistics across the whole examination here. We have looked at the grading of all of the different parts of the exam. We have looked at the percentages that are going there and we have looked at those variations there. I am struggling; I do not think I can just describe all of that information. What I can offer, and we have been offering, is to provide anything that will be necessary to help there, and we have been working closely with Ofqual to give them that information and feedback, but we have been gathering a lot of evidence from schools about it.

**Q64 Neil Carmichael:** Can I ask one more question? How can we map out the variance between schools, which has been referred to since we started this session in one way or another? That is an equally important question, which we discussed prior to your arriving here, and it seems to be still not properly calibrated.

**Brian Lightman:** You have to go back to looking at the pupils’ work and you have to look at comparisons of prepared subject analyses as well to see how the students did across their other subjects. That is going to require some digging back in schools, but I think it is absolutely essential, in a future investigation of this, to get to the bottom of what has happened here. There are urgent issues that need to be addressed here, obviously, in terms of the grades those pupils have got, but there are a lot of very searching questions that can only be answered by looking at that sort of level of detail.

**Q65 Chair:** Can I ask you then to just briefly tell us what you want to happen now, what needs to be done, and who you think is best suited to do it?

**Russell Hobby:** There are two initial things: there does need to be an independent inquiry into this. I do not think Ofqual can investigate itself on these matters because I think they are part of the issue. Now, an independent inquiry may reveal that they have made all the right choices and so on, but at least we would then have the confidence in that.

I think there should be a re-grade for those students affected in June. They are competing against the January cohort. They are competing for places and for jobs and apprenticeships. If comparable outcomes is going to mean anything, it ought to mean comparable within a year as well. They were not given the predictions and the chance to overcome the new grade boundaries. If they had been warned about them, they could have done something or pulled something out of the bag in order to achieve that. So I think they have been let down.

**Q66 Chair:** Do you welcome the re-sit opportunity?

**Russell Hobby:** It is not enough, and it does not reflect that you have to study the GCSE again. You have to start again for parts of that as well. Plus, people are being entered in Level 2 courses when they should be in Level 3 courses in further education. It is an inadequate redress, I think.

**Mike Griffiths:** I think the re-sit is a complete red herring; it has nothing to do with this. The students have already done the paper and they have already performed and whatever else. It is not because of any shortcoming in the students that this has happened, so I think that offering a free re-sit is not useful.

**Q67 Ian Mearns:** The thing that strikes me from the previous discussion is where Ofqual state that “it is regrettable that the publication of grade boundaries for the January assessments could have led schools to assume that the boundary would remain constant, and we will review with exam boards any lessons from this”. Why would schools think that the grade boundaries would be changed within the same academic year?

**Brian Lightman:** Awarding bodies need to have the right to change grade boundaries if there is a different level of challenge between the papers. That has always happened and that makes the awarding fairer, but in this case there was no reason to expect that it would be this sort of extent of change.

**Q68 Ian Mearns:** I think any rational person would assume that grade boundaries would remain constant within an academic year.

**Brian Lightman:** Within an academic year, yes—between years.
Q69 Ian Mearns: Do you really feel that Ofqual have treated young people who have suffered because of this as some sort of collateral damage in the face of the Secretary of State’s concern about grade inflation?

Russell Hobby: I think Ofqual have focussed too much on the statistical arguments and not on the people involved in that. I felt the initial interim report was remarkably complacent about its impact on the individuals concerned. In fact, I do not think they were mentioned within it. We are all getting dragged into statistical arguments here and, indeed, there does need to be some sort of statistical base to this, but it seems like we are letting the statistics and the calculations drive the choices that we make rather than simple justice for the people involved. I think whenever a system has got to this level of complexity, where interested laypeople from all sorts of directions cannot really tell what is happening, we have a system that is out of control.

Q70 Ian Mearns: It feels to me that there is some sort of element creeping in here that there has to be some limit to progression—that only a certain amount of youngsters are allowed to go above that level, so there is a glass ceiling being introduced artificially halfway through an academic year.

Kenny Frederick: That is exactly what is happening. I had to go and speak to year 11 yesterday and try to explain to them what was happening and what is going to happen this year. They are going to sit their exams and yet they do not know what they are going in for. They do not know what the expectation is. They do not know what marks they are supposed to get. Their teachers do not know. It really is not right.

Q71 Ian Mearns: Because of the importance now for the individual, but also for institutions, of English and maths, do you think this has had a particularly strong effect because of what has happened in English this year? You are all nodding.

Mike Griffiths: I think that is true. At the end of the day, the statistics are one thing, but ours, if you like, is a people business. It is about the students themselves. I do not know any other issue that has united the NUT and HMC, the academy chains and local authorities. They are all united in this—that there has been a miscarriage of justice—and that is why we want to see a change. I am sure you have colleagues who have children who have sat the GCSE this year and who have been affected by this. I know of one MP because of a local school, and her son did do the GCSE and has suffered at the hands of this. At the people level, it requires something to be done.

Q72 Mr Ward: First of all, I think the language has been a bit difficult, because adjusting marks and grade boundary changes are not the same thing. There was an opportunity to adjust the marks through moderation in January and the decision was made not to do that. Can I just come to you, Mike? You did not cash in, in January. There must have been some results that you saw in January of which you thought, “That is a bit better than I expected.” Wasn’t it risky to leave it until June? Why didn’t you cash in some of the results where you thought it might be risky if you left it until June?

Mike Griffiths: There is no reason to. My head of department had no reason to imagine that there would be a problem. We did not come into this assuming that there was going to be this level of turbulence in the system.

Q73 Mr Ward: So based upon the results received, with another four or five months to work with the young people, knowing what the boundaries would be, you felt you could enhance the January results.

Brian Lightman: I think there is another point here. Whilst some of that January marking was over-generous, that is a global statement across the whole system. It may well have been that in Mike’s school it was not over-generous, because the moderation was only a very, very small sample of the people who took it.

Q74 Chair: You still have to have 60% of all the marks in a qualification awarded by controlled assessment, namely by the teachers concerned.

Brian Lightman: That is a different discussion to be had in the future and one that we must have, but the issue here is that, if you have that system, you have to have proper safeguards in place. I think that is where the difficulty is.

Q75 Ian Mearns: Kenny, you talked before, particularly in the sort of school that you are head of, of the youngsters who did not make the C. What is happening to those youngsters now? A lot of them will have been disappointed and not progressed and left. What is happening to them?

Kenny Frederick: We have a sixth form, but we do the International Baccalaureate at our school, which is quite a difficult qualification to do. We do some Level 2 courses also. We are working with the colleges and with ourselves to see what arrangements we can make for them. For youngsters who, through good career planning and support, were going into courses that they had chosen, they had applied for and so on and they cannot now do, the colleges are trying to find them Level 2 courses. Occasionally they are being told that they can do the Level 3 course, but they have to spend a year doing the exam again, which takes their eye off the ball of the original course they are doing.

Some are so disgruntled and so demotivated they feel that, no matter what they do, they cannot succeed, because when they keep moving the barriers or the goalposts, you cannot succeed. They are talking about taking a year out. Now, I know with some of my students that a year out means they are going to turn into NEETs. I think this is a big issue; we are going to increase the number of young people not in employment or further training and so on. Some who were supposed to do apprenticeships have not been able to go on to them and we are trying to find them alternatives. My staff are working with those individuals, trying to place them so they are not sitting at home.
Q76 Ian Mearns: According to Ofqual, AQA found evidence of significant teacher over-marking. Should exam boards have been adjusting schools’ marks downwards if that was the case?

Kenny Frederick: First of all, when mine did the exam in January, because our expectations and the results were right, what we had expected to happen did happen. With the controlled assessment, you receive 20% speaking and listening and 40% controlled assessments. They have to be taught certain things, whatever the question is, and they have to work on that, do essay plans and so on. It is moderated, so the exam board asks for samples, they look at the work, they moderate the work, they give feedback. The feedback we had commended our marking, agreed that we had awarded grades and so on. It is not as if teachers just mark. Because of the nature of our school, we made sure that those controlled assessments were checked and moderated by external markers as well as our own staff. So there is no way that the teachers were over-marking, and certainly AQA gave no indication that was the case.

Mike Griffiths: Again, just to reiterate, we were AQA as well and there was no evidence at all that the teachers were over-marking, and certainly AQA gave no indication that was the case.

Q77 Chair: Assuming that is the case, could it be that the schools that were congratulated and had very high standards and did not boost marks to suit their league tables were disadvantaged at the expense of others? AQA are saying, overall, they think there was significant evidence of over-marking, but it was within the tolerances so that they could not intervene. Within the tolerances, people were marking up because they thought they knew where the grade boundary was and they boosted the marks to their pupils in order to boost their apparent performance. Could it be that the schools that did not do that have been penalised as an expense of those that did?

Kenny Frederick: I do not think AQA gave that information to anybody, as far as I know from the feedback we had commended our marking, agreed that we had awarded grades and so on. It is not as if teachers just mark. Because of the nature of our school, we made sure that those controlled assessments were checked and moderated by external markers as well as our own staff. So there is no way that the teachers were over-marking, and certainly AQA gave no indication that was the case.

Q78 Ian Mearns: Lastly, to Mike and Kenny in particular, did your staff attend exam board training events for GCSE English? If so, what information was given at those events about controlled assessment grade boundaries?

Kenny Frederick: They certainly went. They always go to the exam moderator meetings. I cannot tell you precisely what they were told. Whatever they were told, they followed to the letter. They followed the exact instructions, they took advice and we brought people in as well who had come from the exam board to make sure that everything was okay.

Q79 Ian Mearns: What you are saying is, if Ofqual are saying that AQA said that, they kept it to themselves.

Brian Lightman: Yes.

Q80 Ian Mearns: According to Ofqual memorandum to us says that Q78 Chair: feedback I have had from hundreds of schools. information to anybody, as far as I know from the AQA found evidence of significant teacher over-marking. Should exam boards have been adjusting schools’ marks downwards if that was the case?

Russell Hobby: Yes.

Q81 Bill Esterson: Given the importance of getting five or more grades A* to C, isn’t this as much about the league tables and the need to try to get above the 40% floor level as it is about individual students?

Brian Lightman: Of course the league tables and the accountability framework drive behaviour, and schools could not possibly ignore the fact that is what the expectation is. But the issue here—and I think this is why there is such strength of feeling about this—is this is about students. We are here talking about the students today, and we have examples; we have lists of names of people who have lost out as a result of this, and that is what the issue is here. We need to have conversations later about the effects of the accountability system and so on, but they are different discussions. The issue here is about the students.

Russell Hobby: Of course, five A*s to C matters as much to the student as it does to the school, so your interests are aligned. If you have not got that, you are not going forward. But I think throughout this conversation you can see just how much our examination and assessment system is distorting and changing what is going on in schools. Schools are becoming—and are—exam factories. You have seen the strategy and tactics that are used. This is in the interests of the students as well, because if they do not get those five A*s to C they are not going anywhere. But we have a schooling system that is dominated now, from the age of five with the phonics screening check, by tests. Once we have solved the problem for the students in this year, we need to move on and look at what damage is being done to learning by the examination system.

Q82 Bill Esterson: Do you think the reason that English has had a particular problem—I think one of you said earlier that there was slight evidence of something in maths as well—is because of the importance of five A*s to C?

Brian Lightman: Yes, because literacy is the key to every other subject and to every career that you need. So all of the strategies that have been implemented by schools, both at school level and, indeed, at national level if you take London Challenge, National Strategies—all these other things that have gone through in the past—they have all, rightly, focussed on standards of literacy, and that is absolutely what needs to happen.
Q83 Bill Esterson: Is there a tendency in any way for teachers to be, however slightly, generous in marking assessments, particularly at the borderline?

Brian Lightman: I think the quality of assessment has improved absolutely beyond recognition over the years. It has been a major, major focus for training in every school in the country. When I go into schools now, as I do, as often as I can, head teachers show me how they are monitoring the progress of their students and what interventions they are putting in as a result of that assessment. It is not in their interests. Indeed, if Ofsted came in and you were assessing over-generously, you would be, rightly, in for criticism there. It is a science. Assessment is a very high-order skill for teachers and it is something that is, rightly, emphasised as a very important aspect of our work.

Kenny Frederick: It is moderated by the exam boards very carefully, and we have to take back what they tell us and act on that, so it would be in nobody’s interest to over-mark. If anything, teachers probably under-mark.

Q84 Bill Esterson: So the suggestion that could happen is completely wide of the mark.

Kenny Frederick: I see no evidence of that.

Q85 Chair: Except for AQA stating that it is the case and there being an incentive system precisely to over-mark. There is transparency in the system, so you know exactly what the tolerances are and how many marks you can increase over what it should be, and you know there cannot be intervention, and you have all these targets and you are spending every effort to try to reach these targets for the school’s and the child’s benefit. To suggest that there is not an incentive to over-mark is absurd.

Kenny Frederick: As we pay AQA a large amount of money—we spend a huge amount of money on these exam boards—would they not do us the courtesy of giving us that feedback? We take feedback and we act on feedback.

Q86 Chair: It is a different point, though. There is an incentive in the system—which Russell highlights has all sorts of perverse outcomes—within the tolerances to over-mark.

Brian Lightman: What we need to get back to is talking about what standards a student should learn. What should they be learning? Instead of chasing a grade C or a grade B, we should be talking about teaching and learning.

Q87 Chair: I entirely agree. I must bring this to an end. It has been a fascinating session of the Committee this morning. Personally, I am particularly struck by the issue of what might be called the “turbulence” in the system, this variability, because we are talking about the grade boundary as if it is uniform in its effect. This year is peculiar in that the application of standard lines is impacting people in a much more variable way than before. That means that there are thousands of children who are being treated in a way that is different from prior years, and I am not sure we are that much further towards understanding why that variability occurred, regardless of where you put the boundaries and what you do between January and June. So anything you can do to throw light on that would help us understand who has been ill-affected in a bad way and how many, and therefore, that might contribute to any understanding of what needs to be done about it. Thank you all very much; it has been very useful.

Examination of Witnesses

Witnesses: Glenys Stacey, Chief Executive and Chief Regulator of Qualifications and Examinations, Ofqual, Amanda Spielman, Chair, Ofqual, and Cath Jadhav, Acting Director of Standards, Ofqual, gave evidence.

Q88 Chair: Good morning and welcome to our second session this morning looking at the issue of English GCSEs in 2012. Perhaps it might be best to start with as succinct an overview as you can give us, please, as to what has occurred, what is commonly agreed and where there is dispute.

Glenys Stacey: Thank you very much, Chair; good morning. The position is this: GCSE English results were down A* to C this year by about 1.5% overall. That was very much in line with expectations because of changes to the student mix, but there has been an unusual distribution pattern of results: there is a greater variation between schools than expected, it seems. For some schools, these results are a far cry from their expectations, and so we have been looking carefully at what may lie behind that.

I think there are a few important points that we would make. First of all, there has been no political interference; I am happy to talk about that. Secondly, awarding and grade boundary setting worked as it should have done for the English suite and as it worked, indeed, for other GCSEs, AS and A levels. So we played our proper part in regulating standards there, and neither the exam boards nor the regulators did anything other than what they should have done. But it is very clear when we look back that in January awards were, as we have said, generous, but that could not really have been seen by examiners at the time. The issue really seems to be, as we boil it down, that students have taken different routes through these qualifications. There is what you might call a route effect, which we need to understand more. These were designed as modular qualifications with 60% controlled assessment, but it seems that a good number of students took them as if they were linear qualifications. We are looking, for example, I think, at 45% doing that in AQA, the biggest provider, and we may well be seeing here a route effect, and we need to understand more about whether that route effect is pretty well what you would expect for any qualification that was used in that way, or whether
there is something particularly different about these qualifications.

Q89 Chair: Thank you. You have known all about the risks around new qualifications, around modular construction and controlled assessment for several years. Why didn’t you anticipate the situation you now find yourself in and do something about it?

Glenys Stacey: We have certainly known about modularisation and the issues that arise. Modularisation was from a different era, I think, and we have announced that we are making changes to GCSEs. These are coming into effect from September, and move to a linear approach. We are currently consulting on that for A levels as well. That has come from research we have done and published, where we have looked at international ways of assessing. So we recognised that modularisation can have an effect on standards, and we very much want to see a better arrangement there.

On controlled assessment, the arrangements that we inherited provide for different levels of controlled assessment depending on which GCSE subject you are dealing with, and it is a particularly high amount for the English subjects; it is 60%. As you will know, Chair, we reviewed controlled assessment last year and we found from schools some issues about the workability of it, and so we then worked with exam boards to make things a little bit more straightforward. This year, we started another review of controlled assessment, a more fundamental review, looking really at its suitability. We were concerned there about what we were hearing from schools, most particularly in English, about the impact it was having on teaching and testing in schools. In short, the amount of time spent preparing for it, then delivering it and then assessing it was thought to be disproportionate and eating unduly into teaching time. What we can see from the evidence in the English suite this year is that there are other considerations as well that can now feed into that review.

Q90 Chair: We know you are aware of the risks around the introduction of new qualifications. You have well known views. You are reviewing or are scrapping modular controlled assessment. You are totally on board with the risks, and yet we find ourselves in this position. The key question is why you did not anticipate that and what you could and should have done about it, so that we did not find ourselves in this situation. The key question is why you did not anticipate that and what you could and should have done about it. I see; then let me try to explain that.

Glenys Stacey: I am sorry, Chair; I got carried away with modularisation and controlled assessment. We did see risks to the English suite. We approach oversight of these qualifications, where they are delivered, on a risk basis, and we had identified that we were looking particularly this year at new qualifications and those in significant high-stake subjects, and, of course, the English suite was one of those. So we did put in place a particular scrutiny programme right from the beginning, so we focussed on the suite. We have attended many of the meetings for written papers and controlled assessment, for example, as boundary setting is done, many more than we would do ordinarily for the average qualification. We did not see anything at any of those meetings that meant that we needed to take further action, and we are happy to take more questions about that.

Q91 Chair: That goes to the heart of it: why didn’t you? We have just heard evidence that, although some modules had very small entries, the foundation tier, as far back as June 2011, had more than 50,000 people, if I recall, sitting it. So the allegation is that there were sufficient numbers. If you look at A level subjects, which rely on essay writing, some of those have much smaller entries than the entries that you saw in January this year or June last year and, basically, you should have had enough data and information to be able to do this. You were already alert to the problems, and yet somehow we still find ourselves sitting here today. That must mean that you failed somewhere to do what you should do in order to anticipate it. That is the allegation.

Glenys Stacey: I understand and the position is, as ever, a little bit more complicated than it perhaps first appears. The foundation tier paper is, of course, an examination paper, and examinations are, by their nature, different. An examination can prove to be more or less demanding and more or less effective in what it is trying to do, so it is not at all unusual for boundaries on examined papers to change. The critical issue, I think, for us here is in the assessment of speaking and listening, which is part of controlled assessment, where it is notoriously difficult to assess speaking and listening; that has always been known.

But in the speaking and listening units with, for example, the major provider, AQA, in January this year there were 14,000 students that sat that. In June, there were 376,000, so a very large percentage were delaying until June. It is not that different for Edexcel either; I think it was 700 in January and 23,500 in June.

Q92 Chair: 14,000 is large for many subjects. Are we to assume that any entry that is less than 14,000 has unstable and unreliable results? That is what we are struggling to understand.

Glenys Stacey: I see; then let me try to explain that. It is a relatively small amount of the whole and that is not unimportant. The other issue is that, in those early assessments, examiners did not have much information other than the scripts to go on. It was a new set of qualifications, so they were not able to rely explicitly on past examples of performance that they could truly compare in the way that you can do and examiners do do when qualifications are well established. It is always a problem when you introduce new qualifications.

Q93 Chair: Isn’t your key control, rightly or wrongly, the foundation of the comparable outcomes framework? You have developed and discussed and been open about that: to look back at Key Stage 2 results—in other words, how kids do at the end of primary—and then map that forward across decent numbers. That does not stop individuals moving from the bottom to the top, but allows you to check whether you are maintaining comparability. Surely, across
14,000 earlier in the year, you would have applied exactly that and the alarm bells should have rung and something should have been done then.

**Glenys Stacey:** I think the position, Chair, is that the key piece of information available to exam boards by June, but not before that, was a statistical prediction based on prior attainment of the candidates, the cohorts. It was based, as you say, on the relationship between national outcomes in GCSE English and the Key Stage 2 scores for the cohort. If I could just explain the prediction, it makes the assumption that the relationship between prior attainment of a student and outcomes ought to be stable from one year to the next. It was used definitely to guide the 2012 June award in a way that simply was not possible for the awards in each of the earlier units.

**Q94 Chair:** That gets to the nub of it, but could you explain why that was the case? I think a lot of people are wondering how it was possible that did not happen before.

**Glenys Stacey:** I can quite understand the confusion. The statistical predictions that were used to inform the early awards were, compared with June, a comparatively weak source of evidence, because they were thin. They were based on it being a new qualification. We know, for example, that in AQA it was understood and explained very carefully to the awarding committees at the time that this was not as robust an approach. Why not? Because, at the time, you were looking at individual units early on, again with low numbers, and the position is that, at that time, there are several unknowns. There are more unknowns than there are when you have a stable qualification or at the end of the qualification. The unknowns are these, for example: the strength of the correlation between the units, which strongly influences how unit outcomes aggregate to give subject outcomes; the impact of the move from coursework to controlled assessment on mark distributions and, therefore, the aggregation. That move happened in other subjects, but of course here we are talking about 60% and we are talking about English and the assessment, particularly, of speaking and listening. Another unknown until the end was the impact of changes to tiering and candidates being able, for the first time, to get a grade A on the foundation tier. So there were very significant unknowns until June.

**Q95 Bill Esterson:** Coming back to the statistical point, 14,000 out of 376,000 is statistically a very high sample number. If you compare it with the way that public opinion polls or market surveys are done, if you look at public opinion polls, roughly 2,000 people are used to predict the political allegiances of a population of many, many millions to within an accuracy of 95% to 98%, within a margin of error of 1% or 2%. Surely the comparison is that 14,000 out of 376,000 should have made that very much easier to predict than the equivalent in our business.

**Glenys Stacey:** I wish that it was so. Of course, we know that in political polls the sample is very carefully collected to be, so far as possible, representative of the population as whole. It is not like that in awarding.

**Q96 Bill Esterson:** I am sorry, but in the field of statistics—which I can barely pronounce—a large sample size of 14,000 irons out some of those anomalies and, in fact, good analysis enables you to do exactly that by weighting the data yourselves.

**Glenys Stacey:** I will ask Cath to explain a bit more in a moment, but we know that in these early units you may get, for example, particularly bright children being put forward, because they can put that unit out of the way and under their belts and get a decent grade.

**Q97 Chair:** Sorry to take you back, but if we are talking about the comparable outcomes as being the foundation, and the foundation of that is the Key Stage 2, then if the percentage at any particular time happened to be weighted more to the kids at the bottom or the top academically, prior attainment, you would immediately know that. This was not 50 compared with 300,000; this was 14,000, as Bill has said. We are struggling to understand why it was not possible. Was there something missing? Was software not available? Were techniques planned for June that simply were not used in January? I am none the wiser; I may be being dim.

**Q98 Mr Ward:** The 14,000 was not the sample; it was the population. It was 100% of the population. What you are saying is that, out of 100% of the population, you could not do an analysis to say whether the marking was fair.

**Glenys Stacey:** What I am trying to say is that those 14,000, firstly, may not have been representative of the whole, but I understand the point you are making. Secondly, when awarders are awarding they are trying to understand, so far as possible, what they have got in that 14,000. There is some information that can help them, some understanding, school by school or centre by centre, but there is not a lot. So it is much easier to do the job with a greater level of assurance when you have a much more sizeable proportion of the whole of the cohort. I hope that is one element that is understood. That is the first point.

I think the second point is coming back to the unknowns that I mentioned earlier. They are extremely relevant as awarders are trying to gauge standards, particularly if there is nothing else beforehand directly comparable to go on.

Thirdly, and I know this is tricky, English is a particularly difficult subject in terms of setting standards. It is not a hard-edged subject where figures add up and so on. There is a large element of judgment. Now, we know from the scrutiny studies that we did looking at how it was done and from the reporting that we had back from awarding bodies, because we were looking closely at this, that examiners did find it difficult. They did say it has been particularly difficult, but the best judgments that could be made were made. Cath, I wonder whether you want to add anything else to that.

**Cath Jadhav:** As you say, we have Key Stage 2 predictions, but they predict how the students will do...
in the qualification as a whole. The awarding bodies do not know, at the point they are making individual unit entries, how that translates into unit level performance, and that is the difficulty. We know where we expect these students to end up, but we do not know how they are going to get there.

Q99 Chair: So, in effect, it is about the fact that only at the end do all the elements come together, and only when all the elements have come together are you able to triangulate, or whatever the right word is, and ensure the comparable outcomes. I understood that point; that is good. There were some letters between you and a couple of the exam boards that were published this morning or last night. Could you tell us about the discussions you had with exam boards about GCSE English results prior to the publication, to what extent there was a to-ing and fro-ing, and how normal that is? Was this extraordinary in any way? Can you tell us a bit about that?

Glenys Stacey: It certainly was not extraordinary, I can assure you of that, but the best way to approach this to assist is if you explain, Cath, the general approach ahead of the regulator’s awarding meeting, and then I can explain the specific position we found ourselves in with Edexcel, if that is what you are interested in.

Cath Jadhav: Yes. Throughout the awarding period we meet regularly with the awarding bodies via telephone and we receive the data coming out of the award—the emerging data for each subject as the award meeting is completed—and how that compares with the predictions. We review that with Ofqual with the three regulators, and we will go back and challenge or question anything that appears out of the ordinary. Then we pull that together for two separate meetings with the exam boards to look at A levels and GCSEs.

Glenys Stacey: Those meetings occur as soon as we have sufficient provisional data across the subjects—there are about 46 or 47 GCSE subjects—and across each of the exam boards. Just to set the context with Edexcel, they are a relatively small player in the GCSE English market, although their share has risen year on year from 5% to 10% this year, so they are making inroads in the market. The preliminary result came in in good shape and there would have been a different outcry, I suspect, had that been the outcome. They were certainly out of line with the other awarding body results and, of course, our statutory role is to maintain standards over time and across awarding bodies as well. At the meeting, the regulators were there, but also each of the leads in the exam boards. We discussed this at that time, and there was a clear recognition, a view, that this really did need challenge, and so I agreed that I would deal with that with Edexcel on a one to one, which is what you would expect.

I should say that at that meeting I did recognise and say that I knew that awarding GCSE English had been a real challenge for exam boards this year, and I acknowledged that they had been trying very hard to get that right.

Q102 Chair: You did not have to direct them in the end. In a way, they accepted the fairness of the line you agreed, because if they had not, they could have refused and then insisted that you directed them to change.

Glenys Stacey: Absolutely so.

Glenys Stacey: That is right.

Q103 Chair: Did you direct any other awarding body to change in any other subject this year?

Glenys Stacey: No, but we were prepared to; that is our job.

Q105 Chair: It also gives an indication that, if nobody ever forced you to direct them, they are pretty keen on not being directed.

Glenys Stacey: No. You will know that we are there to do a job: to maintain standards. If a direction is required, we will do it. That is why Parliament set us up as it did and gave us the powers that we have. I know that was recognised when our enabling legislation was passing through Parliament—that this would be required to hold the line. Ultimately, it is the exam boards’ decision as to how they bring the qualification in. We basically did a proper job in writing to Edexcel to remind them that it was their duty to do that, and they came and responded to us with information as to how they were going to do it.

Q106 Chair: Can I move you on to WJEC? They are a rather anomalous situation.
Glenys Stacey: Absolutely. I should just say, though, because it is not entirely clear so far in debate, that the proposal that Edexcel came back with still left them out of tolerance. They were still showing year-on-year increases over the year before, and we had a difficult judgment to make as to whether we thought that was tolerable, if you like. We do have to make those judgments and we did, and we concluded that we would accept the proposal that Edexcel made.

Q107 Mr Ward: When you asked Edexcel to bring them into line, did that include their already published January results?
Glenys Stacey: It is not for Ofqual, at that stage, to advise and state what we think they should do. Obviously, we had discussions with them throughout awarding as to the issues. You will see we did not write to them telling them to change grade boundaries in June. We wrote to them telling them they had a requirement to bring the qualification in, rather than at 7% above last year’s results.

Q108 Mr Ward: On the basis that they had already published the January results, and we have more or less agreed that they were flawed from this analysis that was carried out, in your view would it have been inappropriate for Edexcel to bring their results into line using the January results as well?
Glenys Stacey: I do know that Edexcel had considered and discussed with us before the possibility of reopening the January grade boundaries; it was an uncomfortable prospect and, I think, fairly quickly discounted at the time, because those results are out there, established results and students had relied on them as published results when they were building up the rest of their qualification.

Q109 Mr Ward: So if they did not do that, your understanding is that they brought them into line with what we have already agreed to be flawed January results?
Glenys Stacey: No, they did not quite bring them into line with what I would say were generous but right-at-the-time results.

Q110 Mr Ward: Did they claw back the over-generous January marks through adjustments to the June boundaries?
Glenys Stacey: I think I understand what your concern is there: that there was some over-harsh approach to the June boundaries. The position is that if they had clawed back, as you say, to get to truly comparable outcomes, they would have had to make more significant changes to the grade boundaries than they made. They did not do that. They did not come back to us saying they were going to do that. They came back to us saying that they were going to make changes, yes, but they were not of that ilk. They made a judgment; we looked at that; we thought it to be right. An option for us would have been to go back and direct them to make more radical change or, indeed, to reopen the January boundaries. We did not think that to be right.

Q111 Chair: We have limited time and a lot to cover: WJEC as succinctly as you can manage, please.
Glenys Stacey: Right. Well, as I have said before we are applying comparable outcomes across A levels, AS and 47 or so GCSEs. Following the meeting with exam boards to review the provisional A level results, we (the three regulators, acting jointly) challenged WJEC’s results in 33 of the 36 subjects they awarded, and we focused on ten subjects in particular. Following the meeting with exam boards to review GCSEs, we challenged WJEC’s provisional results in GCSE English, and English Language. It is our job to challenge those results and we do.

Q112 Chair: Is this raw, year-on-year comparison, because you do not have Key Stage 2 data for the Welsh students certainly, or are we talking about English students who took WJEC? I might be getting myself confused; I think I am.
Glenys Stacey: WJEC is based in Wales, but it markets its qualifications across the border and, indeed, in English the majority of its students are in English schools and centres, so it is an added complication, but I am talking about all of the A levels and GCSEs that WJEC were providing. Again it is our job to challenge those results and of course, we did so, but we particularly chose to challenge those qualifications where WJEC had a good sample size, if you like, from the English schools and centres and that is what we did. I forget how many it was, but it was six or eight, wasn’t it?
Cath Jadhav: That was for A level; we challenged some A level, and we challenged GCSE English because we had a substantial subset of the entry with Key Stage 2 predictions.

Q113 Chair: The regulator in Wales, the Welsh Government, has come to a very different conclusion from you and has recommended re-grading. Why have they come to such a different conclusion from you?
Glenys Stacey: Our obligations are to regulate to maintain standards in England. That is our statutory obligation. Things are different in different countries, as you know.
Amanda Spielman: I think there is political difficulty in Wales. In what we are seeing, there is a clear divergence in performance between English and Welsh candidates. If English candidates are where we think they are based on our work, the implication is that Welsh candidate performance is not improving. This is a very difficult conclusion for the Welsh to accept politically, hence what we saw yesterday.

Q114 Chair: So as they cannot get results up by improved performance, they are simply inflating results. Is that the allegation?
Glenys Stacey: I think it is useful to reflect on. We had several discussions with WJEC and with the Welsh regulator to try to get to the right position in relation to results in English. It was particularly difficult, because we were able to break the results down by centre or school location, and the results for English students were significantly better than the results for students based in Wales. That is a particularly difficult problem for the regulators and we were very keen to
make sure that we got to some sort of acceptable common standard, bearing in mind that these students are competing with students who are taking the qualification with other exam boards. So, yes, we did write to WJEC, Chair, to answer your question; we wrote a two-regulator letter to WJEC to ask for proposals. They put a number of options to us, to the regulators, as to what could be done. We spoke again with our fellow regulator in Wales and agreed what was, in effect, the softest option, which was to move, by one point, two or three boundaries. WJEC had put up four options, I think, which were more significant as you went up the range. The other options were not acceptable to the Welsh regulator. We took the first option.

Q115 Chair: I think the Welsh review has recommended that there should be a re-grading of WJEC candidates in England. Sorry—right, I have been misinformed. I think that is in Wales.

Glenys Stacey: I think the review said it would be possible in England as well, Chair.

Q116 Chair: Yes, so do you have a view on the re-grading that they recommend for WJEC candidates, wherever it is they lie?

Glenys Stacey: You will understand that this report came in just yesterday and we have not had an opportunity to review it at length, and we would very much like to do that.

Chair: Thank you very much.

Q117 Craig Whittaker: One final question from me. I know we have hammered this to death, but you have already said that the January awards were generous. Can you give us the evidence to support why they were generous? Everybody is saying they were generous, but nobody has given us any evidence yet as to why they were.

Cath Jadhav: I think the evidence comes from having those subject-level predictions when the summer awards were taking place and reflecting back, so it is the benefit of hindsight. Had the awarding committee had that data when they were setting boundaries in the earlier series, they would probably have set higher boundary marks.

Glenys Stacey: In addition, we met regularly with exam boards in English, on a technical level, if you like, and exceptionally, because we knew there were risks with these qualifications, we did ask them to report to us formally after the January 2012 series of assessments. We would not normally do that, but we did this time. We asked them to report any difficulties they had experienced at the time. They told us that they had been cautious in making unit level awards for English. They reported that they had concerns about the quality of controlled assessment marking by teachers, even though they had been putting more effort into giving schools detailed feedback and advice to improve on the quality of their marking. So we were getting that feedback from the January series, if that helps.

Q118 Craig Whittaker: I am still struggling to understand what the physical evidence is to suggest that they were highly inflated.

Glenys Stacey: I will ask Cath again, but I think the issue comes down to the nature of the subject—English—where it is very judgment-based. It also comes down to the specific aspects of achievement that examiners were trying to assess and evaluate. If you are looking at speaking and listening skills, they are notoriously difficult to assess. It is illusive; it is difficult to grasp this, but there were a number of things in January about the subject—about what you are trying to assess, about the dearth of data, about the lack of comparable things from the past that you could truly absolutely rely on—that made those judgments simply very difficult. That is the nature of judgment. There are a number of places in the whole system of awarding and getting to the final outcome where judgment is applied: it is applied at examiner level, at exam board level and certainly at regulator level.

Q119 Craig Whittaker: So what you are saying is there is no specific thing that you could assess, with hindsight, that says, “The reason January was highly inflated was because of A, B, C and D.”

Glenys Stacey: No. I think we can say there is a combination of factors that contributed, but I do not think we would say it was highly inflated either. No.

Q120 Craig Whittaker: Okay. But we generally accept that January was of an easier grade than June?

Glenys Stacey: This is quite a difficult thing to get across, but we were looking very carefully at it; examiners were looking very carefully at it. They were exercising their best judgment at the time. Indeed, some reported they thought they were being harsh. We had subject experts at many of these meetings; they could not see that the standard was not right.

Q121 Chair: Why not? We are still struggling with that, because even if you take a module in which you will not get the whole picture until the end, you then analyse the results and you compare that with Key Stage 2 results and whatever other comparative performance you have that you use. Wouldn’t it have flagged up? I am still struggling to understand why it was not possible. Let me try a different question: seeing where we are today, sitting here, looking back—and hindsight is a wonderful thing; politicians are great at suggesting everyone should have it apart from them—could you have done things differently? Were there techniques, as we are sitting here today, that you wish you had applied or you had had more resource or people to do? Were there things that you could have done that could have given you that insight, or was it literally impossible in any conceivable scenario for you to have been able to predict that the January boundaries were too generous?

Glenys Stacey: I am certainly not aware of any techniques that we could have applied to put more certainty, more assurance, on the judgments that were being made in the best possible way at the time. That
might reflect back and you might ask questions then about the nature of the qualification, but, at the end of the day, speaking and listening is in the National Curriculum and it has to be assessed.

Q122 Chair: If you are right that it is impossible, then it does take me back. I looked up the passing of the legislation, and I think it was in Bill Committee back in 2009 that the then Government basically kept being asked by, ironically, the new Schools Minister, “In order to keep standards constant, does the Minister expect there to be a change in the borderline of grades so that the percentage of students in each cohort getting A, B and C grades remains roughly the same?”

He was asking about grade inflation and saying that the new exams, modular and assessed, would lead to inflation. The Minister very conveniently said, again and again, “That will be the up to Ofqual. They are the independent regulator.” So it was predicted and you are on the receiving end.

Amanda Spielman: If I can come in here, there is a really important piece to get across. It has been really hard and very upsetting listening to schools talking about their disappointment this year, and it is clear that so many schools were expecting much more from the new specs. But whilst some schools are good at predicting, it is clear that a lot are not. In the last year of a very long-established specification, the schools that were using AQA, which is much the largest board, predicted 77% would get a C or better in the written paper; 65% did, so school estimates were 12% higher than the outcome. You might have thought that this year, in the first year of a completely different exam, schools would be a bit more cautious about their predictions and have brought them down a bit.

Q123 Alex Cunningham: So it is the schools’ fault, is it?

Amanda Spielman: This is not saying it is the schools’ fault.

Q124 Alex Cunningham: That is what you have just said.

Amanda Spielman: No, it is not. I am saying that school expectations are in a very different place. We would have expected, in the face of a new specification, to see schools being a bit more cautious. What we actually saw, based again on the data collected by AQA, was that school expectations for Cs went up by 2% this year. They went up to 79%; that is 14% more than achieved a C last year in the same paper. So it seems as though there was an increased level of expectation in aggregate, and we truly do not understand why school expectations were so far ahead of what anybody can realistically have expected to happen.

Q125 Bill Esterson: You get some years where a cohort’s potential results are significantly out of step, either up or down. That surely is the answer. Where your director of standards wrote to Edexcel about this 8% rise, surely that is the point. We have this linear increase in exam results, but that does not reflect the reality of the potential of each cohort. Surely that is part of the issue here.

Glenys Stacey: I can assure you we do see changes in cohort, and the approach that we adopt does allow for changes in achievement as a result. I think in GCSE economics this year, for example, we saw very significant changes in outcomes, positive ones, which we could understand. I think the issue here is in a subject like English, an established subject where the National Curriculum was not changing—qualifications were, or the measurement tool, if you like, but the Curriculum was not—it would have been remarkable indeed if, within that one year, student achievement had shot up by 12% or, indeed, 15%. It would have been very hard to see that over one year.

Q126 Bill Esterson: What about the 8% that you wrote to Edexcel about?

Glenys Stacey: It is the same thing. It is the same thing, because if you think of it, achievement tends to stay relatively steady over time.

Q127 Bill Esterson: Over a number of years it does, yes, but it can go well outside of that linear progression.

Amanda Spielman: At school level, performance can change quite significantly from year to year. At national level, performance changes only very slowly. If you are measuring a glacier and put in a new meter, and the meter tells you that the rate at which the glacier is moving has quintupled in a year, then your first reaction is to check the new meter you have put in, not to say, “The glacier has speeded up five times.”

Q128 Siobhain McDonagh: Our previous guests, whatever we are supposed to call you, brought us back to the fact that it is individual children and individual schools that we are talking about rather than a very advanced degree in statistics. I just want to give you the results from not an academy but a very well established Catholic school in my constituency, in the borough: Wimbledon College, a boys’ school. They normally expect, every year, roundabout, the students to get the same level of grade in literature as language. This year, 73% of boys got five A to C passes in English literature; 32% got five A to Cs in language. In the combined paper that they did, 91% of boys got the A to C grades in the language paper and 88% in the literature. Given that there are years and years of statistical history, years and years of knowledge, that would seem a particularly spectacular result.

Glenys Stacey: If I can just deal with that and say straight away that we need to do more work on the most extreme variances. We do not yet understand them fully. You will see from the report we put in yesterday that we are working very hard to understand what is underlying all of this, but we do need to understand more. But we have to bear in mind that, overall, achievement went down by 1.5% and that was in line with expectations. In fact, underlying that we think there is a slight increase when we look closely at the cohort data. For every school, if you like, that has had a significant shortfall as against expectations, there are others that have had much better results and, of course, we are not hearing so much from those. But I can give you my assurance that, for our next steps, we are continuing to work with ASCL, particularly,
and NAHT and others, and we do want to understand more about why there are these particular variances. We know there are some contributors that might be relevant and different school by school. For example, the routing through, the approach to controlled assessment, the mix of units; these are all very relevant things that we need to understand better.

Q129 Chair: Can you give us any insight? It came out in the first session—all these technical challenges and difficulties, including enhanced expectation, and then this huge variability. It would appear that the key ingredient in what was already a difficult, challenging environment was the fact that there was massive variability. You have outstanding heads with a long history of departments getting consistent predictions and good results suddenly falling off a cliff. It is not surprising that they are going to appear in front of us and write letters and be in local newspapers screaming from the rooftops about something that they just cannot begin to understand, when they have English heads of department who have given every single effort coming in to offer their resignation when there is this inexplicable, for them, collapse in performance in one of the most important qualifications.

Glenys Stacey: Yes, okay. First of all, we are looking to identify nationally the rate of variability. As usual, it is not straightforward. We know there has been variability here. We do not know yet that nationally that variability is particularly out of sync with what we would expect. That is the first thing.

Q130 Chair: I thought your interim report had said that the variability was normally—I might be getting the numbers wrong—about 8% and it was double that this time. Am I wrong?

Cath Jadhav: What we do not know is whether that variability is naturally greater when specifications change, so what we need to do is some more work looking at change situations.

Glenys Stacey: We need to understand how different this is at a national level when qualifications change, and we need a bit of time to do that.

Q131 Chair: Given the introduction of a new exam and given the fact that every year schools are predicting 12%, 14% more Cs and above than they get, and that there is variability anyway between them—so there is always a bunch of outstanding schools that suddenly fall off a cliff—are you saying it is just the collection of all those who do badly in any one year given a reason, once it is in the national press, to believe it is not their fault who are coming out, or not?

Glenys Stacey: Not at all. If we look at it, stand back and ask what the reasons could be here, we know that other qualifications are modular, so it is not necessarily simply the modular factor. We know that other GCSEs have controlled assessment and it came in for the first time for these qualifications. We think there is something about the fact that, for this English suite, 60% of it was controlled assessment. So we are very interested in how schools have approached that and whether there is something to learn from that.

Q132 Chair: Does that mean, in crystal clear terms, boosting marks—‘over-marking’ as you call it?

Glenys Stacey: Not necessarily. There are some suggestions from exam boards that they were concerned about over-marking, but the fact is this is also a completely different way of assessing for schools. It is different from coursework. It is absolutely different, and some of the reports that we have had back from awarding bodies are that some schools were finding that difficult. They were unclear about the rule-setting. So there is something to explore there definitely.

Q133 Ian Mearns: We heard in the previous evidence session that there has been some discourse between you and the awarding bodies about potential over-marking. Why was that never reported back to the schools that are dealing with this and trying to moderate the things for themselves, so that it would have an impact on the learning and the achievement of the children? That is the bottom line here.

Glenys Stacey: It is the bottom line, I agree, and we need to look more at that. We do know that those who were moderating controlled assessment did have conversations when they were returning scripts to schools, but we would like to know more about that, particularly in those schools that are most adversely affected.

There are a couple of other issues, though, I think, that are looking as though they are relevant here as well. Some of the schools that are seemingly most adversely affected—as far as we know at the moment, but we have more work to do—may have a particular student mix and a particular bunching up of students at the C/D boundary level. So they may have, because of their intake, a greater proportion of students who are trying very hard to get across that boundary. So the student mix in each type of institution is really very important.

Q134 Ian Mearns: I put it to you that, if the awarding bodies knew that there was some potential for grade inflation in terms of the moderation being done within schools and they did not report it back to the schools, so the schools could not do anything about it in time for it to affect the grades that were achieved by the children, that is a massive problem. That is a massive problem and, as Ofqual, you have to do something about that with those awarding bodies to make sure that never happens again, but also to make sure something is done about what has happened to the grades of the individual children who have been affected by this.

Glenys Stacey: There are several things there. One is I think one of the lessons learnt here is that communications could have been better.

Q135 Ian Mearns: I think that is an understatement, Glenys.

Glenys Stacey: Well, we know that awarding bodies put the usual health caveats, health markings, on exam board grades from January. We know that they were communicating with individual schools when they were engaging on moderation, but not everything is moderated. It is not clear that enough was done about
getting enough out when it became apparent that there were these problems. I shall come back to the individual student in a moment. Just to complete answering the last question about the particular factors, I have mentioned the particular type of candidature or cohort per school that we need to know more about. The last thing, which I do not think is insignificant, is the routes through these qualifications—the route effect, if you like. A good number of schools were taking a modular qualification in a linear fashion, and we just need to know more about how that correlates to those schools that are most significantly affected by these results. When it comes to individual students, I think you are thinking there about the fair thing for students. Of course, maintaining standards is ultimately about fairness for all students over time. There is no easy answer. There are competing fairness considerations here. We know, for example, that about 95% of the AQA candidates, the majority provider, took their controlled assessment units in June and not January. We also know that, as it happens, one in four students taking the foundation tier paper in June took it as part of their first year of study. So they are carrying a grade through to next year and, of course, they will be competing with other students next year.

We have reflected on fairness and what is the right thing to do very profoundly, understanding the effect that this has had on schools and, of course, on students. We know that awarding in June was regular and proper and, strictly speaking, unpalatable though it is, one might say that the unfairness is in January, but we do not think it right to revisit or intervene in the January results, as you know. Students have already got their grades there and relied on them. We know more about how that correlates to those schools.

Q136 Alex Cunningham: I am sure you are in an uncomfortable position, but this is about the future of young people who sat their particular exam this year, who might not have achieved the A they need for their future university prospects or the C that they need in order to go forward to their next studies. You are acknowledging it is not fair; some people got lucky and it was tough on the rest.

Glenys Stacey: I am not saying it was tough on the rest, because we have had a very careful look at June awarding, and June awarding was right. From the outcomes overall, a 1.5% decrease, we know that, when we strip out cohort changes, that is quite generous.
Q143 Alex Cunningham: You think you would have. Would you have done that? Amanda Spielman: It would have to depend on the arguments that were presented for maintaining that position.

Q144 Alex Cunningham: They have already presented the arguments to you, haven’t they? Amanda Spielman: If they had come back with further arguments as to why they were justified, we would have had to have considered those. It is impossible to say one way or the other.

Alex Cunningham: So they just threw in the towel after you told them that they must downgrade these young people.

Q145 Ian Mearns: Do you think it is possible also that AQA and others might not have bothered to have an argument with you because they knew what was coming? Glenys Stacey: Not at all, not at all. I do not think the relationship is quite like that. If I can come back to Edexcel, we had not seen any compelling evidence that persuaded us or, indeed, anybody else, including the other regulators and other exam boards, that their preliminary results could be sustained. They were out of line and there was no clear understanding or argument put to us as to why those results were different.

Q146 Alex Cunningham: I think you have made that quite clear. Sadly, we are not going to have the chance to speak to Edexcel today, but I hope we will extend the Inquiry in order to do so. If you effectively forced them to ensure that fewer students passed at the higher levels, does that mean the credibility and the professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not professional standing of Edexcel is, at best, compromised and, at worst, totally shot? They are not.

Q147 Siobhain McDonagh: We have already covered some of the issues of what sorts of schools and colleges were affected. Looking in my own area, there is no pattern at all to it. We have roughly the same number of students passing five A to Cs English and maths, with huge drops from the boys’ Catholic schools and increases from the girls’ Catholic schools. We have one academy doing well, and one much more well established, Harris Academy, doing less well. Have you identified what sorts of schools or institutions it is that have had these particular issues? Glenys Stacey: We are beginning to do so, yes. When you are looking at these differences and when we are looking at them as well, we need to understand whether they are differences compared with achievement last year or achievement the year before, bearing in mind last year was an ending spec, which is relevant, or whether it is differences against predictions that were made by schools. I do think that is relevant. We have looked across the types of schools and colleges that exam boards deal with to see a particular pattern. Achievement year on year is matched very closely in academies and in secondary schools, but within academies and secondary schools there will be these differentials that we need to understand. We did see a particular effect in further education colleges, which we want to understand more as well.

Q148 Siobhain McDonagh: What happened in the FE colleges? Glenys Stacey: Looking at the data overall, they saw particular falls as a group of institutions compared with, say, secondary schools or academies. There will be some likely reasons for that. For example, those in further education colleges taking English may well be taking it for a second time. They may well not have been able to get the grade they wanted when they were at secondary school, so that will be one of the factors, certainly.

Q149 Chair: That would not change year by year. It would not explain why there was a sudden drop this year over last year, unless there had been a change in the cohort.

Amanda Spielman: One hypothesis is that, post the Woolf recommendations, more colleges are instituting a policy of having more people do re-sits if they have not already got a C when they arrive. We have a number of hypotheses we have to explore that we simply have not been able to cover yet.

Q150 Siobhain McDonagh: Do you have any idea what group or type of student was affected? We are concentrating on the D/C band, but there is some evidence there was some problem higher up the bands as well. Glenys Stacey: We know the spread of grades, obviously, that were achieved. We can see that there were particular pressures, I think, on the C/D grade boundary.

Q151 Siobhain McDonagh: Are they boys or girls? Glenys Stacey: We have not done a gender analysis.

Q152 Damian Hinds: To pick up where Siobhain left off, is there any difference between foundation and higher?

Cath Jadhav: Obviously there are differences in the grade profiles for those groups of students.

Q153 Chair: Can you tell us a little bit about it, because the Secretary of State a few weeks ago said—it came as news to lots of people, including, shamefully, perhaps, the Chair of the Education Select Committee—that significant numbers of children were on this foundation tier, which meant they could not get more than a C. It looked remarkably like a CSE and looked remarkably like a two-tier system had been in our system all along. What percentage of people take each one in English, seeing we are looking at that, and can you let us know whether the
people taking the foundation tier are peculiarly affected?

**Glenys Stacey:** I think the first thing to say is that things changed this year, because traditionally the foundation tier was for, if you like, the less able student and you could not get more than a C grade if you went down that route. The rule has changed for these qualifications and it is possible to get an A grade on foundation tier.

**Q154 Chair:** As of 2012. So in 2011 the maximum was C, like the old CSE, but in 2012 you can get an A. Is that right?

**Cath Jadhav:** As always, it is slightly more complex than that. It is to do with the proportion of the assessment within the English suite that is un-tiered. The controlled assessment is un-tiered, so a candidate who might be entering the foundation written paper could also, in theory, score maximum marks on the controlled assessment, which, when it is aggregated, means they could, in theory, achieve an A.

**Q155 Damian Hinds:** Is that overall?

**Cath Jadhav:** Overall.

**Glenys Stacey:** It would be unusual to do it, but it could be done.

**Damian Hinds:** Gosh.

**Glenys Stacey:** I know; it gets quite complicated. There were, of course, these common units that students would be taking early on. For exam boards, that is particularly difficult because they do not know, at that stage, what the ultimate destination is in terms of the tier of qualification for the student. It is another complicating factor.

**Q156 Damian Hinds:** There are different questions around the impact on different schools, as Siobhain explored very effectively, and then what happened at the aggregate level. You have mentioned a couple of times the changes in the cohort, and we know there were 23,000 fewer pupils from independent, selective schools and there were perhaps more pupils doing re-sits—perhaps. Do you have such a thing as a like-for-like analysis? Obviously, you cannot say like for like, because they are different children each year, but if you say among maintained, non-selective schools, last year X% of pupils got A* to C and this year it was Y%.

**Glenys Stacey:** This is matched data really, isn’t it?

**Cath Jadhav:** We know all things are never equal, but the matched candidates, the candidates with their Key Stage 2 scores, give us, as far as possible, that like-for-like comparison, and that was up by between 1% and 2%.

**Q157 Damian Hinds:** For the avoidance of doubt, if you strip out, insofar as you can, the changes in the mix of pupils, there were, nationally, more children who scored grade C or above at GCSE English in 2012 than in 2011.

**Glenys Stacey:** Yes, that is right.

**Q158 Pat Glass:** I have sat very quietly all morning and the Chair is now cutting me off because we are running over time, so I know that, in fairness, he will give me more than my fair share with the Secretary of State tomorrow to make up for it.

**Damian Hinds:** Tactics.

**Pat Glass:** I have two questions and they just require yes or no answers. You said, right at the beginning of your supplementary memorandum to us, that there was no political interference. We accept that there is no phone call between you and the Secretary of State, but if, as a Committee, we decide to extend this Inquiry—and I think we should—would you be prepared to publish copies of correspondence, emails, text messages and phone calls between your staff who are involved in this and senior staff at the DfE who are involved in this and special advisers, ministers’ spads?

**Glenys Stacey:** Absolutely; we would have nothing to hide.

**Q159 Pat Glass:** Thank you; that is a yes answer. Secondly, this question is about the people who are affected by this. This is not just numbers on a piece of paper. There is substantial evidence of the difference in life chances between children who get five A to Cs and those who do not, and those who get Cs and Ds, and it is not just about their academic qualifications. Children who get five A to Cs are less likely to get divorced, are less likely to get cancer, are less likely to end up in prison or homeless, and a whole range of other things. So this is about what is going to happen to these young people for the rest of their lives. Given that, and given what we have heard about what the Welsh regulator is doing, and given that we know that, as this is largely the C/D borderline, it is likely to affect fewer children in grammar schools and independent schools—it is going to be kids in comprehensives, on the foundation level, on those C/D boundaries—are you not prepared to look again at the issue of rebounding, given the long-term impact of this on children’s lives?

**Glenys Stacey:** I think the problem we face with that, given our statutory framework and obligations and objectives, is that, if we apply the January boundaries to the June students, we would have inflation of, we think, about 5% or 6%; we have not worked it out objectively, is that, if we apply the January boundaries to the June students, we would have inflation of, we think, about 5% or 6%; we have not worked it out entirely.

**Q160 Pat Glass:** These children worked just as hard, and their teachers worked just as hard. This is not their fault.

**Glenys Stacey:** But the results would be inflated.

**Q161 Pat Glass:** So the answer is no.

**Glenys Stacey:** What we have seen so far, with nothing to doubt it, and we do not expect to doubt it, is that the June boundary setting occurred properly. That does, as I say, leave us all in a very uncomfortable position. We have thought, as I said, carefully about fairness and we keep on thinking about it. We think that the right thing to do is to offer a re-sit opportunity for those students, and we have been working very hard with exam boards to make sure we can do that quickly.

**Pat Glass:** Some of them have lost the will to do that now.
Q162 Chair: I must bring this to a close, but in terms of conducting an inquiry into this, is that something that you can properly do when you are such a key player in it? Aren’t you effectively investigating yourself?  
Glenys Stacey: We have reacted very promptly to concerns expressed to us by schools and colleges, and I know that they recognise that. We have been very open with them about the data information that we are collecting and the conclusions that we are reaching. We really want to do more to get to the root of this. It is our job to make sure that standards are right, and so we want to do that, but of course we do not object if there is any other sort of inquiry—not at all.

Q163 Ian Mearns: Chair, I think anyone, an objective observer watching this from outside, could draw the conclusion that there are lots of unknowns, there are things that you are thinking about doing, and there are some things that you are beginning to do. Some objective observer could draw the conclusion there is a lack of urgency in Ofqual in terms of dealing with this. How would you respond to that? As Pat has pointed out, this is having an impact on individual children’s lives.

Glenys Stacey: I would most certainly and robustly refute that. We were notified, I think, on the Bank Holiday of concerns. We responded immediately. We reported in our initial report inside of a week. There is lots of midnight oil being burnt at Ofqual and we are giving this our top priority, I assure you. Indeed, no one, until now, has suggested otherwise. All of the resource that I can devote to this is being devoted to it.

Chair: Thank you very much for giving evidence to us this morning.
Tuesday 12 March 2013

Members present:
Mr Graham Stuart (Chair)

Neil Carmichael
Alex Cunningham
Siobhain McDonagh

Ian Mearns
Mr David Ward

Examination of Witnesses

Witnesses: Andrew Hall, Chief Executive Officer, AQA, and Ziggy Liaquat, Managing Director, Edexcel, gave evidence.

Q164 Chair: Good morning, and welcome to this session of the Education Select Committee, looking into the GCSE English exam results in 2012. If you are both happy, we will maintain our informal style, which has the great convenience of meaning we only use first names, and save all pronunciation difficulties. Were last year’s English GCSEs merely a furore, or were they a fiasco, Andrew?

Andrew Hall: What happened last summer was a consequence of something that had been put in a system a long while ago. If you look at the fundamental cause of what occurred, you had the interaction of accountability measures, which have become ever more a focus for teachers and schools over time, with a qualification at the time that was not capable of standing the weight of it, with something like 60% controlled assessment. To a number of people I have described it as taking magnesium and water, both of which are fairly stable components on their own. Put them together and you get an explosion, and that is what we saw last summer.

Q165 Chair: So the primary problem last summer was effectively the cheating within the system, was it?

Andrew Hall: No, it is a combination of accountability measures, those very fine judgments teachers have made, and having a qualification that just cannot stand the weight of it.

Q166 Chair: Yes, but when you say accountability measures, what you mean is that, such was the drive to get the results, in too many cases it was a bit too tempting for teachers, who had 60% of the marks to allocate and their own careers resting on the results of those allocations, to give out marks that were not deserved. Is that not what you are saying?

Andrew Hall: When you talk to teachers day in, day out, as we all do, they make it very clear that they do feel the pressures of the accountability measures, and one of the most common statements you get is: “We, of course, do this properly, but we’re quite nervous about the school down the road.” You do wonder how many roads you go down before you come back to yourself.

Q167 Chair: Was it a furore or a fiasco, Ziggy?

Ziggy Liaquat: I am really clear that we occupy a privileged position. We are dealing with a precious thing: the life chances of young people. The court and the Ofqual investigation have decided that we carried out our duties professionally. However, we are clear that there was real anxiety and distress for students and parents, and I would like to apologise for that. There are improvements that could be made. Particularly to your point, we could have avoided the situation to some degree by communicating more effectively earlier on, making it clearer that grade boundaries are subject to movement. There is a need for us to really step up our game going forward, particularly as we embark on this substantial reform that is heading our way.

Q168 Chair: So is stepping up the game a technical improvement or primarily one of communication?

Ziggy Liaquat: One of the things this challenge has revealed is that the awarding process is complicated, and we have not done a good enough job in explaining how it works in a very simple way. We thought we had explained it well, but it clearly had not landed, and those grade boundary movements were a surprise. For the January series that we have just issued results for, we called every single one of our customers and said, “Look, we’re not issuing grade boundaries for this series,” but we still got calls asking where the grades were. The communication challenge is a significant one.

Chair: Thank you. We have three panels this morning, and a lot to get through. The Committee is a bit denuded by the fact that so many members of this Committee are on the Children and Families Bill Committee, so they are scrutinising legislation as we speak. We have a lot to get through, so I ask for both questions and answers to be as short and succinct as possible. Thank you very much.

Q169 Mr Ward: We have already touched on some of the elements of the first couple of questions, but specifically, for your own organisations, there was clearly a loss of confidence in the system. What specific measures are your own organisations taking to try to restore some of that confidence?

Andrew Hall: There are a number of things I would look at. I will try to be brief on this. Clearly, we along with others recognise that the risk of this particular qualification awarding early units was there. We wrote to the regulator. I was not there at the time; it was something like three and a half years ago. We were not anything like forceful enough, I think, in making our point heard. People have not accused me of not making my point heard, generally, so that is a lesson we have learned. If we have got issues with the design of the qualification, we must be very clear about it.
Much of this issue was around teachers marking to the best of their judgment within a tolerance. Those tolerances were too wide; they were 6%. We have significantly reduced those, and that has been important. The other thing is our feedback to schools; it was previously very much a free text. I have probably read more moderation reports over the last few months than I would care to. They were free texts, and that was designed to be helpful. It had things that had to be covered, but it was possible to interpret those in different ways. You would talk very positively about the administration, and people would maybe interpret that as a positive remark about the marking generally, so we can now go on to a slightly tick-box, less-text approach, but specifically addressing the issues.

Looking forward, we have to protect the next groups of students. I wrote to Ofqual in December, suggesting there should be a radical change to the structure of the GCSE as soon as possible, removing the speaking and listening component, and making that an endorsement on the certificate. It is about having the courage to step up and do those things, and I have been pleased that Ofqual and the other boards seem to think that it is quite a good idea if we do that. So it is that type of approach.

Ziggy Liaquat: So as not to repeat, I will add to what Andrew has said. I will come back to the point around really engaging students, teachers and parents in how the system works. That is going to be critically important as we move to linear, because of course, although modular assessment goes away, there is a residual risk that there are expectations around grade boundaries year on year in a linear system. We should be relentless in communicating the fact that these grade boundaries can move.

The other point I would make is that it would be a real shame if this challenge, this issue that we have faced, is a deterrent to providing appropriate assessment for qualifications. For all of us, the last thing in the world we would want to see is a move to a system where everything is externally assessed in a very safe way, in a kind of multiple-choice way, because we have to offer the right assessments for qualifications that stand up to the assessments that we see in the rest of the world. That is critically important. Our jobs would be a lot easier if we did not do that, but it is right for us to assess the skills and knowledge that employers and higher education need. It is important that we look forward and use this challenge in a positive way.

Q170 Mr Ward: You talk about the system, but do you think your own organisations have suffered a loss of confidence, as organisations?

Andrew Hall: To be clear, there were a number of people that were genuinely upset, and there would be no doubt about that. Yes, there was a loss of trust, and we have to build that back. Part of building it back is being very open and transparent about the issues as we see them. That does not mean just giving in to everything that everybody wants. That does not build trust and confidence over time. You have to accept there was genuine frustration.

One of the most encouraging things, in a way, is that whilst there have been a number of investigations, reports and judicial reviews, the dialogue with teachers has continued. There has been a greater understanding, day in, day out, and the dialogue and the involvement with the leadership of teacher associations and trade unions has continued. We have started that process. Fundamentally, we have to have a system that people trust.

Q171 Mr Ward: I am jumping all over a little bit here, but, Ziggy, you mentioned the issue of not revealing grades for the January assessments. Will that in itself make a difference in terms of avoiding some of the problems?

Ziggy Liaquat: I think so. Fundamental to the challenge we saw was the expectation that was set in January, and certainly for our own organisation, we had to award based on 700 candidates. We did the best job we could in awarding those grades in January, but of course there was substantial movement in June. So moving to a system where we run the awarding or grading of those results at one time gives us more data, more information and more candidate work to look at, to make sure that we get it absolutely right.

Q172 Mr Ward: Moderation seems to have gone wrong here. As I was saying before you arrived, when I was sitting on exam boards, if you are shifting marks at the end, then something has gone wrong in moderation. That clearly must have happened.

Ziggy Liaquat: On the nature of awarding, the central point for me is that the primary focus has to be the quality of the work that you see in front of you; you compare series upon series, year upon year, to make decisions around grade boundaries, and I think that is right. There are improvements to be made in any system, and we are working hard with Ofqual to make sure that we learn the lessons from the challenges that we have faced over the last six months. The spirit of looking forward and looking for improvements in any way we can is consistent across all of the organisations here, and in the Department and Ofqual.

Q173 Mr Ward: More of a technical question now, on the use of comparable outcomes, and in particular the use of Key Stage 2 data as a way of trying to get some consistency year to year. Do you think Ofqual was right to insist on its comparable outcomes approach?

Andrew Hall: From our perspective, AQA has used past performance data for a number of years. This is not something new to us as an organisation. It is very clear to me that you need some sort of guidance as to: “What is there? What is the strength of data now?” Key Stage 2 data is the best available for GCSE.

Mr Ward: It is a long way away.

Andrew Hall: Yes. People will tell you—and they are right—that over five years pupils’ performances change. However, it is valid across a national cohort of 600,000 students. If you are trying to use it to assess what you should reasonably expect in a classroom in the middle of Essex, then it is clearly not appropriate, but using it for national guidance is different. The key is in understanding its uses and
limitations, and it is only used as a guide. Examiner judgment is an important part of the overall awarding process. One of the things potentially misunderstood at some stages was that the Key Stage 2 data was generating a quota. That absolutely was not the case, and the best evidence I can give you on that is for AQA's GCSE Mathematics, where we went way outside the statistical guidance. We found something in the mix of our cohort, the nature of the entries, and we submitted a very detailed technical report in the summer to Ofqual, as the regulator, and said, “For this reason our examiner judgment, the other evidence we’ve got, says that comparable outcomes isn’t inappropriate here.” I would happily share that paper with the Committee afterwards if you want. It proves it does happen. If you use it as an absolute limit, then that is clearly not what it is designed to do.

Q174 Mr Ward: Ziggy, do you think it should continue to be used for GCSE and A-Levels?  

Ziggy Liaquat: It is not perfect as a system, and there are improvements. It is as good as any system that I have come across. One of the key points, and potentially one of the misunderstandings publicly, is that somehow an individual’s Key Stage 2 performance caps their particular grade in GCSE. As Andrew said, it is a guide; it is a test to look at the overall performance of the cohort. It is just one of many tests—many of many checks and balances. Like Andrew, for GCSE English in fact, our award was outside of the prediction, and as with AQA, we had good and valid reasons to explain why that was the case. Our award was out of tolerance, and that was okay. Those two examples do demonstrate that it is simply a guide, and that the judgment of senior examiners is key to making those decisions.

Q175 Neil Carmichael: Good morning. With the forthcoming reforms to GCSEs and A-Levels in mind, what are the key lessons for us to take note of following the 2012 GCSE English results?  

Ziggy Liaquat: I will return to one of the points I made earlier. Clearly we have learned technical lessons around tolerance. We will be moving to linear, and we need to really engage and educate teachers, parents and students on how the process works; that is key. Fundamentally we share the ambition of the Secretary of State. It is absolutely right to have a qualifications system that stands up next to the best international studies that we are providing to the Department. We are also in conversation with Ofqual on more technical issues, so none of those things are set in stone yet. What is positive is that engagement is substantial, and we have a lot of data and information to bring to the table, which we are committed to doing.

Q176 Neil Carmichael: So in broad terms you agree with the Government’s general direction of travel on this?  

Ziggy Liaquat: Absolutely. We are still in the middle of the engagement with the Department around content. We have a lot of research from our international studies that we are providing to the Department. We are also in conversation with Ofqual on more technical issues, so none of those things are set in stone yet. What is positive is that engagement is substantial, and we have a lot of data and information to bring to the table, which we are committed to doing.

Q177 Neil Carmichael: Andrew, you mentioned earlier the issues of accountability, and I certainly have some sympathy with that. So my question to you is: what are the lessons for us in terms of the relationship between Government, the regulators and indeed exam boards?  

Andrew Hall: It is one of getting fundamental clarity of purpose on what you are doing. You are right that I refer to accountability. You cannot look to design new qualifications that are going to be used in an accountability system without truly understanding the consequences of that and what will happen in the accountability system. This all needs to be done at the same time. The idea that you can reform a national curriculum, reform qualifications and reform accountability and have them as discrete operations is one I quite openly challenge. You need to have a clear understanding. It is also about understanding the limitations of what assessment can do. It is to some extent an art, not a science. There will always be margins of error in any assessment, as that happens in anything you do, and you need to understand those. However, it is also about making sure that there is progression in the system—that what you are developing at each stage of the curriculum leads into the qualification. Fundamentally, get it aligned. Let us align accountability.

I have always felt uncomfortable with something like 60% teacher assessment in a qualification that is going to be used for an accountability measure. It is quite interesting: if you take French GCSE, which has an almost identical structure to English GCSE, it does not bear the same accountability rate. It is about looking at that, and understanding what is really going to stretch pupils and encourage them to learn. Having reform is appropriate, but it needs to be reform to a point. The one thing we must do, if we are changing the standard in this reform, is be very clear what the standard change is. If the ultimate goal is to make it more stretching and challenging, if I can say that, be clear about how much more stretching and challenging, and do not have it as some hidden message that creeps out two or three years into the system.

Q178 Neil Carmichael: So you want to see more clarity and transparency in the system, don’t you?  

Andrew Hall: Yes, clarity and transparency between the parts of the system. It is impossible, when someone has told you they did not understand
Andrew Hall: a large number of the problems arose because of a number of the interactions that needs to be tackled. Ziggy Liaquat: Can I just add something very briefly? One of the things that certainly struck me, having been to Singapore recently, is coherence. There is a single purpose for education, there is absolute clarity, and every organisation is lined up behind it, so you do not have these different tensions that Andrew talks about. In terms of engagement with the Department and Ofqual, we all accept that expertise in qualification and assessment does not reside in one organisation. It is incredibly important that we do work together in the best interests of learners in that respect.

Q179 Neil Carmichael: You have both touched upon this already, but Ofqual’s analysis was essentially that a large number of the problems arose because of a combination of design flaws in the GCSE qualification for English. Do you tend to agree with that?
Andrew Hall: The fundamental problem is the design, set against the accountability measure. You can have an educational debate as to whether controlled assessment is a good thing or a bad thing. You can have a debate about whether modular assessment is good or bad. We might all have particular views around that, but it is that fundamental interaction.
Chair: That is clear, Andrew. Ziggy, do you agree?
Ziggy Liaquat: There are four points Ofqual made; one was around the pressure accountability brings, and one was around design, and I think that is absolutely right. Of course it is quite natural for teachers to strive for their students to do well, and particularly with English, where it is a headline measure. The proposals around the new accountability system are progressive and well thought through, and it is going to be interesting to see how they play out.

Q180 Neil Carmichael: The Welsh exam board raised issues about the 60% weighting, and it was concerned about the controlled assessment aspect of it. Did you have similar concerns, or were you aware of their concerns?
Ziggy Liaquat: At the stage of design, absolutely. Just to be brief, when we were designing the qualification we asked for there to be 40% controlled assessment in English.

Q181 Chair: It was done in about 2009 or something, wasn’t it?
Ziggy Liaquat: Exactly.
Chair: It was the old QCA, so it pre-dated the current Ofqual.
Ziggy Liaquat: It was the interim Ofqual.

Q182 Ian Mearns: Judging from your answer to the earlier question, which was, “With hindsight, what would you have done differently?” isn’t there an element of “you wouldn’t have started from where you started” about that?
Andrew Hall: Probably, yes. There were some fundamental challenges. The point I made before shows the concern we had at the time: our technical team wrote expressing concerns when this qualification was designed and suggested to the then interim Ofqual—which is, to be clear, not the current regime we have—that it would be wise not to award grades but just to issue marks. The regret I have is that we were not more forceful about that, as I said before.

Q183 Chair: It is good to have that mea culpa from you, but WJEC wrote and said, “This doesn’t make any sense.” Edexcel wrote and said it did not make any sense, and AQA wrote and said it did not make any sense. You guys can chastise yourselves for not being more forceful, but why were your warnings not heeded? Who called the shots on what led to some of the difficulties we saw last year?
Andrew Hall: The legal reality is that the criteria are designed by the regulator. I just want to draw the distinction between the previous regulator and the regulator we have now, Ofqual. At that time there was an interim regulator that the then Secretary of State had pulled out of the QCA and created as a small body. I am on public record quite often saying it did not have sufficient resources to focus on standards, and that is where the issue arose.

Q184 Chair: Was it independent of Government, or was it basically up to the Secretary of State to decide at that stage?
Andrew Hall: I am not sure I am in a position to judge that. I saw it as an independent organisation. I see the current regulator as more independent.
Ziggy Liaquat: We do bear some responsibility. Ultimately, if we had not met the design criteria, which were 60% control assessment, although we were concerned about the controlled assessment aspect of it, did you have similar concerns, or were you aware of their concerns?
Q185 Chair: It is going to be interesting to see how they play out.
Andrew Hall: It was the Ofqual of the moment.

Q186 Ian Mearns: From both of your individual perspectives at Edexcel and AQA, are you both absolutely confident that we are not going to find, at any time in the near or medium-term future, a repeat of what has happened over the last 12 months?
Ziggy Liaquat: I am confident that we have looked hard at the issues and challenges, learned lessons, and we are moving to fixing them. Again, central to this was this expectation that grade boundaries would not move substantially, series on series. We are moving to linear, and we are not issuing grades in January. For English, as long as we continue to relentlessly communicate how this works and what it means for students and teachers, I think that will be okay. With
Andrew Hall: occurred over the last 12 months. to have learned collective lessons from what has capacity, in conjunction with your partners in the field, organisations. Both organisations should have the 

Q187 Ian Mearns: You both work for intelligent organisations. Both organisations should have the capacity, in conjunction with your partners in the field, to have learned collective lessons from what has occurred over the last 12 months.

Andrew Hall: My perspective is that you have this: you review what is in your portfolio, you look, and you see if there are other things that concern you right now. We are going through a process of reform of the qualification base that is there now. The most sensitive thing we are doing is the change in standard on GCSE Science, which we have talked about with this Committee before. I believe we have communicated that change as a group of organisations. It was covered in the media beyond belief at the time it was being created. Do we need to keep reinforcing that? Yes. Is that an area that I have concerns about, in that some schools may not have understood it? It is in the worry mix, of course it is, because you have seen it, so you keep reinforcing the message. Outside that, it is in the process of looking at the GCSE reforms and the A-Level reforms we are going to tackle, to make sure we do not create something again.

Q188 Ian Mearns: From what both of you have said earlier as well, one message you want to get across to everybody out there is: ‘If your exam board is communicating with you as a school, listen very carefully to what they’re saying.’

Andrew Hall: Yes, you are right. Schools receive an immense amount of information, not just from us but from Government and from other agencies—teacher agencies. Actually, you go into schools in different places. We will communicate with exams officers, and we will communicate with heads of department. For standardisation for teacher moderation, a member of staff will be sent. We rely on the communication network in the school as well, and that is one of the things we are certainly saying: “Let’s not just rely on one route into a school.”

Q189 Ian Mearns: Now, in terms of the technicalities of GCSE English, do you think GCSE English in future should continue to have an element of internal assessment for speaking and listening?

Andrew Hall: I wrote to the regulator in December and said, “I don’t think that’s appropriate going forward.” It is an important part of the course, but I believe it should be dealt with through an endorsement, which says you have to teach it; you get a separate recognition on your certificate but it does not count towards the grade. That is where I believe it is, partly because of the challenges of moderating it. It is an ephemeral activity.

Ziggy Liaquat: We definitely need to strengthen the quality arrangements around speaking and listening; that has come through as a clear lesson. You can still do that through internal assessment with strengthened controls, and the key has to be the right kind of assessment for speaking and listening. We just need to maintain that testing and assessing skills in an appropriate way is where our focus should be, and then making sure our controls do exactly that. That has got to be the approach.

Q190 Ian Mearns: Did Edexcel agree with Ofqual’s analysis that teachers over-marked the controlled assessment?

Ziggy Liaquat: We have looked at this, and looked at this very carefully. The thing to bear in mind is that we only work with 10% of the cohort. We have certainly seen inaccurate marking. We have made adjustments to internal assessment decisions both up and down. The thing we have to bear in mind is that Ofqual are looking at the data of the entire cohort, and they have reached the conclusion they have. Although we see inaccuracy of marking in the data we are looking at, it is difficult for us, looking at that data, to make the leap to over-marking. However, we only have 10% of the data to look at.

Q191 Ian Mearns: Yes, but 10% is a pretty good sample, isn’t it? If you do an opinion survey, 10% would be regarded as well within the bounds of discrepancies on either side.

Ziggy Liaquat: It is a substantial number. However, that has to be set against the 100% that Ofqual are looking at. We are in conversation with Ofqual about what we found in our analysis and what they found.

Q192 Ian Mearns: Did you find any evidence at all of significant over-marking?

Ziggy Liaquat: I think we adjusted downwards 8% of marking, and we adjusted up about 5%. So there was inaccurate marking both ways.

Q193 Ian Mearns: Yes, but what I am asking about there specifically is: do you think there was significant over-marking in any individual cases?

Ziggy Liaquat: There was a degree of variety. I would not say there were clear cases of significant over-marking, no.

Q194 Chair: The data shows that outstanding schools tended to do worse in 2012 than in 2011, and in fact the weaker the school, in terms of Ofsted grading, the better it did. You would expect outstanding schools to have better systems of monitoring and maintenance of standards, and is it not the case, therefore, that the weaker and, in some ways, the more desperate the school, the more likely they were to over-mark? Or am I reading too much into it? Are you too reluctant to say it anyway, because these are your customers?

Andrew Hall: No, but if I listen to what a number of head teachers have said to me, they have said similar things, and it is hard, on occasions, to escape that conclusion. After the summer we put a lot of our technical resource into a very detailed investigation looking at mark distributions, and looking at marks needed to get to a particular grade. The thing that has
caused me most grief on a personal basis through this is that we think there was some tendency to over-mark in places. We have said so, and I have explained why we think that was. I can well understand that a very confident school that perhaps did not over-mark may feel that its students were somewhat disadvantaged, and it is hard not to have sympathy for that viewpoint.

Q195 Ian Mearns: Were the 2012 papers that you are talking about there, in terms of over-marking or the potential for over-marking, drastically different from those of previous years?

Andrew Hall: It is the first time we have had a modular approach to this qualification in this format; that was the fundamental difference. In the past there has been coursework—it was coursework, not controlled assessment—that had similar degrees of potential challenge in it, but it was not delivered in a modular system. It was delivered with all the grading coming at the end, and therefore it meant some of the issues that could lead people into temptation were not there.

Q196 Chair: So because of the modular nature, it meant they thought they knew where the boundary was.

Andrew Hall: Correct.

Chair: Where there might be a general temptation—because of accountability, to look good—to mark up slightly, without the firm knowledge of what the additional two marks would do, they were less likely to do it. Once they knew that all they had to do was find two more marks and magically the student would become a C instead of a D, perhaps that was another factor that just contributed to over-marking.

Andrew Hall: That information and knowledge must have been a factor in there somewhere.

Q197 Ian Mearns: Did you pass on any information to Ofqual about what, in essence, your customers were doing?

Andrew Hall: The process of how this works, if I can take a minute on this—

Chair: We do not have it, but go ahead anyway.

Andrew Hall: Moderation is a process. It is not re-marking scripts. It is going and trying to see if the school has followed the standard. You have a tolerance because, particularly on subjects like English, History and Geography, there will be a degree of different judgment between teachers. We now know that 6% was too high. Where someone was outside the tolerance, those marks were adjusted; that is a matter of course. The particular challenge was those at 3%, 4% and 5% within the over-tolerances. We only see that at the time when the award is made. The idea that we have some sort of hidden knowledge halfway through the term of where the grade boundaries are going to be because of the moderation work just is not the fact, so you see that. When we made the award, we adjusted our grade boundaries, as did all the awarding bodies. If this was one awarding body, it might point to something different, but all the awarding bodies did it. In the review meetings we had with the regulator, before the results were issued but after we made our award—which is where, if you like, we are held to account by the regulator—we made specific reference to this point.

Chair: Ziggy might have a different point of view on this.

Ziggy Liaquat: The drivers are there: an incredibly important C grade in English for accountability measures, incredibly important for progress for these people into further education and work, and teachers want their students to do well. The drivers are certainly there. The question I ask myself is: do the data show me evidence of giving marks away to get over a boundary? I am not there based on the sample we have, but I do have to say, again, it is set against the data that Ofqual are looking at, which is 100% of the cohort.

Q198 Chair: Have you seen the WJEC results? In Wales they do not have the same accountability arrangements. There is a chart showing the results of exactly the same exam, and in Wales it is a smooth bell-curve, and the number of people at each mark is exactly where you would expect it to be. Then you see England, and there is a spike at each of the grade boundaries as you go along, which would seem to suggest, unless you are peculiarly different from WJEC, that they provide that snapshot of one world where there is accountability at the boundaries and knowledge, because of modularity, of where those boundaries might be, and one world where it does not matter, and the variation in marks is markedly different. Have you seen that?

Ziggy Liaquat: I am happy to provide the Committee with the data and the analysis that I am talking about. We do see peaks in marks, but way beyond the boundaries eventually set in June. I do not have the data to hand, but I am happy to provide it.

Q199 Ian Mearns: Do you agree with the idea put forward by Ofqual that schools that marked controlled assessment accurately may have seen their students being penalised because the boundaries were raised to address inflation caused by other schools marking generously for their students? Is that an acceptable outcome for an exam system?

Andrew Hall: It is exactly why we should not have 60% controlled assessment in a very high-stakes qualification. We are required to set a national grade boundary for each thing. We cannot set a grade boundary for Sussex and a grade boundary for Surrey, or any particular grouping you like. So absolutely, when you have got that sort of weight—and I know I keep going back to that—you should not have 60% controlled assessment.

Ziggy Liaquat: I will just go back to the point I made earlier, which is that the decisions we make are driven by the quality of work we see in front of us against the grade descriptors. There are other points of data that we consider. Where we ended up in comparing what were two qualifications—there are now three qualifications in English—was outside of the tolerance where we have to explain why we see an increase in the number of students getting to C. That demonstrates the fact that it was not about grade inflation; it was about getting a fair set of grades based on students’ work.
Q200 Siobhain McDonagh: You have already made some reference to trying to tell the regulator about identified problems, but Ofqual has said that as exam boards, you “could have done more to identify and report emerging concerns to the regulators”, and that “monitoring by exam boards could have been stronger and more intelligent”. What do you understand by this, and what will you be improving as a result?

Chair: Who wants to have a go at that first? Ziggy?

Ziggy Liaquat: Thank you. I have said earlier that I think we identified the problem quite early on. We were quite clear that in a modular system grade boundaries are subject to move, particularly as we are compelled to grade with 700 candidates for every series we offer. As you will have seen from the exchange that we had with Ofqual, which is a matter of public record, we had a very robust and passionate discussion about getting to the right outcome. I accept that there is more that we can do to communicate more effectively with schools, teachers and parents, and there is more we can do to make Ofqual aware.

One of the interesting things in this, and one of the things we need to think about going forward, is that by and large the technical expertise resides within the awarding organisations, and we work together with Ofqual to help them understand how these things could turn out and how we need to manage them. So there is certainly a need and a desire to invest more in that kind of collaboration.

Andrew Hall: From my perspective—and we all have different strengths—the thing that caught us most by surprise was the variability between school groups. The overall English outcome, when we looked at it—and bear in mind we look at the overall outcome—moved by something like 1%, which we could understand from a number of things, such as the changing of the cohort. We had not seen the variability between school types when we are setting the grade boundaries, and thereby lies all sorts of risk—to have an early warning system. Would it have fundamentally changed the outcomes this summer? No, because there is certainly a need and a desire to invest more in that kind of collaboration.

Andrew Hall: That is a fair challenge, and I agree with Andrew. It is something that we need to think carefully about certainly for other GCSEs, but particularly as we look to re-develop GCSEs for 2015.

Q201 Siobhain McDonagh: How much is cost a consideration when you are establishing moderation procedures for a qualification?

Ziggy Liaquat: I do not think it is a significant consideration.

Chair: If that is your answer, Ziggy, that is clear and fine, Andrew?

Andrew Hall: No, it is not. It is about effectiveness.

Q202 Siobhain McDonagh: The marking tolerance was tightened from November 2012, reducing it from the previous 6%. Did you welcome this change? Who set the new tolerance, and does it now apply across all subjects and exam boards?

Chair: Nice short answers, please.

Andrew Hall: Yes, we did. We all recommended it to the regulator in one roundtable meeting. We needed to tackle it.

Ziggy Liaquat: Yes, and it does not apply to all qualifications.

Q203 Chair: Which qualifications does it not apply to?

Ziggy Liaquat: It applies to English.

Chair: Right, and it does not apply to the other subjects?

Andrew Hall: No. We are doing a piece of work right now testing where it should be for each different subject.

Q204 Chair: When will that be concluded?

Andrew Hall: I need to come back to you on that, unless someone behind me can tell me.

Ziggy Liaquat: It is one of the technical considerations and conversations, particularly around the reform of GCSEs.

Q205 Siobhain McDonagh: Given all your assessment expertise, why did no one question whether a standard tolerance of 6% was appropriate for such a high-stakes qualification with a higher proportion of controlled assessment?

Andrew Hall: That was because we did not understand sufficiently when we were setting out on this course the impact of accountability measures. That has been a learning journey for awarding organisations, in my view.

Ziggy Liaquat: That is a fair challenge, and I agree with Andrew. It is something that we need to think carefully about certainly for other GCSEs, but particularly as we look to re-develop GCSEs for 2015.

Q206 Alex Cunningham: As a result of your poor communications in relation to the English exam, many people did not get the A-Level courses they wanted, or they did not get the apprenticeships or the chances that they had there. Some have seen their lives set back a full year because they have to re-sit things, and others have suffered other disappointments. You have said you have learned lessons from this, but what consolation do you think that is to the young people out there who, through your failure, have suffered and seen their lives put on hold?

Ziggy Liaquat: I absolutely understand that it has been a distressing time, and I feel for those individuals. Certainly as a parent, I absolutely understand the distress that it has caused. The key point has to be the point that Andrew made, which is that the decision we made around the grade boundaries in June was the correct decision.

Q207 Alex Cunningham: But you failed to communicate it, so the young people lost out.

Ziggy Liaquat: We did communicate it, but we did not do it effectively enough. Certainly, as I said, we called every single English department we are responsible for and said, “Look, we’re not going to
issue grades in January," but we still got calls. So there is much more we can do, and I absolutely accept that.

Q208 Alex Cunningham: So what responsibility are you going to accept? It is a message; we have seen young people’s lives damaged as a direct result of what you say is your failure to communicate correctly.

Ziggy Liaquat: The important point is that those candidates who have a C grade in English have the skills and expertise necessary to warrant that C grade. We have a responsibility to those students, but if we step back, we also have the responsibility to ensure that English as a qualification retains its currency and value as a qualification for the UK system, but also for how we stand against other systems in the world.

Q209 Alex Cunningham: But is it okay for it to have a different value between January and June?

Ziggy Liaquat: No, it is not okay. As I said, with hindsight, because we had to make an award based on 700 candidates, the January award was generous.

Q210 Alex Cunningham: Hindsight is not any use to the young people who got a grade D instead of a grade C. Do you not think somebody in the system should have maybe lost their job because of this?

Ziggy Liaquat: I am not sure. If we take a step back and take a view of where we are now, I completely accept the distress that our lack of communication has caused, and I apologise for that. Everyone in awarding organisations, Ofqual and the Department has learned lessons about how we need to educate people much more effectively about how the process works, and that is the thing we should be focussed on, to make sure that we do not get into this situation again.

Q211 Alex Cunningham: Apologies are no use to the young people out there, but I accept what you said. What steps are you therefore taking to improve communications with schools about the possibilities of changes to grade boundaries along the way?

Andrew Hall: It is about more than just grade boundaries. I made a point before about the number of routes into schools. One of the particular concerns that I have is in the reform process going forward now. We came in front of this Committee once before on the subject of teacher seminars, as I recall, and I think in the end AQA was proved to be innocent on that. However, there is something about how we do launch new specifications in schools, and I really worry that if we overreact to some of the things that others may have done there, we will not have effective communication. On a number of subjects, I have looked at the launch information we give, and I can provide you with lots of it. It is very clear documentation, and there are clear messages at training seminars: “Grade boundaries will change.” Yet I accept not everybody heard that, because they have told me that they did not all hear it. We have to really fight hard when we launch something, because the biggest risk is at the time of change, and it is at the time of change for subjects or new teachers entering the profession that we have to be able to provide really powerful support that does not go to help teachers teaching to the test or fools telling people what an easy exam it is or something.

Q212 Alex Cunningham: I wanted to return briefly to the marking issue, and this over-marking in places. Teachers have been accused of cheating. Is it your view that that has happened, or is it just a case of teachers working within the system but not cheating?

Andrew Hall: “Cheating” is not a word I would use. I have had it made very clear to me by teachers that they do feel the pressures of those accountability systems. When they are making very fine judgments, I can well understand that it influences how they feel. I do not think someone sets out to do something fundamentally bad, but we put them in a position where we are asking them to make really quite complex judgments with 60% of the marks in a high-stakes exam.

Q213 Alex Cunningham: Yes, but the extra mark is the difference between a D and a C. Are teachers cheating by thinking, “Yes, we’ll just allow them the extra mark”? Andrew Hall: As I say, I would not use the word “cheating”.

Q214 Alex Cunningham: What word would you use?

Andrew Hall: I think they are being put in a position where their judgments are influenced by the pressures of the accountability system, and that is natural human behaviour. If you set up a system of measurement, do not be surprised that people use it.

Q215 Alex Cunningham: So they are just being nice? Ziggy, have you got a view?

Ziggy Liaquat: Yes, I think I made my position clear. We have looked at the data and we certainly see inaccurate marking, but I cannot see evidence of teachers giving away marks to get over boundaries. That is based on the 10% of the data that we have.

Q216 Alex Cunningham: So all the stuff we have seen in newspapers and elsewhere suggesting that teachers have cheated is not true? They have just been working within the system?

Andrew Hall: We have not used the word “cheating”. What we have talked about is showing you the peaks and troughs of marks near grade boundaries, and that is hard evidence. That is about fine judgments being influenced, but I would not use the word “cheating”.

Q217 Alex Cunningham: So, on that basis, do you believe that teachers are very aware of the marking tolerances? There is very clear evidence that they know what they are doing.

Andrew Hall: When we halved the tolerance, what happened was quite noticeable, because you clearly monitor the impact that has, and how many schools went outside tolerance, and the number of schools you need to adjust. Yes, that tolerance was there, and schools were aware of it.

Alex Cunningham: Do you agree, Ziggy?

Ziggy Liaquat: Schools are aware of the tolerance. The thing to bear in mind is that many of these
teachers do act as moderators and examiners for organisations like us, so there is an awareness. I will make it clear again. I go back to the data that we have looked at and analysed, and have provided Ofqual; I cannot see that phenomenon happening.

Q218 Alex Cunningham: So you are both satisfied that feedback and moderation will be much sharper in future?

Ziggy Liaquat: The reduced tolerances help focus attention, absolutely.

Q219 Mr Ward: It was almost a blinding flash of the obvious. If the outcomes of football games were determined on the number of corners as opposed to goals, there would be different results. It was so obvious, was it not? Wasn’t there an awareness of how obvious it was that it was going to impact on behaviour?

Andrew Hall: I am in danger of repeating what I said before. When the qualifications were designed, it was driven by curriculum expertise. Whether there was enough expertise in the interim regulator at that time to understand the standards around it is open to debate, and I have said so on many an occasion. For awarding bodies the accountability pressures have grown over time. That is not any one party or another; they just have become more and more focussed, and I think there is a recognition of that, which we did not understand three or four years ago.

Q220 Chair: You have taken a lot of stick, and AQA has taken probably a heavier share of the stick. You are the largest awarding body in English. You have relationships with schools who are your customers. Is there not an issue here—when we go back to Alex’s point about miscommunication and children being left with a sense of grievance, which may or may not have been right—that since all this came out, schools have not always behaved appropriately? Do you feel that those who went to court and lost behaved not always behaved appropriately? Do you feel that they have created miscommunication and suggested that, for instance, grades given in June were wrong, when in fact the court, Ofqual and the united bodies have said that the data seem to suggest that broadly, within these tolerances and other issues, they were right? Have you got any criticism for them?

Andrew Hall: What I found interesting was that not every teacher organisation chose to take part in that organisation. Some trade unions did, some trade unions did not, so they all clearly made their individual judgments. I am not always sure about lessons to be learned, because I do not particularly like that phrase. However, what is noticeable now is that for the non-graded exams that were taken in January, the raw marks have been issued to schools. A number of us are already having to hose down the parents and more pupils at the school, and the exam boards want to attract more schools. Is there not a great concern there that sometimes the truth loses out in all of this, and we are all not willing to criticise each other because we are all in a sort of customer-client relationship with each other? Is that not an overriding concern, when what we are talking about here is a system of trying to educate the nation’s young?

Q221 Chair: So did some of them use this as a sort of smokescreen? A lot of schools did better last year than they had done the previous year while some did really pretty badly, and that always happens when there is a lot of change brought in. However, fundamentally is that because some schools read the spec and some did not, and those that did not, and let down their children, then used this furore or fiasco as a smokescreen to hide their own underperformance? Is there any fairness to that criticism?

Ziggy Liaquat: It is entirely understandable, for the reasons that Alex outlined—the distress and anxiety—that they represent the views of their pupils and their institutions. One of the helpful things both the Ofqual investigation and the High Court judgment have revealed is how the process works, and how we apply that process professionally and diligently, not for a second undermining the feeling of injustice. So if it helps inform the greater public about how the process works, that has got to be of benefit to all of us in the future.

Chair: A very politic answer.

Q222 Ian Mearns: We now seem to be developing customer and client relationships between exam boards and schools, and between schools and the parents, because the schools want to attract more parents and more pupils at the school, and the exam boards want to attract more schools. Is there not a great concern there that sometimes the truth loses out in all of this, and we are all not willing to criticise each other because we are all in a sort of customer-client relationship with each other? Is that not an overriding concern, when what we are talking about here is a system of trying to educate the nation’s young?

Andrew Hall: All I would say is that it has not felt for the last six months that people are shy of criticising, and we have had our share of both being critical and being criticised. There is something about maturity, of being prepared to say what needs to be said, and that is why I think—it is not just because I am here today—AQA have been prepared to stand up and say: “Actually, before it was fashionable to say it two years ago, we were saying results are really bad, and early entry is really bad. This is hurting our income line, but this is bad education.” Organisations have to have the courage to do that, and you have the naive hope, or the hope, that over time that credibility bears its fruit.

Ziggy Liaquat: Our primary focus is and has to be the longer term value of these students having these qualifications for their careers and their further education. Otherwise, these qualifications have no currency. The challenge and the scrutiny is absolutely right. However, as an organisation, we are in this for the long term, so that relationship issue is also with maintaining a standard, year on year, for qualifications, and also looking at that other education systems are producing for their young people. That has to be our focus, and there will be challenges along the way. I am sure of that.

Q223 Ian Mearns: Andrew, you said earlier on, and it is just something that stuck in my mind, that you
12 March 2013  Andrew Hall and Ziggy Liaquat

did not really like the idea of learning from the experience, or learning a lesson?  
Andrew Hall: No, what I meant is that it is the phrase I hate. You do need to learn from everything, but it is too easy a phrase sometimes, and used to deflect: “We’ve learned lessons.” It is about really being prepared to learn the lessons, to look at things and account for it. It is just as a phrase.

Q224 Ian Mearns: Do you think it is a throwaway line? Is that what you are saying?  
Andrew Hall: It can on occasions be used inappropriately. It is about having the real courage to look at something that has happened and honestly learn it, and just to talk about lessons learned—it is kind of a view. It is a throwaway remark, no more.

Q225 Ian Mearns: So saying it and doing it are two different things?  
Andrew Hall: Yes.

Q226 Ian Mearns: Do you think that Ofqual were able to conduct as thorough and impartial investigation into what happened as was necessary or, effectively, do you think it was investigating itself in some respects?  
Ziggy Liaquat: You have identified a challenge. Ofqual is obviously deeply involved in the setting of grade boundaries for English. It is an odd situation that they are investigating an issue that involved them. However, if you look at the outcomes of those investigations and the fact that they initiated the process early so that we were able to offer a re-sit in November and to tighten up controls in January, that is a really good outcome from that investigation. If you look at the final outcomes of their investigation against the findings of the High Court, you have to say that they did a good job in identifying some of the issues.

Andrew Hall: It was fine, because you have to look at where the expertise is to carry out an investigation. During the process it certainly did not feel cosy. It did not feel that the investigation was carried out by someone who was integral in the process. It felt very independent, and we were challenged a lot. We put our technical resource behind it, hence the reports you are seeing.

Q227 Ian Mearns: Do you think Ofqual could have been any more preventive in trying to stop all of this happening in the first place?

Andrew Hall: I have made my comment about how I felt when the qualification was designed, and you have to draw the distinction between Ofqual as it is now and Ofqual as it was in its interim phase, particularly in terms of the increased experience and skill that Ofqual have got in the standards arena now, which they did not have when it was formed in 2009 in an interim form.

Ian Mearns: Ziggy, do you have anything to add to that?

Ziggy Liaquat: I would say Ofqual are a young organisation, and it is a challenging time. Like all of us, in hindsight—and I will not say lessons learned—there are things we maybe could have done differently.

Q228 Ian Mearns: Not as a throwaway line, but do you think there are real lessons that Ofqual can learn from this whole sad experience?

Andrew Hall: I think there are. It is about being sure you have the right technical skill in your organisation to understand the issues you are facing and those that are coming. Everything I have seen about the recruitment, frankly including my having to lock down some of my own staff so that they do not get poached, shows that that is happening.

Q229 Ian Mearns: Pretend for a moment that Glenys Stacey is not listening to this. Do you think that they have the wherewithal within them to do the job effectively in the future?

Andrew Hall: They have now, but they did not in their interim form.

Q230 Ian Mearns: Do you think it is right that the fairness of an exam process should ultimately be decided by the courts of justice?

Andrew Hall: I do not think any of us wanted to be there.

Q231 Ian Mearns: As an alternative, for instance, do you think the Secretary of State should have set up an independent inquiry?

Andrew Hall: That is a very different thing. The court is a matter of absolute law, about process and “Did we do the right thing?” There was an Ofqual review, and they did it thoroughly in my view. If you read the reports, they are very detailed. The job was done.

Chair: Thank you both very much for giving evidence to us this morning.
Examination of Witnesses

Witnesses: Mark Dawe, Chief Executive, OCR, and Gareth Pierce, Chief Executive, WJEC, gave evidence.

Q232 Chair: Thank you both very much for joining us today. You both listened to the last session. What was said by the witnesses in the last session, either by Ziggy or Andrew, that you most disagreed with, Mark?

Mark Dawe: I do not think I disagreed with too much of what they said. The most fundamental point was around the design of the qualification. If there are lessons learned it is that when all the awarding bodies are saying, ‘This will not work. This will cause problems,’ they should not be ignored.

Q233 Chair: Yet they were in 2009 before Ofqual had been made independent, so it was basically a creature of Government, and it nonetheless just ploughed on, and did it anyway.

Mark Dawe: It sounds like they wanted a neat system where controlled assessment was zero, 25 or 60, and you had no option of anything else, so in designing new qualifications we have got to be very aware that every qualification is different. Obviously with threshold measures on English and Maths, there will still be the same pressure on those qualifications. Therefore the design of those probably needs to be different from some of the other GCSEs.

Q234 Chair: Basically, the design at that point was effectively under ministerial control. Do you think we have now got the balance right—before I turn to Gareth—between Government and independent expertise in assessment?

Mark Dawe: You were talking just now about the expertise within Ofqual. We hold years of expertise, with Cambridge offering in around 160 countries. We understand that if we work together and come to an agreement, then that is the best way forward. In the past awarding organisations have made it a skill to make a silk purse out of a sow’s ear, or a quality assessment out of political and regulatory aspirations, so we have had to do that in the past. It is much better if we all agree on what the best way forward is.

Q235 Chair: But is the settlement whereby Ofqual is independent of Government through statute a better situation than we had before?

Mark Dawe: It is. I was not part of that in the past, but from what I have heard, we are in a much better place than we have been.

Q236 Chair: So ultimately having Government deciding on grade boundaries is not healthy?

Mark Dawe: I do not think that is a healthy way forward.

Chair: What do you think, Gareth?

Gareth Pierce: On that point, I agree that regulation should be independent of Government, and also independent of awarding bodies. We have corresponded fairly extensively on that point.

Q237 Chair: Was there anything you particularly disagreed with in what was said?

Gareth Pierce: No, not in terms of disagreement. 2012 was essentially almost a collision between complexity and simplicity. There was huge complexity in terms of the qualifications; there were two parallel but fundamentally different qualifications in English: GCSE English and GCSE English Language. There was complexity in terms of the different uses made of early assessment opportunities, and complexity in terms of the demand in the moderation elements. Then there was simplicity, on the other hand, in the use of data in the regulatory framework for making judgments about awarding standards.

Chair: Excellent. Thank you. We will probably cover some of the same ground we covered in the first session, so please do not feel the need to go on at any length if you essentially agree.

Q238 Mr Ward: We talked earlier on about the need for public confidence and loss of confidence in the system as a whole. You were not involved in the judicial review, but there has now been a judgment. Have you reflected on what that means in terms of public confidence in the system, and what actions are needed?

Mark Dawe: Public confidence is vital and, again, that is the role of the regulator. The regulator and the awarding organisations believed that we followed a process that was appropriate. It is very sad that it went to court, and that is one of the key issues to look at: how on earth we ended up in court—not us, but the other awarding organisations with Ofqual. Where in the process could we have stopped that happening? It was appropriate that Ofqual investigated it. However, when the interim report came out, it did close down quite a few options. It did not really share the issues and the dilemmas that Ofqual were having to deal with, and at that point, if everyone had been prepared to sit round the table, understand deeply what was going on and realise that there was no perfect answer but that the system needed to come up with an answer, the other bodies would not have been in court.

Q239 Mr Ward: The issue of fairness is often discussed. Russell Hobby from the NAHT said “the system was so flawed that the regulator had no choice but to be unfair to some group of students”. Do you think that is true? There was simply no choice—somebody had to be treated badly?

Mark Dawe: If you read the court findings, it reflects what a number have said, which is that the January session was hard to award, and with hindsight was probably slightly generous, and in the summer session the right grades were given and the right boundaries were set. There was an overall fairness, but with some issues around the January session.

Gareth Pierce: From our perspective, we did not have any issues in terms of major differences between January and June with WJEC. However, there are complex questions to do with the impact of the ways in which different schools in England chose between the GCSE English and the GCSE Language route. That was a considerable factor, I think. Linked to that, was the use of data sufficiently robust to ensure that
there were not any aspects of unfairness in the way data was used?

Chair: Can you expand on your first point?

Gareth Pierce: I mentioned that there were two qualification routes, and it is unusual in a core subject to have such different routes. Therefore schools were making balanced judgments on which of those routes to use, in most cases, for different subsets of their candidates. Now, on reflection some schools may consider that that was an important factor that may have affected their outcomes and the choices they make this time round; there are those issues as well.

Q240 Chair: One was easier than the other? Is that the point you are making?

Gareth Pierce: It is not so much that they were easier, but they were different. The demands were different, and therefore, in the light of experience, schools will know how different candidates would have fared in those two different routes. Some may be reflecting that first time round they came to one view, and second time round, in the light of that first experience, their view may be different. That is a complexity that is very unusual in a core subject and especially, as we heard before, in one that features so prominently in the various measures of accountability.

Q241 Mr Ward: Still on the issue of fairness, what do you believe should be the overriding fairness consideration when administering examinations?

Mark Dawe: It is about the standard, and the expectation of what a student will have achieved when they get a grade C, a grade B or a grade A. That is where the judgment of the examiners is vital, and we balance the examiner’s judgment, looking at the student’s papers alongside a whole range of statistics, to make sure it is set at the right level.

Gareth Pierce: Fairness also includes ensuring that the learning objectives that teachers and learners have been working towards are fairly reflected in the assessment method used for that course. This is obviously very important in terms of discussing the fate of, for example, the speaking and listening element. If that is a key part of the learning outcomes and objectives, then to be fair to learners, that element must have a fair status in the eventual assessment, so then there is fairness in the design of an assessment to reflect the curriculum intention.

Q242 Mr Ward: I think you were here for the earlier session; I asked a question about the comparable outcomes. It is the same sort of question, but do you have anything new to add in terms of the value? Was too much weight placed upon the Key Stage 2 base for the comparable outcomes?

Mark Dawe: Certainly our approach is, as I said, to look at the examiner’s judgment and then the stats. The judgment normally gives a range, and then the stats help home in on a final boundary or final marks. Regarding the comparable outcomes, if the boundaries had not changed in the summer, the results would have gone up 7%, and we would probably still be sitting here, saying, “How on earth could students’ performance go up by 7% in one year?” There has to be a way of measuring the standard of the students, and it seems to be the best way at the moment, combined with some other factors.

Gareth Pierce: The primary focus has to be on the quality of students’ work, and there is a risk of placing too much reliance on one single statistical indicator. Of course the Key Stage 2 has largely been used in that way, on its own. Therefore the use of that was instrumental in determining whether it was comparable outcomes for England.

Q243 Chair: You do not have that data for Wales—is that right?

Gareth Pierce: No, that is right. The point for England was that comparable outcomes was interpreted as saying the outcomes should come down by one percentage point because of what that predictor model was saying. Now, we do not agree with that view. We think that is over-deterministic as a use of a single predictor model, and our view is that there is always a need for a balanced use of a range of relevant statistical information.

Q244 Chair: The last panel suggested that there was. Surely when you bring in changes in every single aspect of an examination, as we saw last year, then that is where you rely particularly on something like comparable outcomes. Otherwise you have no lodestar by which to make sure that there is not a serious change in boundaries year by year.

Gareth Pierce: Yes, it guides that process, but comparable outcomes is not a neatly defined concept, as we saw in its use last year. For such a large population of candidates, the decision was that the outcomes should come down by a percentage point. That is a particular interpretation of comparable outcomes for England, in the light of, in particular, one piece of predictor methodology, which is the Key Stage 2 one.

Q245 Mr Ward: We have already covered this, but, just quickly, on January/June and the delay, presumably you are supportive of that change, and the delaying of the January grading?

Mark Dawe: In the original discussion we were not.

Gareth Pierce: It was a discussion, and we gave our view. We did not feel that OCR had an issue in awarding, and we felt we would have been able to award in January, especially as another year had gone by, and as a qualification gets older it is much easier to award. However, we were not the majority, and it was important that there was a set process for the country as a whole, so we agreed that, overall, that was the way to go. Others felt it was important not to do that.

Gareth Pierce: Our position is identical. The view we expressed was that there was no need to hold back from grading in January but, again, our perspective was based on 2012, where WJEC did not really have any issues between the January and June comparisons. However, we accepted that the ruling had to be binding on all awarding organisations, and across the two countries.

Q246 Mr Ward: Just a different area now: the implications for WJEC and the English exam boards of the recent announcement by the Welsh Government
that there will be separate question papers and grade awarding meetings for GCSE English Language candidates in Wales. Any comments on that?

**Gareth Pierce:** We are obviously acting on that direction or special condition. The view that our awarding committee will take is that they are awarding to a single set of standards, so that will be their point of departure. In a sense, the fact that there is a separate question paper that candidates in Wales would have taken is a bit like having a January paper and a June paper. Their aim will be to set the same standards across both papers, so technically it does not add any particularly different issue for them. It just happens to be another paper for them to look at.

**Q247 Chair:** So there are not any implications?

**Gareth Pierce:** At this point there are not any. They will look at it as a paper that has to be awarded to the same standard. Whether there will be any implications or not depends on what the data will show.

**Q248 Chair:** Are universities, the CBI, IoD and other business organisations and university organisations in Wales happy with the direction of travel and this separation, with Wales having different standards, and arguably possibly lower standards over time than England?

**Gareth Pierce:** Our line at WJEC is that we are working within a common framework of standards, and that 2012 was one instance where a decision was made by the Welsh Government that in one subject the standard would be different. Now, heading into this year’s awards, we are working within a three-country GCSE framework, and our view is that that three-country framework represents a single set of standards. That is where our awarding committee will start from. If there are issues in the data for our candidates from England and candidates from Wales that pose issues for that awarding committee, we will have to have discussions with the two regulators. However, at this point, we are working to a single set of standards.

**Q249 Chair:** Do you think there is a temptation for politicians in Wales? It would appear that Welsh educational outcomes are falling behind England, and do you think that may have played a contributory part in their decision to allow the re-grading and, some would say, the artificial boosting of results in Wales compared with the results in England?

**Gareth Pierce:** There are two factors we are aware of that were instrumental in relation to the Wales results, and the Welsh Government are aware of these as well. Firstly, schools in Wales made less use of the early entry opportunities—January units and so forth—compared with schools in England. Secondly, there has been a Welsh Government initiative to support some teaching departments of English in Wales in responding to the demands of the new specification. In other words, it is acknowledged that not all schools in Wales responded as perhaps they might have done to the demands of the new specification. So there are those two factors, which did contribute to the scenario we had with the data for Wales last year.

**Q250 Chair:** Had pupils in Wales been falling behind compared with pupils in England?

**Gareth Pierce:** Not in the way that we look at the data from an awarding perspective, but of course there is a wider perspective on that, which relates to other aspects of performance data to do with cohorts in Wales and England. I am perhaps not the person that should be asked about that.

**Q251 Chair:** Going back to my point about universities, have you had people expressing concern, from businesses and universities, about the possible divergence of standards between Wales and England?

**Gareth Pierce:** No. The 2012 scenario is at the moment recognised as a single situation.

**Q252 Chair:** No concerns have been expressed to you by universities or employers?

**Gareth Pierce:** No, not to WJEC, and we are projecting the view that, in these shared brands of qualifications, both GCSE and A-Level, as an awarding organisation working in England and in Wales, we are very much part of that joint use of those brands to a single set of standards. That is our message to universities, employers, parents and learners.

**Mark Dawe:** The point about a single set of standards is really important, and there is a danger that we end up with a qualification with the same name with different standards, and we have to avoid that at all costs. Really, it is down to the three regulators in Northern Ireland, Wales and England to sort this one out between them, as to how they are going to regulate. Either there are different qualifications in each country with different standards set, or there is an exception, so there is one qualification with one common process for standards.

**Q253 Chair:** Are you confident that that is going to happen?

**Mark Dawe:** Not at the moment, because I wrote a while ago to all three regulators asking for clarification, and we have not received that yet. As JCQ, we have done the same, and we do not have clarity yet.

**Q254 Mr Ward:** The issue is that there was a view, after the judicial review and the deliberations, that that was it. You know, it was a fiasco, we are over that now, and we need to learn the lessons. Isn’t there some temptation to believe that is not the case now—that the uncertainty is going to continue, and comparability will be very difficult, particularly for employers and universities? It is still going to go on.

**Gareth Pierce:** Comparability is an issue when you have got differences in specifications, and it adds another dimension to the complexity, but it still has to be addressed if it is the same brand. For example, in 2014 there will be a change in the specification for Wales, for GCSE English Language, a year earlier than a similar change in England, if I have got my dates right. Now, they are still going to be within the GCSE brand, so the challenge for regulators and for awarding organisations is to still work together on standards, and still sustain public confidence that a
grade C and a grade A, and so forth, means the same even though there are differences in the specifications. As a set of awarding organisations, we have always had this as something we needed to work with, and there are ways of doing so. However, the regulators need to work together on it as well.

Q255 Neil Carmichael: With the forthcoming reforms in GCSEs and A-Levels in mind, what do you think the key lessons are from this business with the 2012 GCSE English results?
Gareth Pierce: It is pointing towards greater care being needed in the design of qualifications, and the join between curriculum intentions, assessment design and the pedagogy that is going to be delivering that curriculum has to be thought through pretty carefully when a qualification gets reformed. The choice of assessment method has to be based on fitness for purpose, but also has to be robust, especially in the context in which qualification outcomes are used as data in accountability contexts, and so forth. The lesson learned is that we have to take that rounded view of issues impacting on qualification design from the outset and reach good collective decisions.

Q256 Neil Carmichael: So from what you have seen, do you think the Government is heading in the right direction?
Chair: Mark?
Neil Carmichael: I was thinking Gareth might want to answer that question, but Mark can certainly have a go.
Mark Dawe: The most important thing is what goes on in the classroom, and what the students come out with at the end in terms of their learning, ability, skills, knowledge and understanding. That is what we are looking for, and obviously qualifications have a role to play, but that is about the quality of the teaching, the way of learning and the use of technology. There is a whole range of things. We have got to get that right, and then we have to design an assessment that then facilitates that learning, identifies where it has taken place and encourages high levels of achievement. As long as the different parts of the education system talk to each other and work together, we will get there. The problems have occurred where it has taken place and encourages high levels of learning outcomes, and therefore they get reflected in the curriculum, you design your assessment depending on what your learning outcomes are within that grade C and grade A, and so forth, means the same even though there are differences in the specifications. As a set of awarding organisations, we have always had this as something we needed to work with, and there are ways of doing so. However, the regulators need to work together on it as well.

Q257 Neil Carmichael: The last panel had an interesting discourse about the tensions or relationships between the Government regulator and exam boards. Have you got anything to add to that? What are your thoughts about that relationship, and how do you think it should develop?
Chair: As briefly as you can.
Mark Dawe: It is important that the three work together and sit round a table. Now franchising is out, it means that everyone has come back round the table and is sharing things, sharing their experience and their expertise, rather than all sitting in a dark room preparing for competition. There are definitely benefits for GCSEs there, and we are seeing the same sorts of conversations happening with A-Levels. As long as everyone is willing to put in the work and share, we will get the right outcomes.
Gareth Pierce: The way forward is now allowing awarding bodies to work collectively, in the way we have done in the past, including working collectively in dialogue with regulators and Government Departments. The regulators then need to have a clear understanding within these shared brands of what they need to work on together, what the scope is for doing things differently and, when they do things differently, how they can still join up on a common understanding on standards.

Q258 Neil Carmichael: Mark, the Welsh board did raise concerns earlier about the 60% weighting. Did your organisation do the same?
Mark Dawe: Yes, we did.
Neil Carmichael: What sort of feeling did you have about the response you got?
Mark Dawe: I was not there, but from talking to people we were just told, “This is how it will be.” As I said earlier, it was trying to fit a neat model. The regulator at that time said, “We have this model and your qualification must fit into one of these.” That forced English to the 60% controlled assessment. It goes back to lessons learned; I quite like the phrase. We must not try to squeeze every qualification and every subject into a small number of rigid models. It is not the way it should be.

Q259 Neil Carmichael: In general you would prefer an examination system rather than an assessment system. Is that the view of both of you?
Gareth Pierce: Are you saying an examination system means written exams at the end?
Neil Carmichael: Yes.
Gareth Pierce: No, I personally do not. Judgment has to be made on the basis of fitness for purpose. Depending on what your learning outcomes are within the curriculum, you design your assessment instrument to reflect those. In many subjects it is important that there are other assessment elements than just an exam at the end of the course.

Q260 Neil Carmichael: So would you want to say that speaking and listening should be included in some sort of assessment in English?
Gareth Pierce: Yes, if they are a key part of the learning outcomes, and therefore they get reflected in the pedagogy for that curriculum, there needs to be a credible way of including them in the overall assessment. Now, obviously that is going to be up for debate, as to whether that can be done robustly. However, fitness for purpose is the guiding rule with assessment design.

Q261 Neil Carmichael: So, Mark, where would you strike the balance between an examination in the traditional sense and assessment?
Mark Dawe: Again, it is subject by subject.
Neil Carmichael: Well, in English, because that is really what we are talking about.
Mark Dawe: In English, like the others, we said that 40% was appropriate. We have to look at whether we really have to assess everything that goes on in the classroom. It is a sad situation if we cannot have teachers teaching in science labs or in speaking and listening, and things like that, unless they are being assessed on it and focusing on that assessment. So we have to look at whether the assessment contributes anything to the grade as well. In some of these cases the assessment does not add anything; it is very narrow, and you do not really get a good result out of it. We have to balance all of those things.

Q262 Alex Cunningham: Way back in the last century when I was at secondary school, I did the GCEs. Alongside that there was a certificate in science teaching. I think GCSE English in future should continue to have an element of internal assessment for speaking and listening. Do you think the GCEs. Alongside that there was a certificate in science teaching. I think GCSE English in future should continue to have an element of internal assessment for speaking and listening. If so, could that moderation be strengthened in some way?

Mark Dawe: We both made comments. I would favour having a separate recognition of speaking and listening but not necessarily a contribution to the final grade as one way forward.

Q263 Alex Cunningham: So in the early 1970s it was right and in the 21st century it was wrong?

Mark Dawe: Not at all. I do not think I said that.

Chair: Only the dress sense is regrettable.

Q264 Alex Cunningham: Moving on now, do you agree with Ofqual’s analysis that teachers over-marked controlled assessment? Did you find any evidence of that? Has there been inflated marking?

Mark Dawe: At OCR we did not find evidence. The curves you were talking about earlier were reasonably smooth; there were peaks through the whole range. We had had a modular assessment previously, and we believe that our centres understood how that worked, and how the controlled assessment works in there, so it comes back to Andrew’s point about training and support. We are still waiting, as awarding bodies, to it. We have to balance all of those things.

Q265 Alex Cunningham: So you would join others in dismissing the suggestion that teachers have cheated?

Mark Dawe: No, not at all. I do not think I said that.

Alex Cunningham: Sorry, I asked the question the other way round. I was surprised at that answer there. What I was saying is that you would accept that teachers have not been cheating.

Mark Dawe: Sorry, they have not cheated. No, they have not cheated.

Q266 Alex Cunningham: Is that your view as well?

Gareth Pierce: Yes, we would agree with that. The extent to which we had to make adjustments for schools being out of tolerance was no more than we would have expected. Obviously with a narrower tolerance going forward, we expect to make more adjustments in future but, again, we do not regard that as having been a major factor in 2012.

Q267 Alex Cunningham: So there was not anything bizarre about the way that marking took place last year?

Gareth Pierce: No.

Q268 Chair: Are people reading too much into the chart showing the marks for your English exam in England and Wales, and the fact that it looked spiky around the grade boundaries in England and smooth in Wales?

Gareth Pierce: Possibly, yes. The smoothest curve was for our Wales data, the next smoothest was for our England data, and then the least smooth was for AQA’s England data. Our England data was between the smooth and the very obvious patterns. That is why it was less of an issue for us, and it is shown in that data.

Q269 Alex Cunningham: Mark, you have suggested that your teachers got it and were fine, so do you think that some schools have been penalised as a result of problems elsewhere when they have averaged things out or they have raised the boundaries to address inflation?

Mark Dawe: It is really hard to know school by school what was happening. I think it was Andrew who was saying that the biggest surprise to us all when we looked at all the data together was the variability between schools year on year in 2012. There was a concern there about what had happened. It was one of the factors in this—something did not seem right there.

Q270 Alex Cunningham: Ofqual said that as exam boards you “could have done more to identify and report emerging concerns to the regulators”, and that “monitoring by exam boards could have been stronger and more intelligent”. What do you understand by that, and what are you going to do as a result, bearing in mind that you think you had the ducks in a row?

Mark Dawe: I would disagree with the black and whiteness of that statement. We did a lot, and we do a lot. At the awarding meetings early on, it was clear that English was going to be difficult. We awarded and did not change anything after that awarding. We awarded before this issue really started to flare up. There are regular meetings with Ofqual both during the awarding period—every week we are sitting down together, identifying where there may be issues in
particular subjects—and then through the year, where there are particular subjects where we may say, “Well, given what has happened in English, are there others we should be looking at?” We are all doing that, so I do not think it is fair to say that we were not monitoring and keeping a close eye on things.

**Alex Cunningham:** Is that your view as well?

**Gareth Pierce:** Yes. I would agree with that. What perhaps took us all by surprise was the significance of the issues that emerged in English, and they did emerge very late. Apart from those design issues, which were way back, it was only very late during the awarding cycle that there was an awareness of the extent of some of these issues. We now know what can become an issue, and therefore there is more of a dialogue on those throughout the year now.

**Q271 Alex Cunningham:** So the lessons have been learned this time?

**Gareth Pierce:** Absolutely.

**Alex Cunningham:** We cannot expect some of the spikes and disappointments that we have seen over the last year. Is that where you are? I see heads nodding. That will be fine.

**Mark Dawe:** Yes.

**Q272 Alex Cunningham:** I will go on to the next question. The marking tolerance was tightened from November 2012, reducing it from the previous 6%. Is that a welcome change? Who sets these tolerances, and does it apply across the piece, to all subjects and boards?

**Gareth Pierce:** Yes. I am just going to confirm the answer you had in the first session: the setting is by agreement across the awarding organisations with regulators, and the change is only for English Language at this stage. However, the benefits that are going to come from that change will cause us, I am sure, to reflect on the extent to which that change should be applied elsewhere.

**Q273 Alex Cunningham:** If that is such a great idea, why wasn’t someone saying this 6% tolerance was a bit bizarre?

**Gareth Pierce:** There are issues of practicality whenever a tolerance is set, and also it is in the context of what contribution that element makes. Now, on reflection perhaps we should have realised that an element contributing 60% required narrower tolerances than something contributing half that.

**Q274 Alex Cunningham:** They existed for a long time, though, didn’t they?

**Gareth Pierce:** Yes, they did.

**Mark Dawe:** They did, and there was not a problem. There also comes a point where the narrower it gets, the more you might as well just mark the papers yourselves.

**Alex Cunningham:** Hopefully we will never get to that stage. The Secretary of State wants to go the opposite way, and just have one exam; that is certainly something that I do not agree with.

**Q275 Siobhain McDonagh:** What steps are you taking to improve your communications with schools about the possibilities of changes to grade boundaries?

**Gareth Pierce:** We are building on the dialogue that has been happening over the last six to nine months with schools, especially in terms of English Language. Obviously the unusual arrangements that have surrounded the January series have been part of that communication. There is considerably greater awareness both through our deliberate correspondence with schools, and through the general public discussion. There is a greater understanding now of what grade boundaries mean and the extent to which they do genuinely vary from series to series.

**Mark Dawe:** Likewise, the concern Andrew voiced around students being told by schools what their grades are, even though we have not issued grades, just shows again that they probably have heard the message, but they are very worried about this threshold measure and the performance of their students. They will be looking to see whether a student has got a grade C or not in their eyes, and whether they should be putting them in again. So it does not matter; we can tell them as much as we like, but there are still other pressures and other issues they will be looking at that will affect behaviour.

**Q276 Siobhain McDonagh:** Ofqual has said that “moderators’ feedback to schools was not always sharp enough”. What do you understand by this, and what steps are you taking to ensure the feedback is sharper in the future?

**Gareth Pierce:** We have certainly looked at the kind of narrative that we have been putting in moderators’ reports, and one clear distinction we need to make is between saying that their work is perfectly acceptable administratively, which is one thing, and the different comment of saying they were absolutely spot on on the standards. In the way we have been drafting and finalising these reports, we are clearer now in terms of making sure that messages are distinct in that feedback. Some of the messages are to do with process, and some are to do with standards.

**Mark Dawe:** After the issues arose, we reviewed our guidance and what moderators were writing, and the majority of them were pretty good, and we were not getting this issue of schools feeling like they had been given guidance that they were doing the right thing. However, like all these things, we sharpened it up and we repeated the messages, so that hopefully there was no misunderstanding.

**Q277 Mr Ward:** I have some questions about the role of the regulator and the ministerial action that was taking place. You did express concerns about the independence of the regulator in its investigation of itself. That was justified by the regulator, but in your view was Ofqual able to conduct a thorough, impartial investigation into what happened?

**Chair:** Who would like to go first? Were they the right people to investigate this?

**Gareth Pierce:** In one sense, they were reviewing things that they had not been accountable for from their origins. For example, there are aspects of this that go back to 2009, aren’t there, so the fact that they
were not responsible from that stage helped them, in a sense; they could take a detached view on some of the origins of these issues. Of course, they were probably best placed to take an overview of what we had all been doing as awarding organisations, because they had clear access to our information. So probably, on balance, their position was appropriate to report and investigate. Of course, whether that satisfies all parties is a different matter.

Mark Dawe: They were set up to do this sort of job, and to some extent if they are not given that role, then why bother with the regulator? As Andrew said, it felt very uncomfortable in terms of the amount of stuff we were giving and the questions we were being asked, and that was very technical data. You needed someone with the expertise to understand what we were handing over to then be able to judge what had happened, and I do not think anyone outside the regulator would have been able to do that job appropriately.

Gareth Pierce: One other point I could make is a joint regulatory investigation would have been perhaps even better.

Q278 Mr Ward: The regulator argued that the issues surrounded schools and your organisations, rather than the regulator itself. Is there anything the regulator could and should have learned from it?

Mark Dawe: This is where the body changes. They say, “It is not us,” but it was the original form of the regulator that caused some of these problems with the original design, so as a whole there is an issue there that the regulator, in its previous form, should take responsibility for.

Q279 Mr Ward: So it was blame-free in terms of what happened last year, in your view. It could not regulate it.

Gareth Pierce: It was independent, to a large extent, of the complexity of the scenario with which we were working. Therefore, it was well placed to comment on that and its implications, but of course it was very much tied into what I described at the outset as too simplistic an approach to the use of data. We are learning a lesson from that as well, and there is a greater openness now to the importance of taking a balanced view of different parts of the statistical evidence.

Mark Dawe: I would go back to the point I made right at the beginning: the regulator had the role of getting everyone around the table and stopping this court case from happening. There possibly were things that could have been done halfway through this process that would have enabled those discussions between all the parties outside court rather than in court.

Q280 Mr Ward: The previous regulator or regime was responsible for the introduction of the system. Was there nothing the current regulator could have done to foresee what would happen?

Mark Dawe: It goes back to Andrew’s point that, staffing-wise, the expertise in the previous one was very limited. The current regulator has been building up their staff base. This was happening right at the beginning of the regulator’s life, and I do not think they would have been aware of many of the complexities that they had inherited.

Q281 Mr Ward: Was this just a train crash waiting to happen?

Mark Dawe: We said so in 2009. We said this qualification was going to be a problem.

Gareth Pierce: On the credit side, both Government and regulators in England and in Wales are now recognising what actions need to be taken going forward, and those actions are happening rapidly for English and they are influencing the debate on other GCSEs. It is to the credit of all involved that lessons are being learned and acted upon very quickly.

Q282 Mr Ward: Can I talk to you, Gareth, about the re-grading in Wales and your position with regards to that? What are you views on the re-grading?

Gareth Pierce: Our position is it is something we were told to do, and therefore we did it, as a direction from the regulator. It is documented that our awarding committee did not find that an appropriate position for them to be in. That is why, in approaching this summer’s awards, they will be getting back to the point of wanting to apply a common set of standards. To an extent, the decision in Wales can be seen as a response to the two factors I mentioned earlier. There were two issues that our data and their investigation of teaching departments would suggest were material factors. It could be interpreted as an approach to give comparable outcomes for the Wales candidates despite those two factors. The two factors I mentioned were: less use of early opportunities, and the issues in terms of some teaching departments in responding to the demands of the new specification.

Q283 Mr Ward: So, in your view, was it the right thing to do?

Gareth Pierce: It was not the right process for our awarding committee to be involved in, because it is not tenable to have different standards for the same qualification. So the question has to be asked in the wider context of those factors affecting candidates in Wales. Having said that we are an awarding organisation that applies common standards, it is for other people to decide whether those other contextual factors made it the right decision.

Q284 Alex Cunningham: Can I just ask a question about the value of the grade now? We have the value of a grade in January, we have the value of a grade in June, and we have the value of a grade re-marked in Wales. What is the value of that grade? Is the value of a grade re-marked in Wales the same as the mark in January? Where is the value? What I am saying is there are three different elements to that mark and they all have a different value. Would that be your opinion?

Gareth Pierce: Yes, within the WJEC award there are two slightly different awarding standards applied to the grade C, and the grade C alone. In a sense, they reflect the two regulators’ different views on comparable outcomes between 2011 and 2012.
Q285 Alex Cunningham: But to the employer, one is worth more than the other.  

Gareth Pierce: I suppose if you get down to that level of detail there is that difference, but it is a small difference. It is statistically a small difference, but you are right that there is a difference.

Q286 Chair: You said that two points contributed. One was that in Wales the nature of entries was different from England a bit, and the other point was...  

Gareth Pierce: Teaching departments responding to the new specifications.  

Chair: Exactly, but in England exactly the same thing happened. I know in my constituency I had a school that did even better than normal and another one that fell off a cliff. That would be because of their interpretation, understanding and recognition of the new specification, I assume.  

Gareth Pierce: Yes. The reason I am mentioning it specifically for Wales is because I am aware that the Welsh Government took a specific initiative to support teaching departments. That is in addition to the support that awarding organisations would give in general terms for teaching departments, so that is why I am mentioning it as a factor. Regarding the other one, the use of early units, there is quite a significant difference statistically in the extent to which Wales centres made less use of the January opportunity compared with England centres.

Q287 Chair: Sorry; so the Welsh Government gave additional help that you would not have seen in England, say, to help schools make sure that they understood the specification of the new qualifications.  

Gareth Pierce: They set up an initiative to provide that support in Wales, yes.

Chair: Prior to the whole—  

Gareth Pierce: After the 2012 series; well, from the autumn term 2012 onwards.  

Chair: Thank you both very much indeed for giving evidence to us this morning.

Examination of Witnesses

Witnesses: Leighton Andrews, AM, Minister for Education and Skills, Wales, and Chris Tweedale, Director of Schools and Young People Group, Welsh Government, gave evidence.

Q288 Chair: Good morning. It is an unusual and pleasant experience to have a Minister from Wales coming before this Committee. Would you like to make any opening remarks, Mr Andrews?  

Leighton Andrews: I am very happy to go straight to questions, Mr Stuart.

Q289 Chair: Okay, great. Can you tell me about your relationship with the Secretary of State for Education in England?  

Leighton Andrews: I would be very pleased to meet the Secretary of State again. I met him in June 2010, shortly after he became Secretary of State. I first met him when he was a television reporter in Scotland in the late 1980s or early 1990s. I have always found him very charming and engaging but, as I say, both the Northern Ireland Education Minister and I would very much like to meet him to discuss these three-country qualifications.

Q290 Chair: What are you doing to improve ministerial relations?  

Leighton Andrews: We have made very great efforts to have conversations with the Secretary of State. The Northern Ireland Education Minister and I wrote in August last year asking for a meeting to discuss the series of announcements that he had made on GCSEs and, indeed, on A-Levels, which we tended to read about in the pages of the newspapers or heard about on The Andrew Marr Show. Given these are three-country qualifications, we felt—and I know that your Committee has asked—that there should be a ministerial conference to discuss these three-country qualifications. We think that would be very valuable indeed. I have to say I have very good relationships with Ministers in other parts of the UK Coalition Government, and certainly in the case of, say, universities I would expect to have regular dialogue, and indeed I have had several meetings, with Mr Willetts.

Q291 Ian Mearns: From what you have said there, do you take any implication from the Secretary of State’s stance? The fact is you have not met him since June 2010; you have asked for a meeting with him. Do you feel as though in some way he does not have any significant regard for the powers or the remit of the Devolved Administrations?  

Leighton Andrews: Well, there is a need for us to meet together as the three Ministers with responsibilities for jurisdictions that have GCSEs and A-Levels. I met with the Northern Ireland Education Minister to discuss this last year. We have had several telephone conversations about it, the last being last Thursday. As I said, we jointly wrote to the Secretary of State in August asking for a meeting. As I said, your Committee has suggested that there should be a ministerial conference on GCSEs and A-Levels on a three-country basis. There are divergences between Wales and Northern Ireland on the one hand and England on the other. Of course, the Northern Ireland examining body has withdrawn its GCSE and A-Level qualifications from study in England as a result, to quote them, of the “emerging policy differences between England and Northern Ireland”. We are now taking steps, as you are aware, to ensure that we establish different examinations in English Language, for example, at GCSE level. So there are issues here that require discussion. There are also issues, of course, that require discussion between the different regulators.
Q292 Ian Mearns: But regarding the divergence to which you refer, the Secretary of State is on record as saying that in itself is not a bad thing. I think I am right in saying that.

Leighton Andrews: Of course it is entirely for the Secretary of State to determine a course of action in England, but these are qualifications that are jointly owned. They are owned by the CCEA in Northern Ireland, by Ofqual and by the Welsh Government. Now, it is rather strange for us, therefore, to read announcements about GCSEs and A-Levels by someone who does not own the qualification.

Q293 Ian Mearns: So, in education terms, you think he is acting a bit like a high commissioner for education for Britain, dictating power to the Devolved Administrations.

Leighton Andrews: He is certainly not dictating to us. We have the policy autonomy to follow our own course of action, and we are doing that.

Chair: Mr Andrews is more than capable of making his own quotes.

Q294 Ian Mearns: Chair, I was merely suggesting a scenario that had sprung into my mind. The Secretary of State has described your intervention on this issue and on re-grading as irresponsible, mistaken and a regrettable political intervention in what should be a process free from political meddling. First of all, I would guess one of your answers would be you did not start it.

Chair: Let him answer.

Ian Mearns: Was your decision based on regulatory or political concerns?

Leighton Andrews: I thought the Secretary of State’s remarks were inconsistent, inflammatory and uncharacteristically ignorant when he said that, because of course I was acting as the regulator in Wales. That is why he was inconsistent. He made a great play in his evidence of saying that he could not comment on matters that were properly for Ofqual and then sought to denigrate the action that had been taken by the Welsh regulator.

I received a report from my regulatory officials in September. It was a sober and serious report. It made it clear that the outcomes of the results for candidates in Wales were unsafe and that we needed to take action. We therefore took that action. Let me say, on a political basis, since the politics of this has been raised, the action I took has been endorsed by three of the political parties in the National Assembly for Wales, so it is not a partisan matter. This is a matter of trying to achieve fairness for candidates in Wales on the basis, let me say, of comparable outcomes, and I would be very happy to discuss the way in which the notion of comparable outcomes has changed over recent years.

Q295 Ian Mearns: You asked for a meeting with the Secretary of State last August.

Leighton Andrews: That is right.

Q296 Ian Mearns: I take it you have asked again for meetings. What has been the response to that?

Leighton Andrews: There has been recent correspondence between me and Mr Gove, and John O’Dowd, the Northern Ireland Education Minister, and Mr Gove, in relation to further recent statements he has made about qualifications, which again we read about in the newspapers or were asked about by the media on the back of the announcements. So you can assume that on the back of that correspondence, we would certainly welcome the opportunity to meet with the Secretary of State.

Q297 Ian Mearns: Of course, not everything that appears in the newspapers necessarily emanates from the Secretary of State himself, although he probably does not know where it does come from.

Leighton Andrews: I could not possibly comment, Mr Mearns.

Q298 Ian Mearns: The re-grading of GCSE English Language candidates in Wales means that candidates in Wales will have received a higher grade than their counterparts in England for the same work. Do you consider that fair?

Leighton Andrews: It was action that we took to deal with an unfairness, and if you look at the judgment in the High Court, of course, what they were talking about in relation to Ofqual was how you address a balance of unfairness, to use their phrase. Everybody must accept that 2012 was unhappy all round, and we took action that we felt was appropriate. Ultimately, in the region of 2,300 candidates were re-graded as a result of the direction we gave to the WJEC.

Now, I would have to say—it goes to the question that was asked by your colleague earlier on—people will ask what the difference is between a Welsh C and an English C, but then in England they will ask what the difference is between a January C and a June C, so there are issues here. 2012 was clearly an unhappy round, and we took action that we felt was appropriate. Ultimately, in the region of 2,300 candidates were re-graded as a result of the direction we gave to the WJEC.

Q299 Ian Mearns: Given the portability of the jobs market between Wales and the rest of the UK, and particularly England, you have said that your intervention secured the fairest outcomes for learners in Wales, but do you think there are likely to be any long-term consequences for the currency value and portability of GCSEs achieved by Welsh candidates?

Leighton Andrews: Portability is very important, and we have just had an independent review of qualifications in Wales, which involved people from the further education sector, the higher education sector, employers and others, and that has charted a course forward for our qualifications system in Wales. We want to build on our existing Welsh Baccalaureate but also on GCSEs and A-levels.

One of the things that we asked that group to look at was portability, and they certainly looked at how our qualifications would be seen, for example, by universities across the UK. The evidence that we have is that university admissions tutors have to cope with entry qualifications from a whole series of different
countries. Scotland of course has an entirely different set of qualifications; there are international qualifications and so on. But portability is important to us, and that is why we had an independent review of qualifications involving employers and higher education people.

Q300 Ian Mearns: In a nutshell, though, in terms of where we began, would I be right in thinking that you would welcome direct contact with the Secretary of State rather than having communication directed at you via the media?

Leighton Andrews: Indeed.

Q301 Chair: We have heard from the awarding bodies that came before us today that they all told the Government in 2009 that the way the proposed GCSE qualifications in English were being constructed was a mistake, and yet they were all ignored. As a co-owner of the GCSE brand, did you say anything to the then Labour Government about the fact that they were going down the wrong track?

Leighton Andrews: Our officials certainly felt that 60% assessment was an issue. We need to be clear: there was, in fact, of course, a regulator called Ofqual back in 2009. I have minutes of the work that was done on maintaining standards in changed qualifications, with representatives of Ofqual not only attending those meetings but chairing them. That is indeed where the first consideration of comparable outcomes comes in, back in 2008. So the discussions would have been between the regulators at that point.

Q302 Chair: Yes, although the regulator at that point was not independent of Government. That was not until the passing of the Apprenticeships Bill.

Leighton Andrews: Well, there was an interim regulator, as one of your previous witnesses said, and it appeared to be called Ofqual.

Q303 Chair: But as a co-owner, I just want to explore that a little more. If your experts were telling you there were concerns about it and all the awarding bodies said that it should not have gone ahead, and if that has been a key component in the problems we had last year, trying to get to the bottom of how on earth it happened is important and, as a co-owner, you are partially responsible for allowing it to happen, are you not?

Leighton Andrews: I was not the Minister at the time, so I do not have full insight into everything that was going on, but my understanding is that as a regulator we made clear our concerns about the high level of that form of assessment. Indeed, we have now acted, of course, in respect of English Language for the cohort that began in September 2012, to ensure that the level of controlled assessment is 40%, not 60%.

Chris Tweedale: Representing the officials, our subject officer for English Language at the time, going back to the origin of the specification, raised concerns about the 60% controlled assessment, but the officer within Ofqual, which at that time was within the QCA, which was an NDPB, overruled that.

Chair: A non-departmental public body, aka a quango.

Chris Tweedale: Indeed. So we did make representations at that time, but we were overruled by Ofqual as it was then.

Q304 Chair: Do you accept that the re-grading intervention has made it more difficult to maintain common standards across the borders, and will perpetuating a split standard by having separate exam papers and grading arrangements for GCSE English Language candidates in Wales help this situation?

Leighton Andrews: We are seeking to ensure that in future learners in Wales are not subject to a Key Stage 2 indicator that has clearly failed to provide comparable outcomes in 2012 for learners in Wales. Key Stage 2 indicators are not being used in Northern Ireland and they are not being used in Wales, so we are certainly not going to put our candidates in a position where future judgments about grading are made on the basis of a Key Stage 2 indicator, and that is what we have done. Now, it is entirely possible, of course, that we could end up with tougher standards in Wales.

Q305 Chair: Yes, but given your intervention last year, employers and universities will think that unlikely.

Leighton Andrews: I do not see why that is the case. The issue here really is the coherence of the exam system, the curricula and the relationship to the accountability mechanisms. Now, there are serious issues in relation to the current relationship between grade-setting, as has been adopted by Ofqual, and the accountability systems that you have in England. We have seen that explained in the letters from the Ofqual Chief Executive to the Chief Inspector of Ofsted and to the Secretary of State in August—letters that we only became aware of after they were published in September.

Q306 Chair: You have said that Ofqual is in the grip of an ideological spasm. Given that language and the Secretary of State’s language, what is the future for three-country qualifications and their regulation?

Leighton Andrews: The jury is out on three-country qualifications. There is a serious issue and it goes to the issue of comparable outcomes. When comparable outcomes was designed as an approach to ensuring that, when qualifications changed, a cohort did not suffer by being the first cohort to sit that new qualification, it was really about ensuring fairness to candidates. Indeed, the quote from those minutes was, “The most important factor is a fair outcome for candidates.” These are the minutes of the meeting, chaired by Ofqual in 2008, that came up with the comparable outcomes framework. Since then, Ofqual, particularly in the last 12 to 15 months, appears to see comparable outcomes as a mechanism for containing grade inflation. That, it seems to me, introduces an entirely new element into the approach, as a result of which we had grade deflation, both for A-Levels and GCSEs, last year.
Q307 Alex Cunningham: Have you any idea why that has happened though? Why have Ofqual all of a sudden gone down a completely different track?
Leighton Andrews: When you held your pre-appearance hearing for the Chief Executive of Ofqual, she made it clear that containing grade inflation was her overarching priority, so I can only assume that she has established containing grade inflation as a key performance indicator for Ofqual.

Q308 Chair: You said that you thought there had been grade deflation last year. The Ofqual evidence that we have seen suggests marginally the opposite. Admittedly, it gets into the complexities of comparable outcomes and the datasets you use, but if you control for that, the 2012 GCSE English results were slightly more generous, if you like, on a comparable-outcomes basis compared with 2011. You have asserted the exact opposite.
Leighton Andrews: I would ask you this: why is it that the Higher Education Funding Council for England predicted there would be more AAB grades at A-Level than there were at the end of the day?

Q309 Chair: We have not looked into the detail of that. We have looked at GCSE English.
Leighton Andrews: It seems to me it is a relevant consideration, because you found a situation in England last year where quite a number of higher education institutions were clearly expecting rather more AAB entrants than they eventually got. You have to look at this in the round. There is no question but that Ofqual over the last 12 to 15 months has seen dealing with grade inflation as its priority. That is set out very clearly in the letters to the Secretary of State.

Q310 Chair: It is a statutory duty on them in the same Bill that set them up.
Leighton Andrews: I go back to this point about the original conception of comparable outcomes in 2008. The quote from the then Chief Executive of Ofqual was, “The most important factor is a fair outcome for candidates.” You have to contrast that with what was said to my regulatory officials over the summer by the Chief Executive of Ofqual: that fairness is not a factor she has to consider because it is not set out in the legislation governing Ofqual.

Q311 Chair: I take you back to the point that you thought there was grade deflation. None of the data we have seen bears that out. We have not looked into the A-Level, although it could be said that the reason why universities would over-predict the results is that, given prior inflation and the fact that they kept being caught out by more people getting it than was expected, they necessarily took a conservative view. If you bring an end to grade inflation, you will start to see an adjustment in universities being able to predict more accurately what is going to happen.
Leighton Andrews: These probably require more time than we have, but let us just say, in relation to A-Levels, this is a factor. The question is, if you are trying to contain grade inflation, should you do it all in one year? That is an issue that has come out in the course of this.

Q312 Chair: We are looking at GCSE English in 2012 and you have said you think there is deflation, and none of the data we have seen bears that out. I am asking you: on what basis do you make that assertion? It is nothing to do with A-Levels: GCSE English 2012.
Leighton Andrews: I will take you to the results of the GCSE English in Wales, where very clearly we had a very stable cohort over a series of years. We had a stable series of entrants, and if you look at the results, we suddenly saw a 3.9% decline in the results in Wales. There were clearly not comparable outcomes on the basis that we had been expecting, and that is down, I think, to the introduction of the Key Stage 2 indicator, which has acted as a deflator.

Q313 Alex Cunningham: Are your comments in relation to Wales rather than England, Wales and Northern Ireland?
Leighton Andrews: There was a decline in results in England too, but there we are.

Q314 Chair: No, there were not declining results in England according to the Ofqual data. There was that raw change of whatever it was, 1.4%, but they said that when you had gone through the process, disputable though that might be, if anything 2012 was slightly more generous. Is it not possible to read in that in fact Welsh students were falling behind English students, and that was one of the reasons why, because you are the Government rather than just a regulator, you came to the conclusion you did?
Leighton Andrews: No. You have seen from the WJEC evidence, and indeed some of the evidence that Ofqual published before Christmas, that we had fewer early entries in Wales. We had fewer examples of re-sits. We had far more examples, almost twice as many in Wales, of a linear approach to the taking of these qualifications than in England. The conclusion of the WJEC when it gave evidence to the National Assembly for Wales was that if a similar entry strategy had been followed, on balance, by schools across Wales, there would have been little difference between Wales and England in the overall outcomes.
We took the decision that we did in terms of re-grading explicitly on the basis of the regulatory report. There were clearly not comparable outcomes in Wales compared with the previous year. We analysed the data on the cohort for 2012 against the data on the cohort for 2011, and there was no reason to make a judgment that that cohort should have performed worse than the previous cohort.

Q315 Chair: It still seems, to me at least, a thin argument. England is obviously much larger than Wales. You could have gone through and found all those schools that were similar to the Welsh schools and had taken one structural approach as opposed to another, and gone around allocating, marking their points up. Surely the danger is that it undermines the value that is perceived of Welsh qualifications, and you must have weighed that very carefully before you came to the decision you did.
Leighton Andrews: We certainly weighed that judgment very carefully indeed. At the end of the day,
as I have said, 2012 seems to me to have been an unhappy experience for everyone. There are clear differences for thousands of students in England as to the grade they got in January and the grade they got in June. Everybody knows that is the case; it is confirmed, indeed, in the High Court judgment. We took a different form of action. One of the advantages of the action we took is that the situation was resolved very early on; it was resolved in mid-September. There has been great transparency in the process. I have been answering questions on it in the National Assembly virtually every week ever since and been through committee hearings on it. So our process worked, but it worked in the context of an overall situation that everybody, including this morning, has acknowledged was an unhappy one.

Q316 Siobhain McDonagh: You have said that your current regulatory system allows swift resolution of injustice. Are you satisfied that the new regulatory model will also allow this should problems arise in the future?

Leighton Andrews: We are looking to move to a system very like the one they have in Scotland, which is the model of the Scottish Qualifications Authority. We think that that will provide a system that people cannot argue with in terms of independence. I certainly accept that people are able to argue that at the present time. It will also, of course, end any suggestion that in the context of competing exam boards there is, to quote the Secretary of State, a race to the bottom.

Q317 Siobhain McDonagh: This overlaps with your answer, but is there not a strong argument for the regulation of exams to be independent from Government and also from the body that awards grades?

Leighton Andrews: The question of whether it needs to be independent from the body that awards is the one I would focus on. In Scotland, they clearly have a model that works and they have very clear dividing lines between the awarding end of the Scottish Qualifications Authority and the regulatory end. It is important, of course, that we preserve that in any structure that we create. There is a not entirely the same but quite similar model in Northern Ireland too. We want to be in the situation where there is consistency for learners, teachers know where they stand, and there is proper dialogue within the sector on the development of curricula and assessment and the regulation that goes with that, and that model is one that we can learn from.

Q318 Chair: Just to be clear, are you saying that fairness to students, which is of course an important but nonetheless vague concept, is more important than controlling grade inflation?

Leighton Andrews: I am not sure I am saying that. I do not think anybody wants to see grade inflation.

Q319 Chair: So you explicitly do not want to see grade inflation.

Leighton Andrews: I do not think anyone particularly wants to see grade inflation. You want to see consistency in standards. The issue is if teaching is improving year on year, then you would expect to see growth in the percentage of young people getting good qualifications. So you want a system that allows those improvements in teaching and learning to be expressed.

Q320 Chair: What if teaching is not improving year by year?

Leighton Andrews: If the teaching is not improving, you need to address that through your accountability measures, which we do.

Q321 Chair: That may be harder if there is a presumption and political pressure to make sure the results make it look like that is happening, and is that not a contributor to grade inflation? Most of the independent experts who have analysed the data and have come before this Committee, despite all the heat and light around this issue, suggest there has been grade inflation, which you have said you do not want, and there are political pressures around that have led successive Governments to allow it to happen.

Leighton Andrews: That is a very fair question, and I heard the evidence given by the representative of the AQA right at the beginning, who said there are issues relating to the accountability process and the examination process, and certainly we would accept that. That is why in our grading of secondary schools, unlike, as I understand it, the previous league-table system in England, we have not only a measure that looks at the percentage of young people who get five good GCSEs, to put it crudely, but also the capped point score and other measures too.

Q322 Alex Cunningham: You have made reference to the very different system north of the border—the Scottish-English border, of course. Michael Gove has said he does not regard an increasing divergence in qualifications policy between England, Wales and Northern Ireland as a bad thing in itself. Do you agree with that?

Leighton Andrews: I do not regard it as a bad thing in itself. What I do think would be helpful, in a situation where we have commonly branded and commonly owned qualifications, would be more ministerial dialogue when major changes are likely to take place.

Q323 Alex Cunningham: So do you see a situation where there is still the potential that Wales could decide to do their own thing, to have a different system, as they have in Scotland and similarly in Northern Ireland, and separate themselves because of the difficulties that you have faced?

Leighton Andrews: We were going down our own road of a review of qualifications in any case and had embarked on that around 18 months ago. We already have a Welsh Baccalaureate in Wales. We are looking to grade that qualification from this year. Clearly, Northern Ireland is also having a qualifications review at the present time, but the Northern Ireland Education Minister and I both agree that we have confidence in GCSEs and A-Levels.
Q324 Alex Cunningham: You have mentioned the need to improve ministerial relations. You have also said that the current three-country arrangements and structures are becoming untenable due to unilateral announcements from Westminster about shared qualifications. What more do you think you could do to try to improve that situation?

Leighton Andrews: I have met with the Northern Ireland Education Minister. We have had several telephone conversations on this issue. We have both sought, through the concordat arrangements that exist between our Government and the UK Government, to try to ensure we get better advance information. However, more dialogue would be helpful, and I certainly think a ministerial conference of the kind requested by this Committee would be helpful.

Q325 Alex Cunningham: So your message this morning to the Secretary of State in England is to say, “For heaven’s sake, sit down and talk with us.”

Leighton Andrews: Indeed.

Q326 Alex Cunningham: What steps are you taking to work with Ofqual to strengthen the GCSE and A-Level brands to ensure future comparability of standards across the board?

Leighton Andrews: We are in regular meetings with Ofqual. My regulatory officials meet with them on a very regular basis. These issues are considered all the time.

Q327 Alex Cunningham: Finally, the WJEC has increased its market share in England in recent years in some qualifications, particularly GCSE English. We see the Northern Irish exam board have said they are no longer going to offer a GCSE qualification in England. Should there be some sort of containment? Should the Welsh exam boards be sticking to Wales or should we have cross-border examinations and offer a free market?

Leighton Andrews: The current arrangements were developed in the 1990s. Clearly, an organisation called the Welsh Joint Education Committee, as it was originally, evolved to develop qualifications in Wales. It has been very successful in attracting custom from over the border and it has a very high reputation, both in Wales and in England. It also has a high reputation elsewhere.

Chair: Thank you both very much indeed for coming and giving evidence to us today. I am sorry, Chris, we did not hear more from you but, as we have seen, the Minister is more than capable of answering these questions. Thank you very much.
Tuesday 26 March 2013

Members present:
Pat Glass (Chair)
Neil Carmichael
Alex Cunningham
Bill Esterson
Siobhain McDonagh
Ian Mearns
Chris Skidmore
Mr David Ward
Craig Whittaker

In the absence of the Chair, Pat Glass was called to the Chair.

Examination of Witnesses

Witnesses: Glenys Stacey, Chief Executive, and Dennis Oppiss, Director of Standards, Ofqual, gave evidence.

Q328 Chair: Welcome, and thank you again for coming along to answer questions before the Committee. I think this is our final session, you will be pleased to hear, so thank you for coming along. It is clear from the decision in the judicial review that the judge agreed with Ofqual that the problem was not July; it was the results in January. However, Ofqual is still responsible for the results in January, so what went wrong there? Were Ofqual asleep on the job in January?

Glenys Stacey: Not at all. The judge, during the court proceedings and, indeed, in the judgment is clear that the design of the qualification is at fault. He does agree with our analysis that we presented in our August Report and again in the November Report that, with the benefit of hindsight, one might say that some awarding in January was on the generous side. We have explained in both of our Reports that awarding was exceptionally difficult in the awarding sessions before June 2012. There were three awarding sessions—two in 2011 and then one in January 2012—but still 80% of awarding took place in June 2012. The difficulties were really around the paucity of information, the paucity of examples of student work that could be looked at, and of course, with a new qualification, you have not got work from previous years that you can use as comparison.

In short, as was recognised at the time and since, awarding in a new qualification that is modular, when you have got small proportions of students in the early units, is exceptionally difficult, and a large amount of judgment is at stake. You ask whether we fell down on the job. I would say not. We were aware of the risks in awarding GCSE English, and we had a programme of scrutiny. We were observing awarding meetings; we saw awarding working, and we saw it working as it should.

Q329 Chair: I hear what you are saying about it being exceptionally difficult. However, there are lots of experts around. Did no one see this coming?

Glenys Stacey: The difficulty of awarding? The position is that design rules were set for this qualification back in 2009. It was one of the last qualifications to go modular. I am told that English and Maths were in that final phase, but the vast majority of the 70 or so GCSE subjects went modular the year before. It is true to say that technical experts—in my view, late in the day—raised concerns about the difficulty of awarding in modular qualifications, modular GCSEs. It was not particularly English, but modular GCSEs. By the time those concerns were raised, moderation was already in full swing and, indeed, those concerns—as I understand it from the record—were considered, but no change was made, because one really would have been undermining entirely the concept of modularisation. That was the concept at the time.

Q330 Bill Esterson: You just said that from the observations of the work in January, you were confident the system was working as it should. If that is the case, why were the grades wrong?

Glenys Stacey: The grades were thought to be right at the time. There were very few examples of work, because there was a paucity of students actually submitting work. My recollection, and I will check this, is that for the controlled assessment modules for the bigger provider, AQA, it was just 2% or 3% of the students that actually submitted in January. When you are then trying to work at a particular mark, you have a paucity of examples of that, and you have nothing to refer back to. Normally, it takes several years for awarding to get established, particularly in a subject like English with a new qualification. We are certainly advised by experts that in that situation, making those sorts of judgments, it takes a while with a new qualification to get them right, and it is much harder if there are fewer examples of work submitted.

Q331 Bill Esterson: I am just trying to tie up the difference between the comment that it was working as it should and that ultimately you did not get the results right. I am struggling to really tie the two up.

Glenys Stacey: Yes. I suppose what I would say then—and I know you are wishing to say something, Dennis—is that it worked as well as it could.

Dennis Oppiss: I think the great advantage of the awards in June 2012 was that by that time there was some fairly robust statistical information that the awarders were able to use. If you go back to the earlier series, January 2012 and the two series in 2011, the statistics they had at that point were weak, and the reason they were weak was because you did not have the whole population entered for the exam; you had a small sample. You did not know how representative that sample would be, so the statistics did not help very much there. The people doing the awarding did
not know how well prepared or how mature those candidates were, and there is something about how the effect of aggregation, adding the units together, which you need to know to have a full statistical grasp of the award. If you do not know how those aggregation effects work, which you cannot do at an early stage, then that also does not help you with your predictions. At those early stages, the statistics really were not helping the awarders very much at all, and so really those awards were based purely on the judgments of the awarders on the quality of the work. If you do not know how those candidates were, and there is something about how well prepared or how mature those awarders were, not helping the awarders very much at all, and so really those awards were based purely on the judgments of the awarders on the quality of the work they were seeing.

Q332 Mr Ward: There seem to be a lot of “don’t know” there. Why didn’t somebody know that they did not know? Why didn’t somebody say it would actually be dangerous here? These are people who are experienced, certainly in terms of statistical analysis and representative samples. Why did somebody not say, “Actually, to do something at this stage would be really dangerous, and we are not prepared to do it.”?

Dennis Opposs: To be clear, the early awards were based on the examiners’ judgment. The examiners were told, “These are not good statistics; you cannot go with them.” If we go right back, as Glenys was referring to earlier, to September 2009—just as the bulk of new unitised GCSEs were starting and a year ahead of the English and maths starting—we did have quite a thorough discussion with the technical people from the boards: “There are these new unitised GCSEs coming along. What are the risks in the aggregating?” We went through some of those and talked about whether there were some things we could do. We talked about whether we should stop there being any assessments before the final summer, and whether we should make the results along the way provisional, so you could issue marks rather than grades. There was also some discussion around changing what is called the 40% terminal rule. There is a rule in GCSEs that you have to take 40% of the assessment at the end, but we could have changed that.

So we discussed with the boards and the other regulators the pros and cons of all of those. At the end of that discussion, we decided not to proceed with any of those, and in particular that second option of making the results provisional—issuing marks but not issuing grades. It was thought that was not a way to go, partly because it would undermine confidence in the system and partly because in a lot of the GCSEs there were short-course GCSEs that were being run, and students would be taking the short-course GCSEs after one year. That would have made things very difficult in those subjects, because you had to issue full GCSE short-course results at that time. We did thoroughly explore that. At the time, the decision was made that no, we should not pursue that; we would go ahead as had been planned. That was then agreed with the exam boards’ accountable officers later that month in September 2009.

Q333 Alex Cunningham: Regardless of what was actually said in court, the system was seen as unfair. How can you guarantee that the exam system in future will be fair, and that we will not get thousands of young people distraught that they did not get the grades that they were led to believe they could expect?

Glenys Stacey: Thank you. The first point to make there is that there is an independent regulator, and there is no doubt in my mind that not having an independent regulator meant that it was much easier for these design rules to be set with curriculum drivers really at the heart of thinking, and in my view insufficient attention was actually given to standards, to awarding and to risks to standards. The first point I would make is having an independent regulator focused on those things makes a material difference, and is some assurance for the future.

The second thing is that we have already made changes to the design of GCSE English for the future, to protect us to some extent from these problems, so we are not having work graded from January. Controlled assessment and examinations are marked, but not graded. We will shortly be consulting on further changes we want to make to this GCSE in relation to speaking and listening, to take speaking and listening out of the grade result and to have it as a separate assessment and endorsement on the face of the GCSE certificate. It is very similar to arrangements that were in a decade or so ago. We have spoken with exam boards about the options here, looked at whether we can make speaking and listening assessment much more robust in schools, and there are no easy answers to that, so taking it out of the accountability pressures makes the qualification and the results more reliable.

The third thing is that we are reviewing controlled assessment in English and, indeed, in all other subjects. We will be reporting on that next month. We expect to strike a significantly different balance between internal and external assessment for some GCSEs, this one included. Finally, there are some very sobering lessons from GCSE English that will stand us in good stead as we set design rules for new GCSEs.

Q334 Alex Cunningham: So you are confident we are not going to see a repeat of what happened over the last year.

Glenys Stacey: The position is that this is just the second year of a new set of GCSE English qualifications, and they are still very different to the ones that were before. Therefore, we cannot guarantee that there will be no variations between schools; it is what you would expect.

Q335 Alex Cunningham: The moderation should pick that up, shouldn’t it?

Glenys Stacey: No, not necessarily. Variations will occur within schools for a number of reasons—for example, if teachers or schools have taken the opportunity with the new qualification to change
provider, so they are getting used to a different spec. or, if, indeed, they have got a new spec and they have not understood in sufficient detail the nature of the changes, or if they are making assumptions about the nature of the assessment, which we know did happen to some extent last time. Now, it takes more than one year for such a change to bed in, as I understand it, so we would expect to see some variations between the schools as well. We have taken two very significant steps, in our view, to protect against the situation we had last year. One is not grading in January, and the second is tightening the moderation tolerance to its limit.

Q336 Alex Cunningham: That is helpful. I would just like to take a step back and look at the value of the grades that young people received. Are they worth the grades they were given, both in January and in June, or is one of those sets of grades devalued?

Glenys Stacey: June awarding passes scrutiny. It passes legal scrutiny, so if you had a C grade from June assessments, it is a C grade. It is right. We have said that for the January assessments in some cases there was a generosity in awarding; certainly Ofqual, as it should, considered whether or not it should reopen those grade boundaries and have harsher awarding. We did not think that would be said, in retrospect, to be fair.

Q337 Alex Cunningham: Russell Hobby of the NAHT has commented that the fact the exam system cannot be managed without unfairness is a strong indictment. What is your opinion?

Glenys Stacey: I think for many good reasons, with many good intentions, GCSEs in particular have design difficulties. Modularisation, which has been a fairly late development in GCSEs’ life, has created problems. We can see that. Secondly, controlled assessment, when it is subjected to the current accountability pressures, can weaken results, and I quite understand the pressures there. Coming fairly new to these qualifications, I think they can be substantially improved, and the current design does undermine confidence.

Q338 Alex Cunningham: You just said that we are only in the second year, and all the changes and things are current. When do you think we will have a settled system, where people can actually have confidence in that system?

Glenys Stacey: I look to my colleague here, because I am not sure in recent times we have ever had a settled system.

Dennis Opposs: We have talked about what changes we are making for 2013. By summer 2014, English and all other GCSEs will not be able to be taken in a modular fashion, so students will be back to where they were a few years ago, which is that they have to take all of the assessments at the end of the course. That is another change. From the point of view of awarding, and all the issues that have surfaced since last summer, having all the awards at the end is easier. It makes it more straightforward for the exam boards to manage it, but that does mean that 2014 will also be a slightly different year. I would hope and expect that, once we are past that year, we should be in a position where there is more stability and we can be confident about what is happening.

Q339 Alex Cunningham: Then somebody else might step in and change it all again. The Times Educational Supplement said that details that emerged during the affair relating to GCSE English had not been kind to Ofqual or the exams system as a whole, and that a nasty taste has been left in many teachers’ mouths. What are you doing to restore public confidence, but also the confidence of teachers, bearing in mind what you have already said?

Glenys Stacey: As the Chairman has said, a judicial review of proceedings has shown our judgments to be fair, that we made robust and cogent decisions, and grasped the nettle when we did. Nevertheless, it has been a tremendous shock, I think, to some schools. You will know that many schools did well, but equally some schools did surprisingly badly, and they will quite understandably be feeling the pain of that and will wonder what is coming. It has become plain to us during the experience of GCSE English that not enough is understood about the complexities of awarding and standard setting—that our role is not well understood. The experience of English engendered a great deal of mistrust, and we have a job to do to explain as clearly as we can what we do, how we do it, and why we do it. We are starting that work now, in developments on our website and in our dealings with teacher representative groups. I think there is never going to be enough that we can say to explain what we do, how we do it and why we do it, and that is a key task for us now. That is really important.

Secondly, I think we have a great opportunity with the reform of GCSEs and the redevelopment of A-levels to get those qualifications to be the best that they can be. That will be our determined effort over the next few years, and I hope over time that fresh qualifications that are better designed, more worthwhile to study and more stimulating to teach will do something to restore faith in schools.

Q340 Alex Cunningham: You know the judge said that Ofqual was engaged in an “exercise of damage limitation”; whichever way you went, it was going to be a case of being unfair to someone. What do you think should be the overriding fairness consideration now, when you are regulating examinations?

Glenys Stacey: Fairness will be much more likely to be secured when we have qualifications that are not designed in a modular fashion. Awarding is more robust; it is more straightforward; it is easier and more fulsome and reliable when you are dealing with a linear qualification. I think that will make a very significant difference. We can be much surer about fairness overall. I do hope that the Committee Members understand that once we understood at Ofqual how these qualifications had played out in schools, we recognised the unfairness, and our board met in emergency session to consider whether we could make it more fair. We considered all of the alternatives; no-one has presented any better
alternatives than the ones we considered, and we could not make it more fair. Were we able to do so, we would not have hesitated to direct changes at grade boundaries, but that was not the solution. In fact, the fairest solution was to support a November resit, and that is what we did.

Q341 Ian Mearns: Given what the judge said, and given the boards’ own recognition of unfairness, do you think there has been a conflict between standards, combating grade inflation and fairness?

Glenys Stacey: No, I do not.

Q342 Ian Mearns: But you have said that the board has recognised that for some students it has been an unfair outcome.

Glenys Stacey: The unfairness really fell from the complexity of the design of the qualification, the multifarious routes through the qualification. It seems wrong to me that a student’s results should depend to some extent on a route chosen through the qualification by his teacher. That cannot be fair to the student or indeed to the teacher having to make those decisions. I think that was an element of unfairness.

The work that we do at Ofqual to manage grade inflation we have been doing since before my time. We have done it at A-level and at GCSE successfully. It is not, in my view, the culprit here. What we did in relation to grade inflation this time was, as you know, to challenge results from WJEC and from Edexcel. That is our job. It is our job to make sure that there is comparability between exam board results in a market model, and indeed the evidence has shown that our challenge was entirely right. Even with our challenge, they were out of line still to some extent. Without that challenge, they would have been very significantly out of line, and that would not have been fair to students this year, last year or in subsequent years.

I know it is difficult for many people to accept, but it is right that we have a view to inflation when we are looking at provisional results from exam boards, and it is right, in my view, that we ask them to tell us where their results are above or below a certain tolerance level, and we ask them to explain why that is. That is the process and, as we said in court, there are 42 GCSE awards in the last two years where we allowed inflation, or increase in results, above tolerances, because we believed it to be right. Now, these judgments are difficult, and we are always looking for evidence, because the system on its own will inflate.

Q343 Ian Mearns: I think, from that perspective, we are dealing with the situation we have found ourselves in now, and found ourselves in then, but there is an element of hindsight in coming to those judgments. When these youngsters started down the route of trying to qualify for this particular qualification, the qualification itself must have been judged by somebody to be fit for purpose at the outset.

Glenys Stacey: The qualifications were designed to meet design rules set before Ofqual existed, and they were accredited by Interim Ofqual; that is, in effect, a sub-committee of QCA. It would have been very odd indeed if the qualifications were not accredited, as they were in line and absolutely compliant with the design rules. In my view, the design rules gave undue weight to curriculum considerations and did not give sufficient regard to the complexity of awarding or to standards.

Q344 Ian Mearns: Is it Ofqual’s contention, then, that there have been systemic problems in terms of the way in which the youngsters going through these qualifications have actually been managed through their course?

Glenys Stacey: It is certainly a factor that we mention in our November Report. There were any number of routes through; I believe there were thousands of routes through these qualifications. There were many shared units. You could sit units at different times. You could sit or re-sit and re-sit. That is a level of complexity that no teacher ought to have to navigate through, and we know from our subsequent analysis that some students fared better than others because of the route they took. That, in our view, cannot be fair.

Q345 Craig Whittaker: Gareth Pierce from WJEC has suggested that you relied too heavily on a single predictor model approach. Is he right?

Glenys Stacey: We use a comparable outcomes approach. It is dependent on Key Stage 2 data, which is regarded as a reliable data set, and the evidence is that this is an efficacious approach across the whole cohort. We have traditionally agreed with WJEC, and indeed the Welsh regulator, that there are alternative data sets, one of them being common centre data, where you are looking at reliable schools, if you like, and seeing year-on-year changes. It is a data set that we know the Welsh regulator and WJEC are particularly keen on.

This year, I think for the first time, because WJEC have a significant market share in English and because, for their English qualification, 65% of the students were sitting them in England in English schools, we asked WJEC to report to us both against the Key Stage 2 data and against common centre data, so we had two sets of data. I do not regard that as being unduly reliant on Key Stage 2 data when I am seeking to ensure a common standard for all standards that take WJEC English and 65% of them are in English schools.

Q346 Craig Whittaker: So he is wrong, then.

Glenys Stacey: I would take a different view to Gareth.

Q347 Craig Whittaker: Okay. Can you explain how comparable outcomes differ from a norm-referenced approach to grade awarding, and does the model produce a quota of As, Bs and Cs that exam boards must meet?

Dennis Oppos: Perhaps let me start by taking a step back and talking about the nature of the exam system, because this seemed to be the subject of much discussion back in the autumn in particular. People were saying GCSEs norm-referenced, GCSEs criterion referenced. Just to be clear what we mean by those terms, a norm-referenced exam is one where you simply compare students with how other students have
performed, and you decide up front how you are going to provide the result. You might decide, before anyone has taken the exam, that 10% are going to get an A, 20% are going to get a B. In a sense, it does not matter how well they perform; it is a comparative. IQ tests are like that. They are about comparing yourself with a standard population. In norm-referencing, there is no concept of reaching a particular standard.

In criterion referencing, it is entirely different. You are comparing students’ performance with pre-determined levels. How well you do compared with other students does not then determine your outcome. Criterion referencing sounds like it is great, and it is where it works, but it is very hard to make it work, because the challenge is in writing and defining those criteria as carefully as you can. In practice, with qualifications, it turns out that you can describe those criteria pretty well when you are talking about such things as building a brick wall or something like that, but by the time you are describing how to mark an English literature essay and what to look for in it, it becomes much more difficult in that sort of context. In GCSEs and A-levels, which are closer to my English literature example, we do not have those sorts of criteria.

We have things called “grade descriptions”. Grade descriptions give you an indication of what a typical candidate will get. We have a grade description for grade C at GCSE, for example, and it says, “A typical candidate at grade C will be able to do these things.” But because it is about a typical candidate, it does not mean that is what you have to do, and part of the system allows you to say, “Well, actually, this student could not do that bit, but they were better than required at this other bit.” There is a compensatory system.

So GCSEs, A-levels—going back decades, certainly, for A-levels—are systems that are not norm-referenced or criterion-referenced; some people say it is called “attainment-referenced”. When the exam boards sit there in their awarding meetings and set their boundaries, they use a mixture of qualitative and quantitative evidence. At the heart of the system, certainly if you go back through the history of A-levels, is a kind of rule of thumb that says if the cohort of students taking the exam has not really changed much since last year, you would expect the results to be about the same. That has always been there.

Comparable outcomes, to come back to the point, which Glenys said were first used in 2009 with the new AS awards, have been the version that has been introduced, consistent with that system I have described but putting one or two additional checks into it. Really the idea is to avoid advantaging or disadvantaging one cohort of students over another. The reason that is important is that, if you have got a group of students taking GCSEs or A-levels, they may well be competing for jobs, university or FE places, or apprenticeships with people who took the exam in the year before or the year after. What we are trying to do is avoid some advantage or disadvantage just in a sense because you happen to be born in a particular year.

Now, what we know is that when an exam is run for the first time, teachers will be far less familiar with the requirements, because it is new for them. They will not be as able to prepare their students well for that exam as in the previous year, when they were preparing them for an exam that they have perhaps been familiar with for five years. In that first year, if you are a student taking it in that first year, your teacher will not be quite as well equipped to train you for it—to educate you to face the assessments. In those circumstances, the performance of students is likely to dip slightly compared with the previous year. If the judgments were made purely on the performance of students, that year would get disadvantaged, because that group of students would end up with slightly worse grades.

Q348 Chair: But we know that all year groups are not the same.

Dennis Opposs: No.

Glenys Stacey: That is right.

Q349 Chair: Most year groups are comparable, but every three or four years we get one that is different. I know, because I can remember ringing secondary heads and being told, “There is a Year 4 group coming through; we do not know why, but they are more disruptive, more difficult. We have absolutely no idea.” Every three or four years, they are different.

Glenys Stacey: That does happen at school level; it does not tend to happen at national level. It is a different thing. Going back to the question, norm-referencing is beguiling because it is simple. In fact, some people would say it is crude. The issue one would have with it is that it does not take into account these cohort changes at a national level, when there are changes of people coming out for IGCSE, for example. It certainly does not take into account any changes in the qualifications, and it does not really accommodate a multiplicity of exam boards. It is not quite the perfect solution that we are all looking for.

Q350 Craig Whittaker: Is it fair to say, then, that the comparable outcomes model is probably the best of a bad job?

Glenys Stacey: I would say that is little more condemnatory than would be fair. I think comparable outcomes is regarded by assessment experts as the best possible model at the moment, and it looks good when compared with how these things are done in many other countries. We are very good at knocking how we do things in this country, but actually there is no known better way internationally. It is not perfect; it is not in our view possible to get an absolutely perfect system; but at the moment, it is as good as we can get it. That does not mean that we do not want to keep that under review and always consider these options and have a healthy debate about that.

Q351 Bill Esterson: Thank you for the explanation and the comparison. I think that will be very helpful for the inquiry. Just picking up on Pat’s point, are you saying that the benchmark, which effectively is Key Stage 2, year by year, across the country, sees little variation at all, ever?
Glenys Stacey: Historically, these things tend to move very slowly. It tends to be glacial changes in achievement across a whole cohort. That does not mean that you could not have a more substantial change.

Q352 Bill Esterson: Sorry to pursue the point, but has there been a year where that has not been the case?
Glenys Stacey: I am not aware of one.
Dennis Opposs: If you look at the Key Stage 2 results going back through the last decade or so, there are years where the results in a subject suddenly go up quite sharply. Sometimes, they then go down quite sharply the year after. Science, in particular, seems to be prone to that.

Q353 Bill Esterson: Is that GCSE results?
Dennis Opposs: That is Key Stage 2 results. They have gone up and down. If you look at the general trend, that is different, but from year to year there have been significant changes.

Q354 Bill Esterson: And that would imply a relative increase in GCSE results for that year if that happened.
Glenys Stacey: That is why using Key Stage 2 data, we think, is a solid approach. The real test for comparable outcomes in our view is whether exam boards and examiners can present evidence of achievement above and beyond expectations. When that happens and we are able to say, “Yes, that is true, that has happened; it is there and evidenced, and therefore results will go up this year—more students will get results,” and people believe that there has been a real increase in attainment, that will be a happy day.

Dennis Opposs: Just to add, going back to last summer, it was the first year of a new exam. English language was a brand new exam; there had not been English language before. It would be very difficult under those circumstances to say, “We have got this performance in these new exams; we have got some performance from the year before, and in the case of English language, it was not even under the same title,” and actually look at those and say “Yes, we can really see genuine evidence here of an improvement.” That would be very hard in those circumstances.

Chair: Thank you. Can I ask if we can have more succinct answers, otherwise we are not going to get through everything?

Q355 Craig Whittaker: Glenys, you have just said to us that the national changes are quite glacial in how they move, but in August last year you wrote to the Secretary of State and you said that comparable outcomes can make it harder for any genuine increases in the performances of students to be reflected in the results. Does this mean that containing grade inflation comes at the expense of reflecting genuine improvements?
Glenys Stacey: Not at all, no. We just need to see evidence of it. That is the nub of it.

Q356 Craig Whittaker: Do you not think that is disheartening for teachers and pupils?
Glenys Stacey: I would hope that teachers and pupils would wish to see grades that truly reflect achievement. Within a comparable outcomes approach at national level, individual schools and students can do exceptionally well or exceptionally badly. You get those variations underneath the national pattern. Indeed, we saw that at extremes last year. For individual students and schools, achievement can rise and fall significantly depending on effort, skill and the expertise put into it. I would hope that they would also accept that there needs to be a regulator that is looking to see at national level whether results reflect true improvement at individual school level, or whether there is something else at play.

Q357 Craig Whittaker: I suppose the point to emphasise is that even at national level last year the variation between years was very marginal, between the English GCSEs.
Glenys Stacey: Thank you for raising that. When we looked very carefully at it after awarding—obviously we see provisional figures during the awarding, but it is not until everything is done and dusted that you can see the picture fully—our analysis was that results nationally in English fell by 0.3%. We reported that in November. We stick with that and believe that, and that is well within the normal range of variation. Actually, it is a credit to comparable outcomes.

Q358 Neil Carmichael: Good morning. Leighton Andrews has written to the Committee and has talked about the fact that he spotted a 5% drop in the proportion of candidates in England achieving a grade C between 2011 and 2012. First of all, is that something that you recognise?
Glenys Stacey: Well, I wish Leighton Andrews had written to me to tell me that at the time he might have discovered it. It is certainly news to me. I do not recognise that; I refute that.

Q359 Neil Carmichael: Right. If you did recognise it, what would you describe it as?
Glenys Stacey: I would describe it as very hard to justify or explain. It does not compute. What I would be happy to do—I mean, we can expand on it if you like, but I know the Chair wishes to move on—is to write to you with the detail of how and why we refute that.

Q360 Neil Carmichael: That would be interesting. That would be really useful, because that is something that we have got on the file, so to speak, that needs an answer.
Glenys Stacey: It might be helpful to have something on the file now.
Dennis Opposs: We know that the Department for Education publishes a variety of figures. Although taken separately they all seem very clear, they each tell a slightly different story. For example, back in August when the GCSE results came out—and the GCSE results looked the worst—it was pointed out that there were 23,000 candidates who seemed to have disappeared from the figures since the previous year.
What we now know is that the entry for IGCSE went up from 5,000 to 20-odd thousand between 2011 and 2012, so there is a whole load of students who got their grade C through that route who will not necessarily be reflected in all the DfE figures. Sometimes, that gives a misleading impression. That is one aspect of why the statistics are never quite as straightforward as they might be.

Q361 Neil Carmichael: Well, Disraeli had a word or two to say about statistics, didn’t he? It would be helpful to this Committee if you could write to us.

Glenys Stacey: Certainly. There are several reasons why you are not comparing like with like there

Neil Carmichael: I fully accept that. I have dealt with statistics myself many times.

Chair: If you would care to write to the Committee explaining that, that would be really helpful. Thank you.

Q362 Neil Carmichael: Thank you. Now, another matter that I think is important and may well also be controversial is of course the Interim Ofqual, which began operating in April 2008. The question arises: who was responsible for the qualifications development when the current GCSEs were designed: the Interim Ofqual or another part of QCA? That is the first question. Going on from that, what role did the Interim Ofqual play in the design of the GCSEs?

Chair: No, Dennis wants to say something.

Glenys Stacey: I was just looking back at our November Report, as it was before my time. As I understand it, back in 2007 the then-Government decided to establish an independent regulator to extract that responsibility from QCA, which was the curriculum authority and had regulatory authority as well. That was done by a sub-committee of the QCA board being created, and it was chaired by a member of the QCA board. It was allowed and in fact encouraged to brand itself as the Interim Ofqual, but it remained a sub-committee of QCA. It was bound by a memorandum of understanding as between Interim Ofqual and QCA. It was obliged by those memoranda to follow the strategic direction of QCA, which would include, of course, unitisation and modularisation. The design rules for new GCSEs were proposed in 2009. That was Interim Ofqual and not full Ofqual. Ofqual was established on 1 April 2010. These qualifications were accredited by Ofqual and not full Ofqual. Ofqual was established on 1 April 2010. These qualifications were accredited by Ofqual and not full Ofqual. Ofqual was established on 1 April 2010. These qualifications were accredited by Ofqual and not full Ofqual. Ofqual was established on 1 April 2010. These qualifications were accredited by Ofqual and not full Ofqual. Ofqual was established on 1 April 2010. These qualifications were accredited by Interim Ofqual—i.e., a sub-committee of QCA. It would have been very odd for that not to have happened in the circumstances.

Q363 Neil Carmichael: Given that relatively complex history, are you happy about the traceability of responsibility that essentially arises from that?

Glenys Stacey: It is perhaps not for me to have a view as to how Ofqual was created. What I am happy about is that I understand it, and I have seen the documentation. I have looked at those memoranda, and I understand what happened. It is the job of the regulator, in investigating what has occurred here, to look back and understand, and account for how these things happened. That is what we have done.

Q364 Neil Carmichael: Do you think there are any lessons to be learned from that period between Interim Ofqual and actual Ofqual?

Glenys Stacey: I am sure there are.

Dennis Opposs: When Glenys says Ofqual was set up as a sub-committee in its interim form, legally it was a committee of the QCA board that was set up. Of course, it was operating within the QCA legal framework, so Interim Ofqual only had the same powers that were there in the 1997 Act, which made it a very different organisation with much more restricted powers than Ofqual when it was set up in independent form in 2010.

Q365 Neil Carmichael: Of central importance to this question is that essentially that relatively complicated structure all happened before the difficulties we are talking about. I am really wondering if there is any link that you can think of that we should be exploring, or that you would like to add something to.

Glenys Stacey: Certainly, from my perspective, it is uncomfortable on the one hand to allow a committee of a board to brand itself as an entirely different organisation and then not give it the power and authority to be that organisation. One would wish, in a way, to have a cleaner establishment. I would say that would be a clear lesson from the past.

Q366 Neil Carmichael: That would certainly be the conclusion I would draw. That is basically it. Thank you. No, Dennis wants to say something.

Dennis Opposs: The reason there was that very unusual arrangement, whereby Ofqual was going to be set up but we then had this interim period was because the Secretary of State was very keen to get it set up as fast as possible, and of course the legislative timetable would not allow that to happen for a couple of years, so that was why it was set up in interim form.

Neil Carmichael: So legislative justifications are behind this, essentially.

Chair: We have never seen an occasion where Government gives responsibility to one organisation and power to the other, have we? It has never happened.

Q367 Siobhain McDonagh: Exam board representatives told the Committee that they raised concerns about structural features of the current English GCSE, especially the high proportion of controlled assessments during the qualification design phase. Andrew Hall of AQA expressed regret that they were “not anything like forceful enough” in making their point heard. Why do you think that the concerns raised by the exam boards about the high proportion of controlled assessments were not acted upon?

Dennis Opposs: We have to go back to 2007, I think, to answer this. Back in 2007, when the plans were being laid for the new GCSEs, the QCA executive—so this is before the time of Interim Ofqual—were considering proposals for the new GCSE subjects, and they decided that the different amounts of controlled assessment that were being proposed in the different subjects could not be justified. They decided that what they wanted was a solution whereby subjects either
had no controlled assessment or 25% or 60%. From that point on, that effectively became a rule for the new GCSEs. That was the origin of that. Beyond that time, all subjects seemed able to do was to decide which of those three weightings they went for.

The discussions about English would then have taken place through 2008, but the development of GCSE English and the consultation on the criteria was all in the main part of QCA. Interim Ofqual was only responsible for picking that up at the end of the process in the autumn of 2008. If there were discussions that were going on, it was with QCA rather than with Interim Ofqual.

Q368 Siobhain McDonagh: Did organisational change mean that the regulator took its eye off the ball with regard to qualification design, or was there confusion about roles and responsibilities? There is an overlap with Neil’s questions.

Glenys Stacey: You are thinking about Interim Ofqual and then Ofqual—that change.

Siobhain McDonagh: During the design phase.

Glenys Stacey: It is a difficult question for me to answer, but you may have one.

Dennis Opposs: I can try. If we are talking about 2008, then there were clear roles in terms of what QCA did and what Interim Ofqual did. That is not the same split as there was once Ofqual was set up in 2010. The arrangement in 2008 was that it was QCA who was responsible for developing the criteria, for consulting on them and for finalising them, and then passing them to Interim Ofqual. It was not Interim Ofqual’s job at that time to keep a close eye on, or be responsible for, the development of the subject criteria. Interim Ofqual’s job was to pick up those criteria at the end and then sign them off. Actually, the issue that Interim Ofqual then raised at the end of that process was to do with functional skills. At the time, there was a Government policy based on the 2005 White Paper that said that to get a grade C in English or in maths, you had to pass the Level Two Functional Skill. That was still there in the criteria in autumn 2008, and that was the issue that Interim Ofqual took to ministers. That was the key question at the time.

Glenys Stacey: I do not know whether it helps, but it occurs to me that, yes, 1 April 2010, Ofqual became Ofqual; in September 2010, it established a scrutiny programme for GCSE English, which we have reported on in the November Report. We had an eye in September 2010 to the particular risks of awarding in GCSE English; the evidence that I have seen suggests that there was one eye on it.

Q369 Neil Carmichael: That was your eye.

Glenys Stacey: It was Ofqual. It was before my time.

Q370 Neil Carmichael: Ofqual’s eye. I wanted to probe this a little bit more, in terms of areas of dispute that might have arisen between Interim Ofqual and QCA, who would have adjudicated and what kind of relationship really existed between the two organisations, given the complications and descriptions you have just come up with.

Glenys Stacey: You do have me at a disadvantage, because I joined Ofqual two years ago. I did not live through that.

Q371 Neil Carmichael: I realise that, but I think this is a really interesting area. It is prior to, but clearly influential of, the situation that subsequently occurred.

Chair: Dennis, you were there at the time. Come on, tell us what happened.

Dennis Opposs: It was quite a challenging time.

Chair: I think that is nicely put.

Dennis Opposs: We were trying to create an organisation that was being separated from its parent, so to speak, at the same time as moving from London to Coventry. There were some real challenges in all of that. At the same time, the top priority was to make sure that we kept our eye on all the balls that were spinning in the air.

Q372 Chair: Were there turf wars at the time?

Dennis Opposs: I think the relationship with the parent body, so to speak, became quite difficult. I am not sure it was that difficult in 2008, but as we got closer and closer to vesting, it did become more difficult. I am not sure that it was particularly difficult in 2008. Back in 2008, as I recall it, the senior people in QCA were very supportive of the creation of this new organisation.

Q373 Neil Carmichael: At the back of my mind—and certainly Siobhain’s, and possibly Pat’s—is that the system you have described does sound complicated. Buck could be passed, responsibilities not necessarily seized, and there were eyes on various balls—and you have said there were quite a lot of balls. Is there a chance that the difficulties really began to become enshrined in that period?

Glenys Stacey: I think there is a commonly understood position, and that is that if you are going through very substantial organisational change, particularly creating new organisations, you have to manage that alongside the day business. There is bound to be a risk, because you are imposing an additional demand on your very senior managers. There is bound to be a risk, because you are managing all of that rather than just the business. No one would say otherwise. You would expect me to say this, but I genuinely think there is a very strong case for taking regulation out of the curriculum authority and having an independent regulator. I cannot imagine having the two in one place.

Q374 Neil Carmichael: I think that is a very fair point, and we have learned that lesson in a number of different areas of public life.

Glenys Stacey: Yes, thank you. There is always a case for regulation to be focused on regulation, and not to be distracted with other, more exciting, areas of endeavour.

Q375 Chair: To be fair, Glenys, I think that none of us disagree with the decision. We think that was the right decision. I do not think that any of us disagree. It appears that the way in which it was handled may have made many challenges. What we are trying to
establish here is whether the origins of what happened to lots of young people in this country lie in the challenges of that organisation coming apart.

Glenys Stacey: I do not think so, personally. From the records I have seen—and I have gone to the archives and got the written records out of key decisions made around the design of GCSE English and its accreditation—I have seen no evidence to suggest that any turf war or difficulty of creating the new organisation actually has relevance to those decisions.

Q376 Neil Carmichael: Dennis, what are your thoughts on that?

Dennis Opposs: I would agree with that. There was a lot of attention and care taken in the original design. I think it would have been very difficult, thinking back to 2008, to imagine the sorts of scenarios that we saw in schools in 2012. There was little, almost no, experience of modular GCSEs at that time, so it would have been very difficult to look forward and think how in these circumstances schools might manage this particular modular exam.

Neil Carmichael: Thank you.

Q377 Chris Skidmore: I wanted to pick up something you have mentioned previously about the common desire to have the 0%-25%-60% rule for controlled assessments. Ofqual, in its second report, described how that aspiration overrode in-house technical assessment and regulatory concerns. What does the report mean by that?

Glenys Stacey: Again, it was before my time, but my understanding is that the subject community was concerned about raising the proportion of controlled assessment in English from what had been 40% to 60%. The record shows that they had considered whether you could move to the 25% threshold, but thought that you could not squeeze the curriculum assessment of that properly into 25%. They were between a rock and a hard place: was it to be 25% or 60%? They went, reluctantly, to 60%.

Q378 Chris Skidmore: It still, as you said, overrode concerns. In terms of the steer, was it the steer from the then-Government that took priority?

Glenys Stacey: It was not a Government steer.

Dennis Opposs: I think it was very much a decision that QCA Executive took in 2007, which I presume must have been agreed with Wales and Northern Ireland at the time. That was the point where this 0%-25%-60% rule was set for all GCSEs.

Q379 Chris Skidmore: The reason I am asking is that I want to reflect on today and Government steers of today. Michael Gove wrote to Glenys, I think on 6 February, outlining his policy steers regarding GCSE reform, stating that qualifications should be linear, taken at the end of the course, and his concerns about the current system of tiered papers, where pupils forced to choose between higher and lower-tiered papers places a cap on ambition. I am sure you are familiar with the letter. In terms of lessons from the past, and in particular GCSE English, and moving forward, what lessons are there here for the relationship between the role of Government balancing the independence of the regulator? Do you feel that you still would be able to act independently, regardless of government steers? I know you are on the record as saying that the goal of a single-tier GCSE is quite a challenge. Do you feel that you have the autonomy to push your own agenda?

Glenys Stacey: I have some statutory objectives to which you will hold me to account. You expect Ofqual to achieve those and to put those first; the statutory framework requires us to have regard to Government policy, but not to slavishly follow it. That is the clear position. As it happens, some of the most troublesome aspects of the design of GCSE English were nothing to do with Government. The modularisation of GCSEs, as I understand it, was QCA's decision, and controlled assessment proportions, the up-rating of that for English, was again a curriculum-authority decision. A NDPB made those decisions, and not at the behest of Government. But they were made with good intent, and they were made before anyone, I imagine, could predict quite the strategic interventions that schools now undertake to enable students to give of their best up to C grade. I do not think the past really helps me, in a sense, with that.

What does help me is to understand entirely the responsibilities of Ofqual, and that includes determining, to a large extent, the design of an assessment. It is Ofqual's job to determine whether or not new GCSEs will be tiered or whether there is a better arrangement that could ensure that these qualifications remain universal, as the Secretary of State has asked, and remain as accessible to the same proportion of students as they are now. There is a lot of research on tiering. Experts do take the view that it serves its purpose but there are downsides. One of the clear downsides is that it does cap ambition for some students who are channelled early into lower-tier papers where, generally speaking, the best you can hope for is a C grade. Experts recognise that, as does the Secretary of State. We have a job to do to look at whether tiering is the best approach, or whether there are other ways in which it can be done. Just let me finish.

The important point here is that we know that is our job, and we will make the right decision about what is the best known way to ensure these qualifications remain universal and accessible. It may be tiering; it may not, but we have a job to do to get to the right solution.

Q380 Chris Skidmore: Just on the point of tiering in particular, do you feel that you have the required assessment expertise in place to understand the nature of gaming the system in a way that clearly was not in place several years ago? Down the line, looking at it from a completely objective point of view, if you removed tiering, you could create a comparable situation to what has happened with the GCSE English exam, where there would be several thousand pupils who may have gained a grade C on the lower proportion tier who when put into a universal exam no longer gain that grade C.

Glenys Stacey: Yes. Absolutely.

Chris Skidmore: Are you looking at that in the light of GCSE English?
Glenys Stacey: Absolutely so. I think one of the key lessons from GCSE English is that decisions about the design of the qualification and the detailed design of assessment need to be made in the real world, with a recognition of not simply what is regarded as best assessment practice but how that might play out in the real world of schools. This is why we are so keenly interested in the Government’s current proposals for accountability. For example, if the threshold grade measure is still to remain central to accountability in mathematics and in English, we will have to take that into account in the design rules for the qualifications. Could I just say that assessment experts themselves, I think, were surprised to find quite the extent and sophistication of some of the techniques used in schools in the current day? Assessment expertise is one thing that we need, and certainly the Select Committee recommended, when you reported on assessments for 15 to 19-year-olds, that we should build our expertise. We are; we now have a Standards Advisory Group that is really playing a full part for us there. This is not just assessment expertise; it is actually being on top of what is happening in schools as those qualifications play out.

Q381 Chair: It is entirely possible that this Committee may look at the issue around AS-levels. What do you think are the important lessons that have been gained and that we need to learn around exam design, and that we can take forward?

Glenys Stacey: I think first of all being sure about the purpose of the qualification. It sounds so straightforward, but it is really important to know exactly the purpose of the qualification. We are somewhere with that now with A-levels; I think that is the first thing. The second thing is, when you are looking at purpose, we understand from the Secretary of State that the purpose of AS level will be to ensure curriculum breadth. I am sure you will wish to consider how curriculum breadth is assured.

Thirdly, we do recognise that with the design of new A levels and the decoupling of AS, there is a real opportunity to look carefully at what is a good A level. We could give you some examples of that; we probably have not got time, but we would welcome a chance to brief you on that. If you look at any particular subject and how the AS is embedded, if you are taking it out, how does that A level become a better, more worthwhile qualification to study? It can be done, so I think that is the real prize in what is now to happen: making sure that we have got first class A levels and that AS remains still a valuable qualification, albeit decoupled.

Q382 Bill Esterson: How confident are you that, if your evidence and your proposals differ from the policy steers of Government, your view will prevail?

Glenys Stacey: I have some evidence of that, as we are no longer preparing to accept English Baccalaureate certificates in some subjects, but we have the really important task of designing new GCSEs in all subjects, so there is some evidence that our views are taken seriously.

Q383 Ian Mearns: Do you think there are any further changes needed to the current English GCSE to protect standards, or are you happy for them to run as they are until the first youngsters take these exams in 2017?

Glenys Stacey: I think “happy” would be overstating it. We are about to consult on some further changes that I outlined earlier: taking speaking and listening and ensuring it is certificated separately but on the face of the GCSE certificate. I think that will relieve the pressure on 20% of the assessment, which I think will be worth while.

Q384 Ian Mearns: So, if not “happy”, content, satisfied? Mildly unhappy?

Glenys Stacey: I will be very pleased to see new GCSEs in English. In the meantime, we need to nurture this qualification and to protect it. It has weaknesses in the design. We have to protect it, and to protect standards. We are doing, if you like, some remedial work. We are propping it up by not allowing work to be graded, and by looking again at how speaking and listening is reported. We have met with exam boards and talked through other options, but they are very difficult to implement and bring their own problems with them. It is the best that we can do to see this qualification through, but I will not be unhappy to see it go.

Q385 Ian Mearns: You have used an interesting word there; it rings bells in my mind. My dad used to work in the pit, as a face worker, and he used to talk to me at length when I was a child about how important it was to get the propping right. The whole process of propping—can you just talk to me about that again? You are talking about propping up something that is already there, but trying to modify it as we go.

Glenys Stacey: Well, if I can get out of the mine for a moment and use another analogy that is easier, what has happened in GCSE English is that we set a bar for the performance standard, and that bar must be set and the student must be able to get over it. If you are doing the limbo, that is a bit easier than going the other way.

Glenys Stacey: What happened in GCSE English, if you like, is that that bar bent a bit. Some students got easier than going the other way.

Ian Mearns: If you are doing the limbo, that is a bit easier than going the other way.

Glenys Stacey: What happened in GCSE English, if you like, is that that bar bent a bit. Some students got a bit of a leg-up to get over it. What we are trying to do is to keep that bar straight, and to limit the opportunities for some students to have a leg-up, because that is unfair to all students, and to make sure that bar stays straight come what may.

Ian Mearns: I think we have got the bar thing now.

Glenys Stacey: I need to keep it straight. I think what we have put in place gives us good assurance that we can do that, but there will still be some variation school by school; we must expect that, because it is still a relatively new qualification.

Q386 Ian Mearns: Some exam board reps have favoured a proposal that speaking and listening should no longer contribute to the final grade in English GCSEs, with only recognition on the exam certificates. What would be the likely implications of such a change, and what would happen to the 20% of
Glenys Stacey: You are not the first person to ask it, and I will contemplate that.

Q390 Ian Mearns: Thank you very much. Now, to change tack somewhat, how would you describe your relationship with the Welsh regulator and his officials at the moment? Are you doing anything to improve these, or is everything hunky-dory?

Glenys Stacey: It is understandably testing, given the experience of last year and the Welsh regulator's decision to re-open awarding for Welsh students with WJEC. We are doing something about it. Again, I recollect when you reported last year you suggested that we should work with the Welsh regulator to reach agreement as to how standards are set. We have done our level best to do that, but we have not come to an agreement as to how comparable outcomes should be set. We have real concerns, particularly in GCSE English, as the Welsh regulator has decided and indeed directed WJEC to set different examination papers for students in England and Wales in the summer. That concerns us greatly, and indeed I wrote to the Welsh regulator on 6 March expressing that concern and asking to meet.

Q391 Ian Mearns: Leighton Andrews told us—in this very room, I think—that “the jury is out” on three-country qualifications. How do you see the future of three-country qualifications and regulation, or do you think it is becoming a thing of the past?

Glenys Stacey: We have at Ofqual agreed some principles for regulating qualifications that may cross borders, and I would be happy to send those to you, so that you can see what our approach is. We have, obviously, given those principles to the Welsh regulator. Our stance is that if a qualification shares a common title and is produced by WJEC or anyone else and is sat by students, wherever they are, it should meet a common standard. It is the standard that matters—I am sorry to keep reiterating it, but it is true. As qualification design diverges—so if, for example, in one jurisdiction, GCSEs are modular, and in another they are linear—the difficulty then is setting a standard, because those things affect standards. Our view is that if the designs of the qualifications are different, then there is a point at which it is not the same qualification. It cannot be said to be the same qualification and you cannot set a common standard. At that point, and we are fast reaching that point, you are then looking at how you differentiate those qualifications at awarding on the certificate and, ultimately, through titling.

Q392 Ian Mearns: You have talked about the two different sorts of exam there, and talked about standards. Do you think they are mutually exclusive when it comes to rigour?

Glenys Stacey: Rigour is not a word that assessment experts commonly use.

Q393 Ian Mearns: It is one that has come into common parlance because of the intervention of somebody above all of us at the moment.

Glenys Stacey: Yes. Politicians often use the word “rigour” when speaking to me about qualifications,
but not many politicians have explained to me what that actually means.

Ian Mearns: Some politicians.

Glenys Stacey: Yes. I do think a good-quality qualification has got three key component parts. Firstly, it has really good subject content—and you have views, I know, about how subject content should be determined. Secondly, it has first-class assessment methodologies; I question modularity at GCSE in that respect, and I question the proportions of controlled assessment. There are other issues as well, but those are big enough as it is. Lastly, its performance standard, where that bar is set, should be known, understood and justified, and standards must be maintained. That is a good-quality qualification. You might call it a rigorous one.

Q394 Ian Mearns: To go back to Wales briefly, an independent regulator is due to be established in Wales shortly. What impact do you think this is likely to have on cross-country qualifications and regulations? Do you think that will help the dialogue between the two country’s regulators?

Glenys Stacey: We very much welcome moves in Wales to establish an independent regulator. The model is slightly different from the one in England, in that, as I understand it, the regulator will also be designing and marketing qualifications. I understand they will wish to market those qualifications in England as well as in Wales. That is, as I understand it, again, fairly central to the financial modelling for the organisation. If qualifications are marketed in England and they are to be taken by English students, they will be regulated by Ofqual.

Ian Mearns: Thank you very much indeed.

Q395 Chair: I do not think we are going to get through the brief today, mainly because of hopeless chairing but also because some issues were too interesting to move on. Can we write to you with some of the questions we have not been able to ask you?

Glenys Stacey: Absolutely. We are happy to come and brief you at any time as well.

Q396 Craig Whittaker: I just want to take you back to the judicial review: 167 individual pupils, supported by 150 schools; 42 councils; six professional bodies; and, of course, the obligatory teachers’ unions supporting all that. Mark Dawe from the OCR has been quite critical. He suggested to us that you could have done more to have stopped the court case from happening and that your interim Report should have been more open about the issues that you were facing at the time. Do you agree?

Glenys Stacey: I would be interested to know exactly what the Committee think we could have done. We produced a Report within two weeks. We were looking at the concerns that were expressed to Ofqual. As we said in our final Report, we were focused initially on exam boards and awarding. Why wouldn’t we have been? We regulate exam boards. We were able to report, I think, in very quick time that awarding had happened as it should. That did not address all the concerns, and we then worked extremely hard to get to the root of them. Certainly, in the judicial review, the court has found nothing that is not already in our Report and has found no reason to disagree with any aspect of our evaluation of what actually happened.

During that period between August and early November when we produced our Report, we met with all representative groups, often on several occasions, to go through the emerging data and client data sets to understand fully the concerns. Indeed, we have met since, so I am not entirely sure what more we could have done, save for one thing. Very recently, I became aware that suspicions were raised in some quarters when it became known that we had challenged WJEC and Edexcel’s awarding. We had not actually covered that in our interim Report. Now, we knew we had done that, and we knew it was entirely proper; we were trying to find out what had gone wrong, not what had gone right, but I can understand now that people may think we could have covered it. I think, with the benefit of hindsight, we could have expanded and put a little bit in about our role and how we played it, but we did not really understand that to be at issue.

Q397 Craig Whittaker: Would that have stopped the court case, though?

Glenys Stacey: No. Interestingly, we did ask those that were to take proceedings to wait a few days until we produced our November Report, but they were unwilling to do that. I am not sure that even our November Report would have prevented the proceedings. I think there was a level of understandable concern, and indeed the court recognised that and did say it was a justifiable matter of public concern. I am not sure that even waiting for our Report would have stopped those proceedings.

Q398 Craig Whittaker: As we now know, the judge of course dismissed the application on all counts, but what he did say in his judgment was that there was no question in his view that the matter was properly brought to court. Do you agree with that?

Glenys Stacey: Well, as I say, there was a matter of public concern, and so I understood proceedings being issued. Of course, all regulators are judicially reviewed. It would be an unusual regulator not to be reviewed. From our perspective, it was disheartening to have arguments put during the course of the proceedings that suggested we had a secret policy or that we were fixing results. Of course that was disheartening, and we had to put a lot of effort into showing how we did things. Whilst we were putting effort into those matters, we were not able to do all of the things that we otherwise would have wished to do during the business year. Actually, I think that is probably all right and proper. There was a real public concern about this. It is clear that some schools did not do as expected, and they did not understand how that could possibly be, and they were mistrustful. I am not sure even our Report—not many people read it, understand it—would have appeased that concern.

Q399 Craig Whittaker: Is the court the right place to challenge the exam system, do you think?
26 March 2013  Glenys Stacey and Dennis Opposs

Glenys Stacey: There are a number of places where the exam system is properly challenged. This room is one of them, and a court can be another. As I say, regulators do get judicially reviewed. It is not at all unusual. I suppose what was unusual here was that we were judicially reviewed not by those we regulate, but those that we do not.

Q400 Craig Whittaker: Who then should hold Ofqual to account and intervene when things are thought to have gone wrong? Clearly from this episode you come out with a fairly clean bill of health, but who should be the ultimate people that hold you to account?

Glenys Stacey: We are accountable to Parliament. That is very clear in the legislation, and I recollect appearing before you very quickly after initial concerns were expressed. I have always been happy to appear here and account for what we do.

Q401 Craig Whittaker: You said earlier that Ofqual as an independent regulator is a key strength on-going, but in your November Report you said that regulation could have been more effective in some respects.

Glenys Stacey: Yes, I did.

Craig Whittaker: Are you satisfied that you are able to take a sufficiently impartial and critical stance regarding the regulator’s role in the whole affair?

Glenys Stacey: I can only do my level best, and I do not believe I am known for being a partisan or self-interested individual, otherwise I would not have been appointed here and I am sure you would not have endorsed my appointment in the way that you did. We do recognise that we could regulate better and certainly regulate more intelligently. We are working extremely hard to do that. Indeed, I put a business case in to the Secretary of State this morning for the necessary funding to make sure that we regulate as well as possible during a period of substantial reform.

Q402 Craig Whittaker: What has the cost of this whole episode been? How much has the judicial review cost?

Glenys Stacey: On the judicial review costs we were, as you would expect, very canny in what we chose to pay our legal advisers, so we will come in substantially cheaper than other parties, I believe. It is still several hundred thousand pounds, moneys that we had allocated for other business. As it happens, our costs have been awarded. I am thankful for that, but we did not have them in the year that we needed them.

Q403 Craig Whittaker: Do we have a total figure for the whole process for everybody involved?

Glenys Stacey: I do not, and probably there will not be at this stage. There will be a process of taxation to be gone through in order to come to those figures, but I am expecting other defendants to be coming in at a higher price than that.

Q404 Craig Whittaker: So several hundred thousand from your point of view, and several hundred thousand plus for the others. We are probably talking in the region of £1 million or whatever. Surely, this is not the best way to deal with disagreements. Surely, there must be a better way to put some trust back in the system, so that people understand very clearly what they can do and whom they can go to, without having to spend £1 million.

Glenys Stacey: I agree, and we are working hard at that. It will be a bitter disappointment to me if we find ourselves in court again because people do not understand what we have done. We are working very hard to make sure that there is much more transparency about what we do and how we do it, and how we set standards.

Chair: Thank you for coming along today. I do often reflect that we are never as swift as when we are chased, so hopefully Ofqual will come out of this a stronger, more respected organisation. Thank you for coming along.

Glenys Stacey: Thank you very much for chairing this session today.
Written evidence

Written evidence submitted by Ofqual

BACKGROUND

GCSE results were announced on 23 August 2012. On 25 August Ofqual announced that we would look closely at concerns raised by schools and colleges into outcomes in GCSE English and English Language. On 28 August we published a list of questions and answers about GCSE English 2012 which set out the concerns being expressed, the work we intended to do, and provided additional background information.

Ofqual worked with representatives of schools and colleges, exam boards and other stakeholders to gather and analyse evidence to inform its work Ofqual’s findings were published on 31 August in its initial report “GCSE English Awards 2012: A Regulatory Report”.

This note sets out the actions and meetings that have taken place since the publication of the initial report.

COMMUNICATIONS

The initial report was published on Ofqual’s website on 31 August along with a press release.

On 1 September we published Myths about Ofqual’s report into GCSE English 2012. This document deals with a number of common concerns, and misunderstandings, to help people understand better how things work.

HELPDESK

Since the release of GCSE results on 23 August, and up until 16:00 on 5 September, the Ofqual Helpdesk has received 46 phone calls and 270 emails relating to GCSE English results. Three calls have also been received which relate specifically to the November re-sit opportunity. All calls and emails are being dealt with within our usual turnaround times.

AWARDING ORGANISATIONS

On Monday 3 September, we met with the chief executives and responsible officers of AQA, CCEA, Edexcel, OCR and WJEC. The purposes of this meeting were:

— to give exam boards the chance to challenge our initial report and findings
— to make swift progress in establishing arrangements for resits and to consider how Enquiry About Results (EARs) are being handled across awarding organisations
— to discuss plans for GCSE English 2013

On Wednesday 5 September we spoke to the Chief Executives of AQA, CCEA, Edexcel, OCR and WJEC to clarify our expectations around the processing of EARs and the preparations required for re-sits in GCSE English.

STAKEHOLDERS

On Wednesday 5 September we met with representatives of NUT to allow them the opportunity to review the evidence which we used in reaching the decisions in our report, and to raise any concerns with us. Similar meetings were held with AOC, ASCL, IAA and NAHT on Thursday 6 September.

FURTHER ACTIONS

We are continuing to oversee exam board plans for resits. We need to make some changes to the detail of the usual regulatory rules around resits.

We will reach decisions soon about whether we should make any changes to the GCSE English arrangements for 2013, to secure standards.

We are actively considering two questions:

I. GCSE English A*-C results fell by 1.5% overall, in line with expectations given changes to the student mix. But some schools and colleges have been surprised by their results—with significant and unexpected variations as against their predictions. Why?

II. These were new qualifications. They were the first full suite of modular GCSE English qualifications and the assessment included controlled assessment, a relatively new concept brought in across the range of GCSEs to replace coursework. To what extent did these factors—or other factors—mean that some schools got the results they expected and others didn’t?

---

1 Some callers have also emailed and these figures are included in the total therefore it should be noted that there may be some overlap within this figure.
We are developing hypotheses, and we will be taking advice from our Standards Advisory Group on them, and the best methodologies for getting to the root of things. This work will take a little while, perhaps 4–6 weeks. We will publish a second report.

We have made additional data and analyses requests of exam boards, to help us get to the root of things, and we expect to make more. We will continue to liaise with representative groups of schools, colleges and students.

*September 2012*

---

**Further written evidence submitted by Ofqual**

**INTRODUCTION**

1. Further to the memorandum submitted on 6 September, this memorandum provides the Select Committee with additional information about our work on GCSE English awarding in advance of the oral hearing on 11 September. The body of the paper sets out our view about what happened with GCSE English awarding this year and summarises the issues arising. Annexed to the paper is a more detailed discussion of some of the issues we have been considering, and how we have been thinking about them.

2. We welcome the opportunity to answer Select Committee’s questions tomorrow.

**WHAT HAPPENED**

3. Overall, GCSE English results were down at A*–C by 1.5 percentage points. This was in line with expectations because of changes to the student mix. But there has been an unusual distribution pattern, a greater variation between schools than expected. And for some schools, the results are a far cry from their expectations. We have been looking carefully at what lies behind that.

4. There are three points it is important to make up front.

5. First of all, there has been no political interference.

6. Secondly, awarding and grade boundary setting worked as it should have done for the English suite, as for other GCSEs and A levels. We played our proper role, regulating standards, and using the comparable outcomes approach. Awarding in January and before was generous, but neither the exam boards nor the regulators could have seen it at the time. All the evidence at the time was that awarding decisions were if anything harsh.

7. Thirdly, GCSEs are modular and students’ overall GCSE results are influenced by the routes they take through the qualifications—which units they take, and when, and whether they resit.

**THE IMPACT OF CHOICES**

8. Route choices are generally made at school or college level, by schools rather than students. Schools make those choices using their best judgement as to how to get the best outcomes—to make sure that the students get the results they deserve. These choices are particularly important, particularly sharp in English. As is normal with modular qualifications, students’ results are influenced by the decisions their schools and colleges made about how to approach the qualification as well as by the quality of their work.

9. Some schools chose to put their students through all the modules at the end of the course—to put students through a modular qualification in a linear fashion, with all units sat at the end, in June 2012. The rules allow for this. We can confirm that some of those taking the qualification in this way did worse than they expected. They did worse than those who sat the qualification as it was designed, in a modular fashion, and noticeably worse than the small numbers who finished in January.

10. A small proportion put candidates through the English qualifications early, with final modules taken in January 2012. We are looking into that. Those students did comparatively well, and one of the reasons for that will be the grade boundaries for units sat in January 2012. We now know they were generous, but exam boards could not have known that at the time.

11. Candidates taking qualifications in a modular way will tend to have an advantage—they can see how they are getting along during their period of study, as unit results come in. They are clocking up points. And if they do not do as well as expected in a unit they can resit that unit at a later session. So you can expect them to do better, all other things being equal. That is the case for all GCSEs—indeed all modular qualifications—not just the English suite. And on average they did do better in the English suite, this time, with the effect being exacerbated by the generous grading in January and before.

12. So, some students did worse than expected. Expectations in some schools were raised because they saw the January grade boundaries when they were published by exam boards in March. Schools are told not to rely on those grade boundaries (though there has been some suggestion that some schools may have got different messages from exam boards, and we are exploring that), and they know they can change and do change, but we suspect they were relied on particularly here because the qualifications were new and there was little else
to go on. And, because they were generous, expectations were raised. We know of some schools expecting a 15 percentage point increase in achievement at Grade A*–C, for example.

13. From the work we have done so far using the data we have collected from exam boards, the difference in outcomes at Grade C between those who took a linear route and those who took the modular route is not particularly unusual and not out of line with what we would expect from analysis we have done in the past on this issue. We will do more work to test this.

14. We also know—from the information that schools give to exam boards—that schools were predicting that 15% more students would achieve at least a C grade, as compared to last year. Schools generally over-predict. In usual years they predict, on average a 12% increase at this level—so predictions, and expectations were more optimistic than usual.

15. We are turning now to school variations, and looking at those schools where the results were most out of line with their expectations. We know that the proportion of schools that had very significant negative differences is very small—less than one%. We need to know more, school by school, about what lies behind their results. We know that one consideration will be the student mix, and the proportion of students judged to be on the C/D borderline.

UNDERLYING CONSIDERATIONS

16. There are underlying considerations that are relevant, and that might be particularly relevant to some schools.

17. Firstly, controlled assessment. 60% of the English suite assessment is by way of controlled assessment, supervised and marked by teachers in schools. A third of that controlled assessment is assessing what are known as speaking and listening skills. Assessing speaking and listening skills is notoriously difficult, but speaking and listening is in the national curriculum. Most subjects have controlled assessment as part of the overall assessment at GCSE level. For most other subjects where controlled assessment is used, the proportion of assessment that is controlled assessment is 25%. For the English suite it is 60%.

18. Teachers mark controlled assessment. A sample is then moderated by exam boards, and exam boards allow a tolerance of 6%. We think that controlled assessment is problematic, and we are reviewing it in all subjects, but we can see that it is particularly problematic in English, and for speaking and listening.

19. Teachers received the moderated papers back from exam boards with marks confirmed, in the main. Some teachers assumed then that the mark meant the grade—by reference to January grade boundaries. It did not—but again, expectations were raised. We are looking closely at how controlled assessment worked for the English suite.

20. Secondly, there is a need to make sure that standards are maintained in a unitised qualification. 40% of the assessment must take place at the end, and this rule—known as the terminal rule—gives headroom to enable exam boards to get it right. We know that exam boards followed the rules of awarding. The three regulators played their part, querying and challenging provisional outcomes. We challenged two of the four providers—Edexcel and WJEC—to bring their qualifications in within acceptable boundaries. They did so—we did not need to direct them to do so. There is a suspicion that this put undue pressure on the grade boundaries in June. It did not. It is the regulator’s job to challenge outcomes, and we did. But there is a recognised tension, for modular specifications, between getting the unit grading right and making sure the qualification level outcomes are right.

21. Thirdly, these were new qualifications. The system is used to changes in qualifications, with GCSEs changing every five years or so. The changes in the English suite were not insignificant, and there were changes to entry patterns as well. Schools made choices, student by student between English, or English Language and English Literature, and some of the units were what are known as common units—that is, they could be used to contribute, ultimately, to English or to English Language. Schools could delay final decisions. All this made the job of predicting and awarding complex for exam boards.

22. Lastly, of all qualifications and grades, the most weighty pressures come to bear on English Grade C. It has the most significant role, above any other qualifications measure, in the way that schools themselves are assessed.

23. We believe that the combination of pressures has created particular strains and tensions, now evident.

EMERGING CONSIDERATIONS

24. At this stage some important points emerge.

25. Modularisation has increased the difficulty for exam boards of maintaining comparability across all specifications and pathways through English/English Language GCSE: this is problematic given the particularly high significance placed on Grade C in English (and mathematics) in accountability systems.

26. The ability to defer some of these decisions (ie choice of GCSE (English or English Language) as well as timing of unit entries and resits) until late in the course has seemingly intensified tactical concerns, as
schools gather intelligence about the apparent attractiveness of available pathways. This intelligence includes news from other schools about their successes and disappointments in individual units, as well as previous unit grade boundaries published by exam boards, and also affects school expectations of outcomes.

27. The increase in the teacher-marked controlled assessment (previously coursework) component of the qualification from 40% to 60% of the qualification may have contributed to very high school expectations of outcomes on the new English/English Language GCSEs. Many schools are disappointed, not because their results are down, but because they were not up more.

28. There is some confusion in schools between marking and grading, especially in the context of controlled assessment. Some schools appear to have believed that controlled assessments were graded to pre-set pass marks.

29. Accountability pressures encourage schools to focus on tactical decision-making about their choices from among the proliferating routes through GCSE English/English Language as well as on teaching and exam preparation.

30. The possible consequences of the extensive changes to English GCSE do not appear to have been fully considered at the time that the new set of English qualifications was designed.

Ofqual, 10 September 2012

Annex

FURTHER DISCUSSION

COMPARABLE OUTCOMES

31. In 2010 Ofqual adopted the “comparable outcomes” approach to securing standards in new GCSEs, an approach we had already started to use for the new A levels. The meaning of “comparable outcomes” is best understood from the perspective of the person sitting the exam. The basic principle is that, other things being equal, a young person should have the same chance of achieving a given grade, no matter which exam board their school uses and no matter which year they take the exam. The outcomes should be comparable.

32. The exam boards use data on Key Stage 2 attainment to judge the expect attainment of each GCSE cohort, for that majority of GCSE entrants who have a Key Stage 2 result (“matched candidates”). So the national predictions are adjusted if the mix of candidates changes. This year there were several changes in the entry cohort that affected the mix, as we noted in our report.

33. “Comparable outcomes” are used by exam boards as a check at qualification level. When considering what the grade boundaries should be, examiners see a comparison between the grade outcomes predicted by candidates’ Key Stage 2 test results and the proposed actual outcomes given their judgement about where the standard should be set. This testing of examiner judgement against statistical predictions is central to comparable outcomes, and has the potential to give us the best of both worlds. At AQA, for example, about 0.3% more of this year’s matched candidates reached Grade C or better in English/English Language combined than the statistics would predict. There are tolerance limits agreed between exam boards and Ofqual which allow for some divergence (typically ±1%), reflecting the fact that statistical predictions can never be wholly accurate. Beyond these tolerance limits, the exam boards can justify improvements in grades by reference to genuine educational improvement or deterioration.

34. This approach has been used for some years to improve comparability between exam boards, and also within exam boards to contextualise their judgements.

35. The change that has coincided with the introduction of this qualification is the strengthening of the requirement for exam boards to justify increases beyond tolerance to the regulator. This change was agreed in December 2010, in advance of the first awards of the new GCSEs in most subjects in 2011. This approach has also been applied in A level and AS awarding since 2009 and, in the A level context, has been the subject of a positive independent evaluation by NFER last year. We published details of our approach to try and help schools, colleges and others to understand what we were doing.

36. Questions have been asked about the appropriateness of the application of comparable outcomes in the first awards of a completely new GCSE. This is actually the point at which it is most important to be able to anchor examiner judgement. Their job is to carry forward over time the collective understanding of the level of educational achievement that deserves an A, C or G grade, and to interpret that in the context of changing assessments and other contextual factors. There is always the risk that performance will dip in the first year of a new qualification, as teachers get used to new curricula and assessment approaches, and comparable outcomes is intended to mitigate that—and then make sure that the standard is carried forward into future years as teachers get used to the new qualifications.

37. One of the biggest challenges in maintaining qualifications standards is to distinguish between improvements in outcomes which are attributed to genuine improvements in attainment—as a result of better teaching and learning—from those which arise from other factors, such as greater teacher familiarity with the assessment or changes to the qualification structure (sometimes called “grade inflation”). This is particularly the case when qualifications change. If there were a substantial change in outcomes at national level at the
38. Furthermore we know that in any education system, national outcomes change only slowly, even though variations may be quite significant at school level. In the context of English/English Language GCSE, we know that schools, teachers and the National Curriculum are all much as they were in preceding years. If the change was a fall in attainment, it may be that a lack of teacher familiarity with the assessment approaches or new subject content could explain the change. If a rise in attainment were reported it could be that the assessment structure benefited students—particularly, as with GCSE English in 2012, when there was a move from a linear to a modular structure, allowing for more retakes. For this reason, we think it is right to have a strong framework to enable challenge to and testing of examiner judgements in the first years of a new qualification to help carry forward standards.

39. We believe the comparable outcomes approach was the best way of securing standards year on year in GCSE English in 2012. We know that the exam boards found it unusually difficult to set the standards in the units awarded before June 2012. Given that, it was essential that there was a robust approach to qualification level awarding in the summer, to avoid grade inflation, and comparable outcomes provided a way of doing this. The evidence we have looked at so far suggests that the awarding process in the summer was done appropriately, with the right accountabilities and appropriate engagement with the regulator. This is, though, something we will be looking at further over the coming weeks.

**Maintaining Standards in Modular Qualifications**

40. The comparable outcomes approach is not specific to modular qualifications, but it helps to manage the specific challenges that arise with grading modular qualifications. These challenges are discussed in some detail in the 2009 paper by Isabel Nisbet, Ofqual’s previous chief executive, which has been highlighted in the media in recent weeks.

41. Qualification awards are made by adding up the awards made at unit level. There are no awarding decisions at qualification level distinct from the cumulative awarding decisions at unit level. That means it is critical that unit awarding is done consistently and accurately.

42. There are two issues in particular that need to be managed. The first is that, if the full benefits of a modular system are to be realised, then students need to know how they are performing in each assessment as the course progresses, so they know where to focus their efforts. This means that exam boards need to award grades whenever a unit can be taken (up to four times during a GCSE course), even though some of the earlier awards may be made on the basis of only limited information about expected performance, and with only small numbers of entries and a possibly unrepresentative cohort. This means that many of the checks on examiner judgement which would typically be available when awarding grades are not available. When qualifications come to be awarded at the end of the course, if the statistics lead examiners to revisit their judgements about where the standard should be set (as would often happen particularly with a new qualification) there is no opportunity to revisit the standards for units already awarded.

43. The second problem with a modular system is that different choices can be made about the routes through the qualification which can affect student performance. As a result, the students taking each unit can be a mix of school years and the ability profile of the entry can vary from one exam series to the next. We know this makes it more difficult for examiners to judge where the standard should be set (as would often happen particularly with a new qualification) there is no opportunity to revisit the standards for units already awarded.

44. There is a level of transparency in the GCSE system now which means that schools and colleges are generally in a position to make informed judgements about the most appropriate way of teaching and structuring the qualifications. Information about how awarding is done, the standards that are set, and the assessments that are offered is routinely available to schools, and—because of the pressures of the accountability system—we know that schools focus very heavily on understanding what they need to do to secure the grades their pupils deserve. In general, modular awarding at GCSE has been successful—for most subjects, new GCSEs were awarded in 2011 and this happened successfully; and in 2012, new GCSEs in subjects other than English were awarded without difficulty. But we recognise the significance of the issues with GCSE English this summer: the generosity of the awards in assessments taken before June 2012 has had a serious impact on perceptions of fairness at qualification level. If we were not already planning to remove modular GCSEs after the current school year, we think there would now be a strong case for doing so.
45. Some commentators have suggested that the June awarding must have been harsh to balance the generosity of the January awarding. Had that been the case, this would have further disadvantaged candidates who had taken the qualification on a linear basis. However, we have no evidence that this happened. If it had done so, we would have expected to see two things:

— Examiner reports at the time of June awarding consistently expressing concerns that the statistics were pointing to boundaries that were higher than examiner judgements would suggest they should be. In fact, as set out in our 31 August report, there is no evidence of that; as would be expected in a normal results season, there were a range of different boundary-setting decisions.

— Data suggesting that like-for-like outcomes were identical to the previous year. In fact, there were small (but within tolerance) increases in like-for-like outcomes in English (though the overall result was down as a result of changes to the cohort).

NUMBERS AFFECTED

46. It is very difficult to quantify the numbers of students affected by the issues with GCSE English awarding this year, because there are many reasons why candidates do not perform as expected, and more generally it is difficult to separate cause and effect.

47. There will also have been some students who got a D grade when they might have been expected to have got a C. As set out above, there are a number of reasons why this might have been the case, and it is impossible to disentangle them and estimate the numbers involved:

— Because they chose not to take assessments as they went along, so they could not resit units where they had done less well than expected.

— Because they took all their assessments in June, and therefore did not benefit from the generous grading in earlier assessments.

— Because schools made over-optimistic predictions about their attainment, perhaps based on incorrect assumptions that the January grade boundaries would be carried forward.

VARIATIONS AT CENTRE LEVEL

48. One of the issues that we will be exploring over the coming weeks is that there was an unusual level of turbulence in the results in GCSE English this year. Although year-on-year the results were very similar—as would be expected given our comparable outcomes approach—this masks significant variability, with some schools and colleges seeing results that were lower than the previous year’s, and lower than expected, and others seeing results that were higher. We need to understand this better and consider what it tells us about how the system is used and managed. It may be that some turbulence is inevitable with a new qualification—some schools will be better prepared for a new qualification, or will use approaches which suit some qualification types more than others, or make better choices about how to manage the assessments (for example, the timing of assessments). So we need to understand whether the levels of turbulence, and the reasons for it, are unusual for GCSE English this year. That will require analysis of results at centre level which we are now embarking on, and we will report further in our final report.

September 2012

Further written evidence submitted by Ofqual

This note is an update on Ofqual’s work relating to GCSE English and English language since the Chair, Chief Regulator and acting Director of Standards appeared before the Committee on 11 September 2012. We have set out our work under key headings but in reality these different strands of work have been part of a programme of work that has been a high priority and importance for Ofqual.

FINAL REPORT ON SUMMER 2012 AWARDING

Following our interim report on 31 August, we continued to investigate GCSE English. We looked more closely at exam board awarding and interviewed the Chief of Examiners in English at each board. We interviewed representatives from 100 schools and colleges to understand fully the concerns being expressed, and to identify common features and common approaches to teaching, learning, intervention and assessment. We required exam boards to conduct their own analyses of results. We developed systems in Ofqual to analyse data at school level (in the past our analysis has been limited to the national picture). We evaluated exam boards’ analyses as well, and we sought advice from experts in assessment.

In our final report, published on 2 November 2012 [http://bit.ly/TQ4m6p and http://www.ofqual.gov.uk/files/2012-12-10-gcse-2012-erratum.pdf], we confirmed the findings of our interim report: Grade boundary setting in exam boards had happened as it should, but nevertheless there was an unexpected variation in school results when compared to school expectations, and when compared to more established qualifications.
We were able to show that overall, results were almost unchanged, year on year. The issue was not overall outcomes, but variations, school by school. A significant minority saw large changes, about equally up and down. We went on to conclude that variable results were to be expected to some extent. We showed that results are always more variable when qualifications change, and we explained that these qualifications had changed significantly. This was the first time the new qualifications had been awarded. Some grade boundaries in January 2012 and in 2011 had been generous, but this had not been obvious to examiners or exam boards at the time.

We also concluded that the new qualifications were exceptionally complex and difficult to award, and especially susceptible to pressures, as teachers strove for the best possible outcomes for their students and school. We concluded that putting so much weight on one grade in one subject as part of accountability and performance measures created perverse incentives for schools in the way they marked controlled assessment and led to over-marking. It is our view that no teacher should be forced to choose between their principles on the one hand and their students, school and career on the other.

Fairness

It became clear during our investigation that there had been some unfairness, and that it was not possible to identify which students had been treated unfairly. Ofqual Board considered whether anything at all could be done to redress that unfairness. It considered all known options, including directing grade boundary changes, and concluded that there was no way in which to improve fairness: any action other than supporting an early resit opportunity, which was the option we announced in the August report, would have created greater unfairness.

Judicial Review

Legal proceedings were started against AQA, Edexcel and Ofqual by a consortium of local authorities, schools and students a few days before we published our final report. The claimants were not willing to await our report. The hearing was expedited, and the matter was heard in mid-December in a three day hearing.

The court delivered judgement in the New Year. It concluded that Ofqual had acted properly, that Ofqual’s decisions were cogent and rational, and that we had “grasped the nettle”. The claimants’ claims were dismissed robustly. The Court decided there was no case for Ofqual to answer. Subsequently, our costs and the costs of AQA and Edexcel have been awarded.

November 2012 Re-sit

With our agreement and support, exam boards offered resits free of charge in an exceptional sitting in November 2012. Over 46,000 students entered one or more units. 36,000 of those were students who had achieved a grade D in summer 2013. Approximately 30% (9400) of those students gained a C in the November re-sit.

Contrary to normal practice, we arranged for JCQ to publish a full set of results for these re-sits [http://www.jcq.org.uk/examination-results/gcses] in the interests of transparency.

Actions to Strengthen Future Exams

Our November 2012 report identified flaws in the way the qualifications are designed. We cannot quickly change some of those, but we have taken action, where we can do so safely, to protect the qualification in future exam series, as follows.

1. For November 2012 onwards we agreed reduced tolerances for moderation of controlled assessment marks—that is the difference that can exist between the teacher marks and the moderator marks without any adjustment being made.

2. We published a special condition for exam boards which prevented them from grading the January 2013 units in English and English language. This means that teachers will not know the January grade boundary marks when they are marking controlled assessment work for submission in June. Students who entered units in January 2013 received their raw marks on March 7th but did not receive a grade. Grading for both January and June units will take place in July 2013.

We and the exam boards communicated this to schools well ahead of the March results day. We scrutinised exam board arrangements for implementation of these new arrangements, as there were some system risks for exam boards to manage. We also scrutinised their communications with schools and colleges.

To date we have received 17 complaints into our helpdesk specifically about GCSE English since results were issued following January exams. Seven of these were from parents, five from schools and five from candidates. These are all complaints rather than enquiries. This is a comparatively low rate of enquiry and complaint, in our experience.

3. We had already implemented changes to make GCSEs linear in England for summer 2014. January 2013 was therefore the last January series for GCSE. In future all GCSE qualifications will be taken in a linear
way—all units at the same time—in June. The only exception we made to this was to make provision for a retake opportunity in November for English, English language and mathematics, because of the importance of these qualifications for progression to Level 3 qualifications or to the labour market.

We now know that the pressure to achieve a C in English and mathematics is so great that it sometimes distorts teaching and learning. So we have made clear that from 2014 the November exam series will only be available to students who are re-taking the qualifications. It will not be available for first time entry.

4. We are discussing with exam boards changes that might be made for teaching from September 2013. We know that there are particular challenges in moderating the teacher marks for speaking and listening. We are planning to consult over the next few months on changes that would mean that marks for speaking and listening did not count towards the overall GCSE grade. We have been working with exam boards to model the effects of that change.

**GCSE English Language in Wales**

The approach to awarding in summer 2012 had been agreed between the three regulators in England, Wales and Northern Ireland, as had been the case in previous years. In Wales in September 2012, after the results had been issued, the Minister directed WJEC to re-grade GCSE English language (GCSE English is not available in Wales) so that the overall outcomes matched those for Welsh students in summer 2011. This decision resulted in more favourable treatment for the 2,300 students in Wales (compared with students in England) who were upgraded.

The Welsh Government recently imposed on WJEC a special condition for summer 2013 that requires WJEC to hold separate award meetings for Wales students for January and June 2013, and to provide different question papers for students in England and Wales in June 2013. Different question papers will necessarily mean different grade boundaries. We do not think this is a satisfactory situation. It damages the credibility of, and confidence in, the GCSE brand for all students, but especially for those in Wales. On 6th March we wrote to the Welsh regulator to ask for a meeting to discuss it, but so far we have had no response.

We have been talking to the Welsh regulator’s officials with the aim of finding an approach to summer 2013 awarding that would provide us with the assurance that grade standards in WJEC English language in England and Wales will be comparable. To date we have not been able to gain assurance that common grade standards will be maintained.

To aid decision-making we have agreed a principled approach to regulating qualifications that may be subject to more than one set of regulations (see Appendix 1). If we enter summer awarding without the necessary assurance, we will make sure that it is clear to students and to users of qualifications that common standards are being set for these qualifications taken in England but that the standards in Wales are different. For the future there are likely to be increasing divergences between GCSE qualifications taken in England and Wales. We are working with the Welsh regulator and staff to make sure that users of the qualifications are not misled.

The brand GCSE is jointly owned by the three regulators, Ofqual, Welsh Government and CCEA (for Northern Ireland). When this was agreed it was assumed a common standard could be agreed. This position was tested to the limit in 2012. Looking ahead we will take the steps that are necessary to distinguish the GCSE qualifications that we regulate (and where we can assure the standards), from any that may be awarded at a different standard.

**Our Approach to Awarding GCSEs in Summer 2013**

We have been clear that we have used the comparable outcomes approach since 2009 when the first new AS awards were made (and it is a development of the approach that has always been used for GCSE awarding). It was not new for summer 2012. We have been clear that we use it for two key reasons—to make sure that students taking new qualifications are not disadvantaged in the first years of those qualifications (to allow for teachers being unfamiliar with the qualification), and to make sure grade standards between exam boards were in line so that students sitting with one board do not have an advantage over others. It provides a statistical check that would indicate if awards were out of line with expectations, in order that the exam board can consider whether there is evidence that this is justified.

We have discussed this approach with our Standards Advisory Group and we are clear that it is currently the best available approach. We will continue to use this approach for summer 2013 and we have published further detail on this [http://www.ofqual.gov.uk/standards/summer-exams-2013/].

**Lessons Learned**

The design of these GCSE English and English language qualifications is too complex. Moving to linear assessment helps: it enables safer awarding and eliminates the so-called “route effect”—the likelihood that students will do better if they enter in a modular way. The proportion of controlled assessment—60%—is hard to justify in assessment terms, and we are considering the place of controlled assessment in new GCSE English language qualifications due for first teaching in September 2015.
Qualifications and school accountability measures need to sit well together. We welcome the thrust of the Government’s current consultation on the secondary accountability framework. The current proposal is that a threshold measure is retained in GCSE English and maths. We understand the rationale for that. If that is confirmed, then we will need to ensure that new GCSEs in those subjects are sufficiently resilient.

The views of technical assessment experts in exam boards matter, and they need to be consulted as new qualifications are designed. There is some evidence that this did not happen sufficiently or at the right time when these GCSEs were being designed by our predecessor body. We will certainly make sure that this happens for the new GCSEs now being planned.

More generally, the English 2012 experience has told us a lot about how qualifications are used in schools and the strategies that are used by many teachers, as well as how they interact with the accountability system. That experience—sometimes sobering—has been and will be invaluable as we develop our regulatory arrangements for the future, plan for the implementation of reformed qualifications, and advise Government on its policies on qualifications, accountability and other related areas. For example, it heavily influenced our advice in the autumn to Ministers about the proposals for English Baccalaureate Certificates.

Regulation was effective in ensuring comparable outcomes nationally—preventing students in 2012 from being disadvantaged by the fact that they were taking a new qualification, and in reducing unjustified discrepancies in results between one exam board and another. The High Court subjected our approach in general, and our actions in 2012, to intensive scrutiny, and found they were justified—we welcome that. However the approach is not well understood by schools and colleges, and not generally trusted. The proper steps we took to challenge Edexcel and WJEC’s preliminary results appear to have taken some people by surprise, the suggestion being that Ofqual was fixing results. We have a big task ahead, to explain our approach, and to make it and our actions more transparent. We have already published information about our approach to the awarding this summer and we will publish more information in the run-up to the summer exam series.

Appendix 1

OFQUAL’S PRINCIPLES FOR REGULATING QUALIFICATIONS WHICH MAY BE SUBJECT TO MORE THAN ONE SET OF REGULATORY REQUIREMENTS.

1. Our main priorities are the qualifications taken by learners in England (and Northern Ireland for vocational qualification), the interests of these learners and the interests of others who use these qualifications. We will, however, work to avoid disadvantaging learners taking regulated qualifications in other countries where possible.

2. We should aim, where possible, for the same qualifications to be available in England, Wales and Northern Ireland. We will not pursue this aim where this would prevent us meeting our statutory objectives, and in particular securing standards, or where this is explicitly contrary to the education policy of the respective governments.

3. Where a qualification we regulate is also assessed or awarded in jurisdictions other than England (or for vocational qualifications England or Northern Ireland), the same standard must be maintained, whatever the location of the learners. If this cannot be assured, the same qualification should not be available outside England (or England or Northern Ireland for vocational qualifications). Any alternative qualification should be uniquely identified and distinguishable from the qualification we regulate.

4. We should avoid, where possible, putting recognised awarding organisations in a position whereby they cannot simultaneously satisfy our regulatory requirements and those of the regulators in Wales and/or Northern Ireland (with the result that an awarding organisation may not be able to offer the same qualifications in England and Wales and/or Northern Ireland).

5. We should prioritise managing the risks to the safe awarding of qualifications to candidates in England (and Northern Ireland for vocational qualifications), including where those risks arise from the activities of another regulator.

6. We should explain our principles to awarding organisations and to the public.

March 2013

Further written evidence submitted by Ofqual

Question 1—Are you satisfied that exam boards have taken steps to improve their moderator feedback to schools/colleges? Will it be clear to centres in future if they have marked slightly generously/severely but within tolerance? How do you plan to check this?

As part of our work to monitor this summer’s awards we have made clear to exam boards that they should improve their feedback to schools. We have included this matter in a letter we sent to exam boards recently and are seeking assurances in that they are putting improvements in place.
Our regular monitoring activities incorporate checks on a range of systems and processes, including exam board management of the summer activities of moderation, marking and grading. The monitoring work also enables us to be ready to take any action quickly if necessary. We are checking whether exam boards’ systems and processes can provide the enhanced information to schools that have marked slightly generously/severely but within tolerance. We will then discuss with ASCL and NAHT how desirable this additional information would be to schools, bearing in mind the additional costs that exam boards would incur to provide it.

Question 2—Who stipulates the tolerance to be applied to internally assessed units in GCSEs? Are you planning to extend the tighter tolerance to other GCSE subjects?

Our Code of Practice requires exam boards to moderate marks that have been given to internally assessed units. The exam boards set the tolerance that they use in moderating the marks.

We supported the action taken by exam boards before exams in November 2012 to implement a tighter tolerance (+/-3% of the maximum mark for the unit) for GCSE English and English language. We are gathering evidence of the impact of that change, from the November 2012, January 2013 and the June 2013 series of exams.

The current tolerance (+/-6% of the maximum mark for the unit) is used for all other GCSE subjects and all A level subjects with coursework/controlled assessment. This tolerance was first used by exam boards for curriculum 2000 A levels and was then adopted for new GCSEs when they were introduced in 2003.

We are planning to commission some research that will provide an evidence-based method for determining appropriate moderation tolerances for summer 2014 and beyond. In the meantime there is no evidence to suggest the tighter tolerances should be extended to any other subjects.

Question 3—Have you seen an improvement in exam board communication with schools and colleges this year? Are you satisfied that schools are now sufficiently aware that grade boundaries might change from one exam series to the next and that they understand the reasons for this?

As part of our work to monitor the January 2013 GCSE English exam series, we commissioned an independent review of exam boards’ activities, including their communication with schools and colleges. The review indicated to us that such communication is improving. Additionally, as a result of the review, we asked exam boards to take some additional specific actions. As noted earlier, we will continue to monitor exam board communication.

As well as this, we have published on our website an explanation of the differences between marking and grading (see http://www.ofqual.gov.uk/help-and-advice/about-marking-and-grading/). This is part of our plan for communicating with schools and colleges ahead of the summer 2013 awards.

Question 4—Your second report identified that some schools had been marking controlled assessment slightly generously but within tolerance and that this had contributed to the problems experienced last year. Will you be asking exam boards to monitor this “slightly generous but within tolerance” marking more closely in future?

We are monitoring a number of risks to summer 2013 awarding in GCSE English. This is one of those and we have asked exam boards how they will review, ahead of awarding, evidence of marking trends within tolerance.

Question 5—WJEC has increased its market share in England in recent years in some qualifications. GCSE English is one such example. By contrast, CCEA has said that it will no longer be offering its GCSEs and A levels in England. Should WJEC, or in future Qualification Wales, continue to offer GCSEs to candidates in England? Should English exam boards be able to offer qualifications to candidates in Wales?

An exam board can offer regulated qualifications in England if it meets our requirements as an organisation and its qualification meets the rules set out for that specific qualification eg GCSEs. Bodies that were offering qualifications in England before we were established have been allowed to continue under transitional arrangements. WJEC falls into this category. If we found that WJEC were not compliant with our requirements, we have powers to take any necessary action.

Qualification Wales will be a new body. If it wants to offer regulated qualifications in England it will need to demonstrate that it can meet our requirements and that each of its qualifications also meet the relevant regulatory requirements.

With regard to English exam boards offering qualifications in Wales, qualifications policy is a devolved matter. This means that it is for the Welsh Government to determine the qualifications to be taken in Wales.

I hope that this answers your questions, but if you would like any more detail on our work, we would be happy to provide it to you.
I have also included with this letter a note to clarify the change from 2011 to 2012 in the percentage of pupils awarded an A*-C in English or English language. This follows the letter from Leighton Andrews to the Committee on 15th March 2013.

April 2013

Further written evidence submitted by Ofqual

In March 2013, the Minister for Education and Skills in Wales, Leighton Andrews, wrote to the Education Select Committee reporting a fall in 2012 in the overall outcomes for GCSE English and English Language in England. He stated that there was a drop of five percentage points from 2011 to 2012. A report that we published in November 2012\(^2\) said that like-for-like results went down by only 0.3 percentage points.

This note explains the apparent difference between the analysis we published and that referred to by the Minister.

The results referred to in our November report are the results for candidates in England, Wales and Northern Ireland who were awarded GCSE English or English Language in summer 2012. These results were collated by JCQ and published in August 2012. This was the first time that qualification awards had been made in these new specifications and included all those who had taken units in any of the four series: January and June 2011, January and June 2012. The summer 2012 results we reported on included students of all ages who completed qualifications in summer 2012. This included those who had completed their qualifications in year 11 (at the end of the Key Stage 4), and also those from years 9 or 10 and older students (for example, in further education colleges).

The letter written by Leighton Andrews AM, refers to data that was published by the Department for Education (DfE) in January 2013\(^3\). The DfE’s data is published each year for students that have reached the end of Key Stage 4 in that year, whereas JCQ’s results data, published in August each year, reflects candidates that were awarded in that exam series. As well as this, the DfE’s data reflects results for students in England only and JCQ’s data is for students from England, Wales and Northern Ireland.

It is therefore not straightforward to compare results that are based on the different data sources.

In our report into GCSE English in November 2012, we explained that the 0.3 percentage point drop from outcomes in 2011 to 2012 referred to ‘like-for-like’ candidates. The initial results that JCQ published in August reported a drop of 1.5 percentage points. Included in these results were a new GCSE in Digital Communications, which suppressed overall results by 0.2 percentage points.

It should be noted that there was a significant difference in the student cohorts in 2011 and 2012. There was a loss to the cohort of about 23,000 students from the independent and selective sectors—students that would typically perform rather better than average. We estimated that the impact of this change to the cohort would affect results by a fall of about 1 percentage point.

Taking these factors into consideration, we therefore concluded that, on a ‘like-for-like’ basis, in GCSE English and English language in England there had been about a 0.3 percentage point drop from 2011 to 2012.

The data referred to by the Minister also does not account for results for iGCSE English language. Between 2011 and 2012 the number of students at the end of Key Stage 4 that had taken an iGCSE in English language had risen from 5,574 to 22,254—an increase of 16,680 students\(^4\). Over 90% of these students achieved at least a grade C. This represents about 1.5% of all students that achieved at least a grade C in either a GCSE or iGCSE in English/English language.

April 2013

Written evidence submitted by Leighton Andrews, Minister for Education and Skills for Wales

Thank you for inviting me to give evidence to the House of Commons Select Committee on Education on 12 March. I was pleased to have the opportunity to explain issues relating to the regulation of three-country qualifications and to last summer’s awarding of GCSE English Language from the perspective of the regulator of qualifications in Wales.

In my oral evidence to the Committee, I made reference to what I referred to as “grade deflation” in England. I feel it would be helpful for you and Committee members if I elaborate on that a little.

August 2012 saw the first results for a new suite of GCSEs in English. 2011 saw the final results for GCSE English, which was available in both England and Wales. That GCSE was replaced in England in September

---


\(^{3}\) http://www.education.gov.uk/researchandstatistics/datasets/a00219173/gcse-and-equivalent-results, table 9

\(^{4}\) http://www.education.gov.uk/researchandstatistics/datasets/a00219173/gcse-and-equivalent-results, table 12
2010 by a revised GCSE English and a new GCSE English Language. In Wales only GCSE English Language was available from 2010.

In simple terms I think it is worth noting that in 2012, for the first time in many years, there was a fall in the overall outcomes for GCSE English/English Language in England. Data from the Department for Education in England show that in 2012 63% of the cohort in England achieved a grade C or above in GCSE English or GCSE English Language, compared to 68% who achieved grade C or above in GCSE English in 2011.

I also believe that it is important to set this in the context of a desire to maintain comparable outcomes, particularly when specifications for qualifications change. This principle should have been particularly important in 2012, because of the changes to the GCSE suite for English/English Language that I refer to above. The principle of comparable outcomes was established on the basis that, when specifications change, awarding organisations and regulators should work to ensure similar outcomes between the final awards under the old specifications and the first awards under the new specifications. In so doing, the intention would be to ensure that no candidates are either unfairly disadvantaged, or advantaged by a switch in specifications. In my view this was demonstrably not the case in last summer’s awards in England.

March 2013

Written evidence by AQA

The GCSE English qualification was redesigned for first teaching in September 2010 and was awarded for the first time this summer. This new suite of English GCSEs includes English, English language and English literature. Students in England take either English, or English language and English literature. So while in the past students simply studied GCSE English, teachers are now offered a choice of routes to fulfil the requirement for English set out in the floor target. In a further change, this new GCSE English qualification also required students to study literature which was not the case in its predecessor.

It is the maintenance of standards from the previous GCSE English qualification to the two new GCSE English and English Language qualifications that has caused concern amongst teachers. To understand the issues, it is useful to understand the change in the structure and assessment of the specifications.

QUALIFICATION DESIGN—STRUCTURE AND ASSESSMENT

The structures of AQA’s GCSE English and GCSE English Language qualifications are summarised in Table 1 below. All students sit three units: the first is a written examination, which is taken at either the higher or foundation tier; the second is a compulsory speaking and listening assessment, carried out by teachers; and the third is a written controlled assessment, also carried out by teachers, with different versions for students taking GCSE English and those taking GCSE English Language.

There are aspects of this design that were intended by the regulator (through the qualification criteria) to create a positive teaching and learning experience and to meet the various needs of students. However, they make the setting of comparable standards complex and contribute to the volatility in results that teachers are experiencing.

Only 40% of the assessment is tiered and carried out by conventional examination, while 60% is controlled assessment that is marked by schools. Thus, the ability to set controlled assessment boundaries, exam by exam, is key. Another consequence of this aspect of the design is that it is possible for students entered into the foundation tier of the examination to achieve a grade A overall and although few students will (it would indicate that their ability had been dramatically under-recognised in school), this was not possible in the previous qualification. It is also important to note that the unit 3 written controlled assessments are the key differentiator between the qualifications and hence critical to achieving comparable standards between them. Adjusting the unit 3 controlled assessment standards is necessarily the method by which comparable qualification level standards are set. Clearly, the choices that are made in how to balance outcomes across the assessment units impact on the overall rank-ordering of the candidates.

Table 1

UNITISED STRUCTURE OF AQA’S GCSE ENGLISH AND GCSE ENGLISH LANGUAGE QUALIFICATIONS

<table>
<thead>
<tr>
<th>Unit</th>
<th>Mode of Assessment</th>
<th>Weighting</th>
<th>Jan 11</th>
<th>Jun 11</th>
<th>Jan 12</th>
<th>Jun 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>English unit 1</td>
<td>Written Examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>foundation</td>
<td>tier</td>
<td>40%</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>English unit 1</td>
<td>Written Examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>higher tier</td>
<td></td>
<td>40%</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>English unit 2</td>
<td>Ephemeral Controlled</td>
<td>20%</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>untiered</td>
<td>Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
By comparison, the structure of AQA’s previous (legacy) GCSE English qualification is summarised in Table 2 below.

Table 2
LINEAR STRUCTURE OF AQA’S LEGACY GCSE ENGLISH A QUALIFICATION

<table>
<thead>
<tr>
<th>Component</th>
<th>Mode of Assessment</th>
<th>Weighting</th>
<th>Jun 10</th>
<th>Nov 10</th>
<th>Jan 11</th>
<th>Nov 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1</td>
<td>Written Examination</td>
<td>30%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>foundation tier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 1</td>
<td>Written Examination</td>
<td>30%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>higher tier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 2</td>
<td>Written Examination</td>
<td>30%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>foundation tier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 2</td>
<td>Written Examination</td>
<td>30%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>higher tier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 3</td>
<td>Ephemeral Coursework</td>
<td>20%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Component 4</td>
<td>Written Coursework</td>
<td>20%</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

AQA’s previous qualification comprised 60% tiered, written examination. The November awarding series was used as an opportunity for both resitting and early entry. It is important to note that the grade boundaries for the coursework were agreed at an inter-awarding body level and did not change during the lifetime of the qualification. This was unique to GCSE English. Having 60% of the assessment via written examination was the means by which comparable standards were maintained. This was more acceptable to teachers than a change to the coursework boundaries in which the task had not changed. This, however, also had the unfortunate effect of creating, over time, an imbalance in the outcomes on the examined and coursework assessments: coursework grades increased over time and so the grades on the examinations decreased to compensate. This has consequences for the validity of the qualification.

AWARDING THE NEW QUALIFICATIONS

In awarding the new qualifications, it was necessary to maintain comparable standards on several (sometimes conflicting) axes, including:

- from the legacy to the new qualifications;
- between awarding organisations;
- between the two qualifications (English and English Language);
- between the modules that comprise the qualifications;
- between the tiers of the examination papers; and
- across the multiple assessment opportunities (January and June).

The approach taken to these issues is briefly outlined below.

In order to maintain comparable standards between the previous (legacy) and new qualifications, and between the awarding bodies, statistical predictions based on the prior attainment (KS2 test results) of the students entered for the new qualifications were used to support the judgements of the awarding committee in the subject awards made this summer. This method was successfully applied in the maintenance of standards of those new GCSEs awarded for the first time in summer 2011.

The predictions took into account how the general ability of the students entered for the new qualifications compared with that of the students entered for the previous qualification. Crucially the predictions were used by all awarding bodies and were based on the national outcomes in the previous GCSE English qualification.

5 In 2010 AQA also offered a GCSE English B specification and GCSE English Mature specification but 90% of AQA students were entered for GCSE English A.
6 This specification was first awarded in summer 2004.
7 No work is submitted to the awarding organisation.
rather than each awarding body using outcomes in its own exam only, and so ensured that the different awarding bodies’ standards were aligned. It is important to note that although the predictions were an important source of evidence, the awarding committee were able to recommend boundaries that fulfilled their professional expectation of an appropriate standard of student work.

The predictions were also very useful in allowing the awarding committee to understand the very different kinds of students entered for the new English and English Language qualifications. In terms of prior attainment, the students taking GCSE English Language were significantly more able than those taking GCSE English. The disparity in the results of the two groups is stark but reflects differences in ability (for example, 31.3% grade Cs in English and 74.7% grade Cs in English Language). The challenge of the new literature element of GCSE English would have been considerable for these students.

When there are significant changes to qualifications, both in terms of their content and their assessment, there is an expectation that outcomes may dip as teachers and students become familiar with the new demands. Because it is considered unfair to those students unfortunate enough to be in the first cohort taking the new qualification, the use of statistical predictions is considered a useful method of ensuring that this extraneous change in demand is compensated for. Conversely, in this instance it seems likely that the move from 40% to 60% teacher assessment might have led to an expectation that results would improve. The purpose of the awarding process, however, is to take into account such changes in demand and assessment style.

The early awards of units in January 2011, June 2011 and January 2012 were challenging. While statistical predictions can be useful in supporting awarders’ judgements at a qualification level, they provide only very weak support in making early unit awards. This is because the students entered at this earlier point are a relatively small and unrepresentative sample of the cohort and assumptions can not therefore be made about the likely relationship between their prior attainment and GCSE outcomes. It is also the case that the students entered early will have had varying levels of preparation and will have matured, in both the academic and developmental sense, to varying levels. Hence, these awards relied almost exclusively on the awarding committee’s judgement of candidate performance.

This is problematic because, it must be noted, standards reside at the level of the qualification. The way in which candidate performance on the individual assessment units aggregates to give an overall grade is crucial. For example, if a group of students were to take a qualification made up of six assessments, the grade boundaries would need to be lower at the top end, and higher at the bottom end, than if they were to take a qualification made up of fewer units; these are known as regression effects. Even when it is possible to estimate statistically these aggregation effects, it is impossible for awarders to factor them into their judgement of candidate performance and thus the setting of the grade boundaries. The mental agility required to compare student performance, which is often uneven, to the generic grade descriptions is in itself challenging enough.

In sum, while we are entirely confident that the awarding committee made the best possible judgements given the information that was available to them, the level of confidence we could have in the early unit awards was inevitably lower than that in this summer’s awards in which students certificated. The setting of the controlled assessment boundaries in June 2011 and January 2012 was, of course, particularly problematic because the publication of these grade boundaries naturally influenced the way in which teachers marked student work thereafter. Teachers having a clearer idea of the standard impacted on the distributions of marks seen in later examination series. Despite repeated communications to make the likely change of these boundaries clear to all teachers, the change has nonetheless caused immense concern and has undermined for some their trust in the awarding process, however, is to take into account such changes in demand and assessment style.

There is concern that those students taking a modular route through the qualification have benefited from both lower controlled assessment boundaries and lower examination boundaries, particularly on the foundation tier. We are very much in the hands of the judgement of the awarders in assessing the relative standards between those students making use of the early units and those who took their assessments linearly. This cannot be done with confidence statistically as the two groups are selected samples of the overall cohort so one would not necessarily expect their outcomes to be the same (even taking into account measures of their general ability such as KS2). Moreover, one group had the opportunity to resist units and the other did not, this is particularly problematic since only the better of the attempts counts towards grading.

**Some ideas for the Immediate Future**

During the development of modular GCSEs, in 2009 AQA proposed to the regulator some approaches to the grading (or not) of early units. At the time these were rejected but are worth revisiting for GCSE English before the January 2013 examination series, particularly in relation to the controlled assessment units:

1. Early unit grades could be issued as provisional and only finalised in the summer when candidates certificate.
2. Early unit marks (either raw or standardised, but not uniform) could be issued instead of grades. These would allow teachers and students to assess performance without implying any reference to grade boundaries.
We accept that these options are likely to be unpopular with teachers. An alternative is not to offer the opportunity to submit controlled assessment in the January series, though we recognise that that too would be unpopular with some schools that have planned to submit at that point in the year.

Teachers would clearly need to be consulted on the impact of these options on teaching and learning before any decision is made. We would be happy to discuss these possibilities further.

September 2012

Further written evidence submitted by AQA

Why you cannot apply the comparable outcomes approach in the early examination unit awards

The key reasons why it is not possible to apply the statistical prediction used to guide the summer qualification-level awards to the earlier unit awards are:

1. The candidates entered for early units are less mature, both developmentally and academically, than we would expect them to be by the summer.

2. The candidates entered for early units may or may not be as well prepared as those entered in the summer.

3. The candidates entered are an unrepresentative sample of the final summer entry, and we cannot assume that they ought to perform so as to achieve the same value-added relationship from KS2 as all candidates do. We do not have sufficient information to adjust the predictions to take this into account.

4. If we apply the qualification-level statistical prediction to each individual unit, when we aggregate all the units together we would not meet the qualification prediction. The outcome would be severe at the higher grades and too generous at the lower grades.

The way in which units aggregate together is influenced by 'regression effects'. The more units there are making up a qualification, the greater these effects. The weaker the correlation between candidates’ performance on the units is, the greater these effects. These effects cannot be known in the early unit awards.

September 2012

Written evidence submitted by Pearson

As always, we worked closely with the regulator to determine the grade boundaries for this year’s GCSE English and all other subjects. The decision making process and its outcome are in line with Ofqual’s Code of Practice, and the imperative to maintain standards year on year. However, it is important that schools and students are satisfied that the right processes have been undertaken, and that we capture the lessons which may be learned. We are hopeful that the Committee’s session will support this.

To that end, the paper below sets out the approach Pearson and our fellow awarding organisations have taken to date. We hope this will be helpful to the Committee in providing some more context to this year’s award, explaining the processes undertaken, and making recommendations for the future.

Standard Setting

Awarding organisations, as you know, have a responsibility to maintain standards year on year, and this is overseen by Ofqual. It is important that standards are comparable over time in order for a qualification, and awards at different grades, to retain any intrinsic value for the learners who achieve them and therefore enable them to progress. This is the guiding principle in standard setting.

During the awarding process, senior examiners consider a wide range of materials and reach decisions on the number of marks that candidates need to get certain grades. These materials include copies of candidate work from the current and previous years, as well as relevant statistical data. Decisions reached during awarding must ensure that overall outcomes reflect consistent standards for candidates past and present.

Awarding the new English GCSE in 2012

2012 marked the first award of a new English specification. At a time of new specifications, setting grade boundaries is especially challenging, and the familiarity of teachers with specifications can also have an impact on results.

2012 is also the second year in which an additional statistical measure has been used to predict the outcomes of GCSE cohorts. The “prediction matrix” gives examiners more data about the prior performance of cohorts. In GCSE, the reference point used is performance at Key Stage 2. This means that, at awarding, we are better able to see if a group of students’ performance at qualification level is coming out of line with what could reasonably be expected given their overall ability (and therefore if the demands being placed on them to gain
particular grades may not be comparable to those placed on peers in previous years). Reasonable expectations as to outcomes for a given year are set by Ofqual prior to examinations, in the form of “tolerances” for awarding. Awarding organisations are required to ensure awarding outcomes are aligned to these predictions.

For the new GCSE English, the summer 2012 examination series was the first opportunity where all awarding bodies could consider how the grade boundaries set for individual units contributed to the overall qualification grade. As we began to make summer awards, it became clear that all awarding organisations were reporting results which suggested that the demands placed on students to reach particular grades for some units was insufficient to maintain the standard at qualification level. In other words, awards were moving out of “tolerance”.

**CHANGES TO GRADE BOUNDARIES**

Grade boundaries set with the regulator for the January cohort—a much smaller number of students than completed in the summer (in the case of Edexcel, just 750 students versus 24,000)—did not therefore give confidence that the same performance was being required of students for the achievement of particular grades as in previous years.

We agreed with the regulator that the shifts in outcomes we were seeing on January grade boundaries could not be accepted if the principle of ensuring the intrinsic value of performance at particular grades was to be upheld. In order to maintain standards it has been necessary for all awarding organisations to move grade boundaries.

In Edexcel’s case, in three instances out of nine, the grade boundaries have increased compared to the January 2012 examination series, by 2, 4, and 10 marks. This does not mean that the overall qualification standard has changed, nor that it is any easier or any harder to get a grade in GCSE English.

These changes were discussed with the regulator at some length, because it is also important, of course, that students are rewarded for the genuine performance they show on the day and that examiner judgement on quality of work, rather than a statistical model is the key factor in grades awarded.

Ofqual, Pearson and other awarding organisations therefore worked together to determine what grade boundary shift was required to report results which represented student achievements fairly and delivered comparability year on year. In Edexcel’s case, for example, the boundary shift decided upon reported results slightly above the tolerances determined by the prediction matrix.

We are happy to share our exchange of letters with the regulator on the changes, in which different grade boundary movements were discussed before reaching a decision as to the right award. This is usual practice where an award outside tolerance is being made.

**TEACHER REACTION TO CHANGES**

In considering the reaction of past weeks, it is critical to reinforce the point that the discussion of grade boundaries with the regulator between series, and their revision, is not unusual. Schools and teachers are aware that grade boundaries are always provisional and may move from series to series. We do not advise teachers to use previous grade boundaries as a guide for students as to the grades they may achieve. Finally, it should also be noted that Edexcel’s grade boundary decisions on this qualification are not severe in terms of the outcomes expected by the prediction matrices for the cohort.

What has contributed to the strength of feeling shown in recent days is the extent of the changes made, and that some grade boundary changes have been to “controlled assessment” modules, marked by teachers who may have used previous grade boundaries as a guide and communicated expected grades to students on this basis.

In the case of the Edexcel English GCSE, both can be clearly explained:

- First, due to the structure of the qualification, changes to boundaries in controlled assessment units were unavoidable if standards were to be maintained at qualification level. The size of the shift, too, is partly attributable to the structure of the qualification, which has only a limited number of grade boundaries on which a change can be made (though in other awarding organisations, changes of around 10% have had to be made even where the qualification structure is different).
- A second factor in the size of the shift, as noted above, is the very small size of the cohort upon which the original grade boundary for this new qualification was determined. The shift to the terminal rule and the fact that there are now three English GCSE options on offer to students (English, English Language and English Literature) also meant that the composition of the cohort sitting in both January and June had differences compared to previous years, complicating awarding.

**THE RESIT OPPORTUNITY**

Whatever the explanation, we understand that students not receiving the results anticipated can be unsettling, for them and for their teachers. However, we take very seriously our responsibility to ensure that students are rewarded correctly for their performance, and treated comparably to their peers in previous years.
It is in seeking to reconcile the above that Pearson made the suggestion to Ofqual that a further resit opportunity be provided. The perceived unfairness which stems from the difference in the boundaries between January and June cannot be addressed directly at this point without compromising standards or undermining the broad approach we take to awarding. However, the impression of unfairness also comes from the belief that young people may have been working to a boundary which then moved significantly—exacerbated by the concentration of changes on controlled assessment units. For this reason we have suggested that students have another opportunity to sit the same task, with June grade boundaries maintained, free of charge.

By providing a resit opportunity in the Autumn term, we hope to alleviate the sense of unfairness and to limit the impact on affected students who may be “at risk” of disengagement from education, which has been a key concern cited in communications from schools. The decision to provide resits demonstrates awarding organisations’ collective willingness to enable more students to reach the required standard but, crucially, not to compromise it.

LESSONS LEARNED

There are a number of lessons that can be learned from the issues that have been experienced this summer.

1. Clarity on the status of January grade boundaries

Schools are aware and used to grade boundaries changing between series.

However, new specifications always have an impact on performance as they settle and, given the extent of the changes with the new English specification, and the small cohort in January, we would have expected it to be particularly so in this case.

Communication with schools could with hindsight have been more explicit about the dangers of presuming January grade boundaries stood for the summer. Pearson will ensure this is explicitly addressed in all future series.

2. Accountability

There is a wider point to be made around the issue of the accountability and the extent to which it drives a focus on grade boundaries which is unhelpful. Giving transparency on assessment to teachers and holding them to account for performance, whilst avoiding too much focus on the test is an issue worthy of consideration.

There is a need to balance a right focus on standards and floor targets with the reward of broader measures of improvement and all stakeholders need to work together to develop a system which better facilitates this.

3. Impact of change

The issues we have seen this summer are related to a change in the structure and content of qualifications in a single subject. This is indicative of the complexity of making changes to the examinations system, and the associated risks to standards and potential implications for public confidence.

Proper time and consideration needs to be given to the long term impact of changes at the point they are suggested, in order that they can be resolved. For example, some possible issues around maintaining standards in modular and composite qualifications, relevant to this summer’s events, were explored in 2009 and with hindsight would have merited further debate.

At a time when system wide reforms are being considered, both to the way the exams system is run and to the qualifications which sit within it, we should be cognisant of the likelihood that issues such as this may well arise in any transition to new qualifications and systems. This will be exacerbated if both things happen at the same time and if there is insufficient engagement between all stakeholders on associated risks.

I hope the above reassures the Committee that the right processes have been followed with regard to this year’s results. We are working hard to talk directly with schools and their representative groups to explain our decisions.

We feel acutely the disappointment of learners and teachers, but are confident that the steps we have taken are to the good of the credibility of the GCSE qualification awarded to students this year and for the long term.

September 2012
Written evidence submitted by WJEC

WJEC is the second largest provider of GCSE English and English Language qualifications in England. The main issues which we perceive as arising from the summer 2012 awards are set out below (with some additional explanation provided as an annexe).

1. General Issues
   (a) that overall % A*-C outcomes are down in both England and Wales when under a “comparable outcomes” brief the intention was to maintain consistency in standards.
   (b) that unusually large numbers of centres have seen significant changes in their % A*-C outcomes for 2012 when compared with 2011.

[note: In January 2012, WJEC offered written papers but not controlled assessment, so did not experience the problems reported in the Ofqual report in relation to grade boundaries being changed between January and June for controlled assessment.]

2. Issues in the Regulators’ Domain
   (a) regulatory decision (2008/09) to give controlled assessment a 60% weighting in these specifications (WJEC view was that it should be nearer 25%).
   (b) complexity of the qualifications suite available in England—two qualifications each with written papers available at foundation and higher tier and hence multiple routes through them to achieve grades A*-C (exact number of variants depending on awarding body).
   (c) doubtful robustness of the Key Stage 2 predictor model which awarding organisations are required to use.
   (d) influence brought to bear by regulators on some awarding organisation grade boundaries, based on conclusions reached by regulators using the outcomes of the Key Stage 2 predictor.

3. Issues in the Awarding Organisations’ Domain
   (a) appropriateness of moderating arrangements used, with +/- 6% tolerances—ie centres’ marks remain unadjusted when deemed to be within +/- 6% of the mark which the moderator would award.
   (b) specific or generic issues which may emerge from the Enquiries about Results process (which is currently in progress), including possibly in relation to quality of marking.
   (c) extent to which additional advice needs to be provided to centres regarding the different demands of the controlled assessment relative to previous coursework approach.

7 September 2012

Annexe

1. WJEC Background
   (i) Perspective
      WJEC understands the concerns expressed by candidates and teachers in England and Wales about the results of this summer’s GCSE English exams. Many candidates were disappointed by their results, and we have every sympathy with them.
      Like all the awarding organisations involved, WJEC followed guidelines set by regulators in making the summer awards. Ofqual has confirmed its view that the grades issued in England were fair, and we await the outcome of the Welsh Government’s review.
      WJEC’s prime concern is to ensure fair outcomes for all our candidates; we are continuing to work with regulators in England and Wales to ensure this.
   (ii) Scale of WJEC’s operation
      WJEC is the second largest provider of GCSE qualifications in English and English Language.
      In the summer 2012 series we provided assessments for some 156,500 “cash-in” candidates:
      — 38,600 for GCSE English (a qualification available in England only) representing approximately 23% of the provision in England.
      — 118,000 for GCSE English Language: around 84,000 of these were in England, representing approximately 19% of the provision, and 34,000 in Wales (approximately 95% of the provision).
   (iii) January and summer series 2012
      In the January 2012 series, WJEC did not offer controlled assessment. As we offered controlled assessment in the summer series only, we did not experience the difficulty which is reported extensively in the recent
Ofqual report, ie the need to adjust controlled assessment boundaries upwards between January and summer series.

In the January series, we offered each of the written paper units, which were also available in the summer series.

(iv) Perceptions of grade boundaries

June 2012 was the first occasion on which WJEC has set grade boundaries for the controlled assessment units. In order to provide guidance for centres when undertaking the assessment of these units, the marks available are grouped in “bands”, and relative to these the grade boundaries used by WJEC this summer are not generally perceived by centres as being inappropriate.

Our written papers, which are externally marked, have now been awarded three times: June 2011, January 2012 and June 2012. Whilst there has been some movement in grade boundaries across these series, in both directions, these in the main reflect examiners’ judgements about the relative difficulty of the tasks contained in those question papers and are not different in nature from the variation in grade boundaries that would be seen between series across GCSE units in other subjects.

2. MAIN CONCERNS

Our two greatest concerns in relation to WJEC’s assessments and awards for GCSE English and English Language in 2012 are that:

(i) our awards have contributed to a reduction (see section 3) in the national outcomes for England (% A*-C all awarding body outcomes reducing from 65.5% to 64.2%) and for Wales (% A*-C all awarding body outcomes reducing from 61.3% to 57.4%), neither of these being anticipated under the regulators’ “comparable outcomes” approach nor intended as a specific regulatory requirement

(ii) far more centres than usual experienced a large change in % A*-C outcomes in 2012 compared with their outcomes in the previous year.

3. ISSUES RELATING TO THE AWARDING PROCESS

We have the following concerns in relation to the awarding process for GCSE English and English Language in 2012.

(i) The regulatory requirement to use the predictor model based on Key Stage 2 data for “matched candidates” in England

This is the first time that WJEC has been required to use this approach, and we remain unconvinced that it provides a robust basis for delivering the “comparable outcomes” intended by regulators. We have asked regulators for a meeting to clarify their approach to “comparable outcomes” in the context of GCEs (where we also have concerns) and we would hope to extend that discussion to cover GCSEs.

(ii) The influence which regulators are able to exert post-award based solely on a statistical comparison of outcomes against the predictor model

When WJEC uses statistical information during the awarding process, this happens alongside consideration of evidence of the quality of candidates’ work; however, regulators use statistical information in isolation and the powers available to them are such that it is very difficult for awarding organisations to resist requests to adjust grade boundaries. In three of WJEC’s units, the grade C boundaries for GCSE English/English Language were adjusted upwards at regulators’ request post-award.

Our view is that the above factors have contributed to making WJEC’s GCSE English and English Language awards somewhat too severe in 2012.

4. ISSUES TO BE ADDRESSED

WJEC considers that the following issues are priorities which need to be addressed.

(i) Appropriateness of nature and weighting of the 60% controlled assessment element

During discussions with regulators in 2008/09, when the regulatory criteria for these specifications were being developed, WJEC’s view was that the weighting should be nearer 25%, but the prevailing view was that 60% was appropriate. We would suggest that this needs to be reviewed in the light of the 2012 experience.

(ii) Impact of the qualification and tier choices available

In England, there are essentially six routes available to a Grade C: English and English Language, each at Foundation tier, Higher tier, or through a combination of Foundation and Higher tier units. The complexity of these arrangements has the potential to cause confusion.
(iii) Appropriateeness of moderation arrangements

There is a convention across awarding bodies that “tolerances” of 6% should be applied when moderating controlled assessments of this kind (i.e., an awarding body will not adjust the marks awarded by a centre if, for the sample of work inspected by the external assessor, the centre’s marks are deemed to be within 6% of the mark which the external assessor would have awarded.

Our view is that the appropriateness of this convention needs to be reviewed in the context of a controlled assessment element that carries such high weighting within a high stakes subject.

(iv) Outcomes of the enquiries about results process

This is the main route through which we progress our follow-up work in response to centres’ concerns. As the volume of enquiries is considerably higher than in a typical year, we have more re-marking being undertaken by the most senior examiners and through this we are able to identify and address both specific and generic issues, including those relating to the quality of marking.

(v) Potential differences in the ways in which centres have adjusted to the requirements of the new specifications

The impact of the introduction of controlled assessment is possibly greater in English and English Language than in any other subject, hence it is possible that a considerable part of the change in results at centre level between 2011 and 2012 has been caused by the impact of the considerable additional demands on centres and candidates.

September 2012

Further written evidence from Leighton Andrews, Minister for Education and Skills for Wales

I welcome the opportunity to give evidence to your committee on the subject of GCSE English Language next week. I have, of course, read the transcripts of evidence given to the committee in September, and I understand that the chief executive of Ofqual has now corrected some elements of the evidence which she gave to your committee in respect of WJEC.

Your committee will be aware that GCSEs and A-Levels are three-country qualifications. Indeed, the brands are owned jointly by the CCEA (the examinations body in Northern Ireland), Ofqual, and the Welsh Government. You may be interested to know that John O’Dowd, the Minister for Education in Northern Ireland, and I wrote to the Secretary of State for Education on 1 August last year asking for a meeting with him, but that he rejected this request, which is unfortunate. I will of course be happy to share with the committee concerns which the Minister for Education in Northern Ireland and I have discussed in relation to the operation of GCSEs and A-Levels as three-country qualifications.

The GCSE English Language grading last year has raised issues which go to the heart of GCSEs as three-country qualifications, including the use of KS2 indicators—not historically used in either Northern Ireland or Wales—as factors bearing on the setting of grades in GCSEs, and the question of achieving “comparable outcomes”, an objective discussed by regulators since 2008 in the context of protecting learners from inconsistent outcomes when exam specifications change, but which Ofqual now appears to see as a year-on-year device “to prevent what is sometimes called ‘grade inflation’” (http://www2.ofqual.gov.uk/files/2012–05–09-maintaining-standards-in-summer-2012.pdf).

I look forward to seeing you on Tuesday. I am copying this letter to the Minister for Education in Northern Ireland.

March 2013