

GCSE English 2012



November 2012

Ofqual/12/5225

Contents

1. Executive Summary.....	4
2. Findings and actions.....	17
3. Our approach	20
4. The qualifications	22
5. Responding to concerns about awarding	30
6. Results for individual schools	39
7. Regulating GCSE English and English Language.....	74
8. Ensuring standards	85
Appendix 1 - GCSE English Awards 2012: Report of Findings from Centre Interviews, September – October 2012	88
Appendix 2 – Supporting information.....	105

A letter to the Chair of the Select Committee for Education

Dear Graham

This is our second report on GCSE English and English Language in 2012. It does not change the initial conclusions from our first report, but it goes much further.

Results for English were more variable than usual this year. To some extent this was to be expected, as results are always more variable when qualifications change, and the qualification changes here were significant and account for much of the variation. However we have found other matters also contributed to the variability this year.

We are concerned about what we have found. Major changes were made to GCSEs in 2007–9 with the best of intentions: to increase educational opportunity by making GCSEs modular, like many vocational qualifications. This and other curriculum priorities were given precedence over qualification standards and other apparently more technical concerns. At this time there was no independent regulatory voice.

Modularisation and other changes affected GCSEs in English more than any other subject, and GCSE English and English Language have proved to be exceptionally complex and difficult to award.

We have also found them to be especially susceptible to pressures, as teachers strive for the best possible outcomes for their students and school. With GCSE English currently so central to how schools are judged, this is a significant weakness. We have found that the qualifications are easy to bend out of shape: they can buckle under the pressures of accountability, and the evidence we have is that this did happen to some extent.

We have asked whether that could have been foreseen and prevented. Some controls, especially moderation tolerances that apply to controlled assessment, could have been framed and run more tightly, and communication could also have been better in various ways. Monitoring by exam boards and regulation by us could have been stronger and more intelligent. It has not helped that problems with the predecessor GCSEs were never publicly acknowledged.

However, once this GCSE design had been implemented, we do not believe that the problems seen this summer could have been eliminated, no matter how much more tightly these qualifications had been managed.

Furthermore, these new GCSEs have reinforced the trend of running Years 10 and 11 as a tactical operation to secure certain grades and combinations of grades. This has come to be seen as “what good schools do” despite the awareness of many

teachers and parents that the concept of broad and deep learning can get lost along the way.

Looking at the whole picture, we see that there are choices for the future. One option is that we lock down all qualifications (and by implication all teaching) to content that can be assessed with very high accuracy and assessment models that cannot be bent out of shape by external pressures. Many excellent qualifications could not pass both these tests. To take just one general example, it would be difficult under this approach to include oral assessments in foreign language qualifications.

Alternatively both qualifications and accountability could be rethought. It ought to be possible to have strong and well directed accountability without necessarily assuming that what can be accurately captured in accountability testing represents the totality of what should be taught in schools.

We believe that the latter option has more potential to contribute to making education standards in our schools as high as they could and should be. We would welcome wide discussion of these choices in the context of both the current accountability review and also the development of new qualifications, including English Baccalaureate Certificates.

We have been helped enormously in this review by the many schools who spared time to talk to our interviewers, and by their representative groups who provided us with information. We are very grateful. It has enabled us get to the root of the issues not just for future students, but for students studying now, and most especially for those students who took the qualifications this year.

Signed:



Amanda Spielman

Chair



Glenys Stacey

Chief Regulator

1. Executive Summary

- 1.1 When GCSE results came out in August 2012, it quickly became clear that although the overall English Language and English results were very similar to last year's results, there were some unexpected variations at school¹ level.
- 1.2 We immediately reviewed the evidence of how exam boards had awarded the qualifications. We reported on 31st August this year that they had acted properly when they set grade boundaries. That had been our first concern.
- 1.3 At that stage we were unable to explain the variations fully, so we undertook a further programme of work to investigate and explain them. As part of this work we have interviewed leaders and teachers from more than 100 schools. We have found that the variations happened in large part because the qualifications had changed significantly, as we would expect, but that is not the whole picture.
- 1.4 On average schools' results were almost unchanged. A significant minority, however, saw large changes, about equally up and down.

Changes to the English GCSE qualifications

- 1.5 New GCSEs in all subjects were planned from 2007 and introduced in 2009 and 2010. They were in many cases significantly different from those they replaced, and the changes to English were greater than for any other subject.
- 1.6 The old English GCSEs were well established and exceptionally stable, more so than any other high-volume GCSE. What is more, the way in which students were assessed and to some extent what they could expect in the written exam were highly predictable and so were grade boundaries (within a few marks), especially for coursework².
- 1.7 Previously, all students studied GCSE English and four out of five also studied English Literature. The content of each qualification changed in 2010 when these two GCSEs were replaced with three GCSEs, in English, English Language and English Literature.
- 1.8 Students who would previously have studied English and English Literature will generally now study English Language and English Literature³, with a greater

¹ By "school" we mean those schools, colleges and other institutions where students study for GCSEs.

² Coursework is work undertaken by students in their own time, assessed and marked by teachers in schools and then graded by awarders in exam boards.

³ The arrangements are different in Wales and in Northern Ireland. This report relates to England only.

emphasis on language skills and much less overlap between their GCSEs than had been the case before. These tend to be the more able students. Students who would previously have studied English alone will generally still do so, but the content has changed, with a greater emphasis on the core skills of English, and literacy.

- 1.9 All three new qualifications became modular, each with several units which students can take at intervals over the period of study. The rules changed so that for the first time, teachers were able to calculate the marks each student needed by the end of the two years of study, and could track progress. The coursework element of the English and English Language qualifications rose from 40 to 60 per cent, and coursework became “controlled assessment”⁴, with students now doing the work under supervision in school. Like coursework, controlled assessment is *marked* by teachers and then *graded*⁵ by exam boards.
- 1.10 The qualifications are available in England from any one of four exam boards. Two of those exam boards together have over 80 per cent of the market⁶. On our analysis, the new qualifications generally appear to be better in some respects than those they replaced. They cover more ground, and the exams also appear to be better: they are less predictable and more thorough. One can expect this to be more demanding for some students, as has proved to be the case. What is more, in the old qualifications, students did very much better in coursework than in exams. There were wide disparities between the two, especially for those in the foundation tier⁷. This is much less evident in this year’s results.

⁴ Controlled assessment is work undertaken by students in schools under controlled conditions overseen by teachers. It is marked by teachers in schools and graded by awarders in exam boards, as coursework was.

⁵ The essence of the difference is that “marking” is about determining how the student performed on the particular assessment, while “grading” is the job of deciding the level of performance on that assessment that reflects the achievement of students who can consistently work at the relevant grade level.

⁶ AQA have 62 per cent of the market in England, WJEC 20 per cent, Edexcel 11 per cent and OCR 7 per cent.

⁷ Foundation tier papers are geared towards less able students, those predicted to achieve C–G grades. Higher tier papers are geared to more able students, those predicted to achieve A*–D.

- 1.11 However, the changes brought added complexity for exam boards. They had to maintain comparability of qualification-level standards, balancing a proliferation of different routes through the three qualifications, with awarding at unit level rather than all in one block at the end, and some units which are shared, or common to more than one of the three qualifications. As a result, they have proven to be exceptionally complex and difficult for exam boards to award.
- 1.12 These changes brought added complexity for schools as well. They had many more decisions to take: where previously they merely had to decide whether to enter each student for the foundation or higher tier, now they also had to decide whether to enter each student for English or for English Language and English Literature; when each student should sit units and in what order, and when and whether each student should re-sit any unit or units.
- 1.13 The flexibility inherent in the qualifications' design is virtually unprecedented for such a high-stakes qualification. It enables schools to decide and then plan and track what they judge to be the best route to the best grade obtainable for each student. As unit results accumulate for any one student, schools are able increasingly to predict the student's likely final result, and aim for it. This makes them especially susceptible to pressures, as teachers strive for the best possible outcomes for their students and school. With GCSE English currently so central to how schools are assessed, this is a significant weakness. We have found that the qualifications are easy to bend out of shape, and the evidence we have is that this did happen to some extent.
- 1.14 This summer saw the first awards of these new GCSEs.

What happened

School predictions

- 1.15 Schools generally set target grades for students, and use prediction systems and trackers to help focus their efforts, student by student, and to provide information on predicted school-level outcomes. They update predictions regularly. Some schools were predicting exceptionally good results in these qualifications this year, and in predicting, many assumed that June 2012 grade boundaries would be the same or almost the same as those set in January 2012 and published⁸ in March 2012. This is significant, as these predictions underpin schools' intervention strategies – ways in which schools provide extra support to ensure so far as possible that each student that can do so achieves a grade C.

⁸ Exam boards have always published grade boundaries.

- 1.16 Exam boards collect school predictions of results in late spring, shortly before students sit exams in the summer. Schools historically over-predict to exam boards, especially at the grade C boundary. However the over-prediction was higher than usual this year by several percentage points, despite the newness of the qualifications, and therefore the lack of past history for schools to go on. Schools evidently had a high degree of confidence in their predictions.
- 1.17 When predictions were given to exam boards, most students had already sat a written exam for a module, and received a grade for that module. Many of the controlled assessment tasks would already have been completed and marked in schools. Schools basing their preparation of students on the assumption that controlled assessment grade boundaries in June would be the same as those in January therefore felt very confident of achieving their predictions.

Awarding

- 1.18 To ensure comparability of standards, exam boards use a range of evidence to make judgements about the level of demand of each assessment, and set grade boundaries accordingly. Grade boundaries will often be set differently for different assessments, reflecting for example different levels of demand between those assessments, and that is necessary to secure consistency.
- 1.19 Units can be sat at intervals⁹. Awarding and grade boundary setting in the three sittings before June 2012 was difficult for all exam boards. We set out more detail in the regulatory report we published on 31st August 2012.
- 1.20 We know, from the 2011 and early 2012 exam board awarding meetings which we attended in an observer capacity, that awarding in these early intervals was harder than usual. Awarding was bound to be more difficult than for more established qualifications, as examiners could not rely on any past experience of standards setting for the qualifications. For some units there were few students, and setting standards is harder when there are not many pieces of students' work to compare.
- 1.21 We observed in January 2012 that some examiners grading controlled assessment units thought they were erring on the harsh side. As we explain, that turned out not to be the case.
- 1.22 With many units shared between tiers or between the qualifications, and the need to strike the right balance between units, awarding in these qualifications was more difficult overall than in the past, and probably more difficult than many of the examiners had previously experienced. Exam boards use statistical data

⁹ In 2011–12 these were in January 2011, June 2011, January 2012 and June 2012.

to see how awards compare from one year to the next and that was more complicated here, in part because of the move from two qualifications to three.

Grade boundaries

- 1.23 The evidence we have seen shows that grade boundary setting happened as it should. We set out more detail in our initial report¹⁰.
- 1.24 Grade boundaries differed between January and June 2012. Awarding in June was generally more straightforward than in January, or in 2011 because examiners had the whole picture, and this enabled them to award with more confidence. However, grading was challenging for controlled assessment units. The majority of marks in the qualifications came from controlled assessment units, and four out of five of all students' controlled assessments were submitted in June 2012.
- 1.25 On seeing students' controlled assessment work, and comparing it with that seen in earlier intervals, examiners judged the grade boundaries at a higher level than in January. They judged that setting boundaries any more leniently would be setting them at the wrong standard. From the records of awarding meetings, it is clear that examiners saw considerable evidence of over-marking by teachers. The average mark for a piece of controlled assessment in June was higher than that for the same piece of work in January. As a result, the grade boundaries needed to be higher.
- 1.26 For written exams, grade boundaries reflect the relative ease of the exam: the easier the exam, the higher the grade boundary. That is how exam boards ensure a level playing field, and schools do not expect grade boundaries to stay the same from one exam to the next in any one qualification.
- 1.27 Schools have expressed concerns about a key examined unit¹¹ from the exam board with the majority market share. Examiners set the grade boundary for the exam in June higher than for the exam in January. We explain in section 5 the reasons, and confirm that awarding happened as it should.

¹⁰ See: www.ofqual.gov.uk/files/2012-08-31-gcse-english-awards-2012-a-regulatory-report.pdf (accessed on 31st October 2012).

¹¹ AQA Unit 1 (ENG1F) Understanding and Producing Non-Fiction Tests.

Results

- 1.28 Qualification awarding in June was completed mainly without intervention by the three regulators¹², although our exchanges with Edexcel and with WJEC (jointly with the Welsh regulator) led to amendments to their awards to bring them more in line with other boards. We detail this in pp35-36. The evidence is that these interventions were necessary. We can now see that final results for both these exam boards are closer in line with those of other exam boards.
- 1.29 Results stayed steady overall at a national level, in part because exam boards applied the “comparable outcomes” approach to awarding, in accordance with the regulators’ requirements. The comparable outcomes approach helps to maintain standards over time.
- 1.30 This approach is particularly important for new qualifications. Students may perform less well until schools have enough experience of the new qualifications in practice, and the comparable outcomes approach makes sure that students are not disadvantaged by that. This is especially important when qualifications change but the standard needs to stay the same, as was the case here. Without comparable outcomes, exam boards would struggle to award to the right standard. We explain this approach in Appendix 2.
- 1.31 Results for these qualifications varied school by school, although they stayed steady at a national level. This is normal – there are always variations between schools. Results are always more variable for new qualifications, and so we could expect more variability than usual, but the patterns of variation here are a little more pronounced than that. Moreover, we now know that variability in the old qualifications was unusually low, and so the difference would be particularly noticeable to schools.

What problems have been found

- 1.32 There are some real problems with the qualifications themselves that have contributed to the variations.
- First, the complexity of the qualifications. The combination of tiers and units, sittings and re-sittings and the overlap between units made it difficult for exam boards to set grade boundaries at unit level and to maintain sufficient comparability between units, tiers and each qualification as a whole.

¹² GCSE, AS and A level qualifications are regulated jointly by three regulators: Ofqual (in England), the Welsh Government (in Wales) and CCEA (in Northern Ireland). Until 2010 QCA was the regulator in England.

- Secondly, the large teacher-marked controlled assessment component has made these GCSEs particularly susceptible to pressures such as those from performance tables and other accountability systems. Schools clearly want the best results for students, and for their institutions. With the analysis we have done, and with the help of those we have spoken to in schools, we have seen how this worked in practice.
- Many students already had a written exam grade before June 2012. Schools could then calculate – basing their calculation on assumed (January) controlled assessment grade boundaries – the number of marks on the controlled assessment they assumed would give students a particular grade for the whole qualification. The patterns of controlled assessment marks suggest strongly that this assumption influenced the way that many teachers taught, or the way they administered or marked controlled assessment. The pattern of marks – the unprecedented clustering around perceived grade boundaries for each whole qualification – is striking, as we show on pp.68-73 of this report. There is also evidence of less sophisticated marking of controlled assessment to fit what were assumed to be the unit-level grade boundaries, but the qualification-level effect is much stronger.
- It is also clear that this tendency would have resulted in problems no matter where the January grade boundaries had been set: once teachers assumed that they knew how many marks a student needed to achieve a grade, for many that assumption was bound to affect their teaching, assessment preparation and marking.
- Thirdly, as the qualification has 60 per cent controlled assessment, the standard marking tolerance of 6 per cent is broad enough to take a large proportion of students over a grade boundary without exceeding the tolerance. Exam board moderators check a sample of controlled assessments, ahead of awarding. Moderation by exam boards did not prove strong enough to identify and counter problems effectively. It was not designed to do so. Most schools did not have their marks adjusted on moderation. Moreover, moderation operates at unit level, but the patterns of marking found here are only apparent when the marking is analysed at qualification level, and that can only be done after the event.

1.33 There are aspects of the interactions between exam boards and schools that are relevant:

- First, feedback to schools from moderators created problems in some cases. Some schools assumed that moderator confirmation of their marks

meant that their grade predictions were also confirmed. Others have said that moderators suggested where grade boundaries would be set.

- Secondly, communication between exam boards and schools did not work as well as it should have done, most obviously around the possibility that grade boundaries set for the controlled assessment could differ. While this was communicated by exam boards, and schools did receive and note it, it may have seemed like pre-flight safety demonstrations, referring to a possibility so remote that it attracted almost no attention.

1.34 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 a 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.

1.35 Drawing all these threads together, we see that there are many contributory causes of the variability seen in GCSE English and English Language this year. None of these causes is unique to the subject: all GCSEs are currently modular; science has overlapping single, double and triple specifications; modern foreign languages also have 60 per cent controlled assessment; mathematics is an equally high-stakes GCSE; and moderation tolerances are currently the same across all GCSEs and A levels. It is the cumulative impact of these and the other factors outlined above, in conjunction with the external pressures of accountability which come together uniquely in English.

1.36 These qualifications are very different from the ones they replaced and we would expect a level of variability towards the unusual end of the spectrum. Some schools will have adapted to the new qualifications better than others, of course. From the evidence we have seen, we conclude that the reasons for the additional variability are:

- Some students found the qualifications more demanding than others, and this was particularly the case for less able students.
- A small proportion of students overall (some of the students who sat units in January 2012) appear advantaged because of the route they took through the qualifications.
- Reliance of teachers on published grade boundaries:
 - In designing intervention strategies, many teachers will have assumed grade boundaries for June 2012 would be carried forward

from January. Intervention is often finely tuned, to bring students to what teachers estimate to be the total marks they will need for a particular grade for the whole qualification when units are put together.

- Known January 2012 grade boundaries influenced teachers' subsequent administration and marking of the remaining controlled assessment units in schools in some cases, at individual and collective unit level. In June 2012, examiners saw significant evidence of over-marking of controlled assessment units.

Effects on students

- 1.37 The introduction of new qualifications inevitably disrupts student and school performance to some extent, even in the absence of any particular problems. The comparable outcomes approach, explained above, compensates for that at a national level, but not school by school.
- 1.38 After reviewing awarding, we studied students' results at school level, to look for evidence of unfairness, and to see the impact of the differing grade boundaries that has been suggested by some schools and their representatives as the key issue.
- 1.39 Only a small proportion of students who submitted controlled assessments in January were slightly advantaged by the grade boundaries set at that time. Adjusting June students to January boundaries would advantage almost all of this year's students, relative to those in previous and future years. And it would mean ignoring the evidence of over-marking – the fact that average marks for similar pieces of work were higher in June than January.
- 1.40 It is also not feasible to adjust controlled assessment grades to reflect what students might have got if their schools had ignored or acted differently on information about January grade boundaries: there is no possible way of identifying those students, or determining how those students would have performed in different circumstances.
- 1.41 It has been suggested that students who sat all modules in June 2012 were penalised when compared to those that took some modules earlier. Student choices about when they take units and in what order, or when they re-take units, will inevitably affect their results, and this is so for GCSEs in all subjects. It is not possible to know whether any individual student would have done better, the same or worse had they entered in January. It is clear, however, that a large proportion would be significantly over-graded if June boundaries were changed to match January boundaries.

- 1.42 If the June controlled assessment units were re-graded on January grade boundaries, the overall results could not be justified by reference to educational achievement, and it would introduce substantial new unfairness. For example, almost one in three students sat the most popular written exam in January¹³ and submitted a controlled assessment in June. If the June controlled assessment boundary was moved to match January for these students, then the proportion receiving a GCSE grade C would be 85 per cent, compared to 64 per cent who actually achieved a grade C in the new qualifications and 65 per cent who achieved grade C in English last year.
- 1.43 For both controlled assessments and exams, we believe the early re-sit option already offered by exam boards offers the best balance between on the one hand giving students who believe they have been affected the chance to improve their grades quickly, and on the other hand introducing significant new unfairness into the results.

Wales

- 1.44 The approach to awarding these qualifications and the approval of the original awards by all exam boards was agreed by us and the Welsh and Northern Ireland regulators.
- 1.45 The re-grading of WJEC's GCSE English Language for students in Wales directed by the Welsh regulator was for stated reasons essentially unconnected with the concerns that have been raised in England around fairness between awarding between January and June 2012.
- 1.46 After awarding the regulator in Wales subsequently produced a report expressing the view that there was no known reason why students in Wales should have performed worse than in 2011. The report was accompanied by a direction to WJEC to adjust grade boundaries for students in Wales so as to make the grade distribution as close to the previous year's distribution as possible.
- 1.47 This unilateral decision by the regulator in Wales has resulted in more favourable treatment for the 2,300 Welsh students who were upgraded than for their English counterparts. However this was entirely outside our control or that of the regulator in Northern Ireland. It signals significant problems for the future, if we are to maintain common standards across borders.

¹³ AQA Unit 1 (ENG1F) Understanding and Producing Non-Fiction Tests.

What could have been done better?

1.48 We have identified fundamental issues in relation to the qualifications design:

- Excessive complexity. The decisions relating to qualification design were the responsibility of the Qualifications and Curriculum Authority (QCA). We have reviewed QCA records of discussions that we have been able to locate, and have 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. It appears that curriculum aspirations over-rode in-house technical assessment and regulatory concerns, which at that point had no independent voice.
- The qualifications were designed with a high proportion of controlled assessment, administered and marked by teachers, and there is little evidence of awareness of the tensions this would create, given their central role as a performance measure for schools. The GCSE five A*–C per student measure was amended to mandate inclusion of English and mathematics from 2007, so it should already have been clear when they were designed that these new qualifications would be extremely high stakes at school level.
- More broadly, there was a lack of openness at the time about the problems with coursework and the extent to which English GCSE was already under stress as a result. The move from coursework to controlled assessment dealt with concerns about external help and plagiarism but did not adequately address concerns about the support provided by teachers. Had these issues been more fully and openly acknowledged at the time, there might have been more scepticism about the likely effectiveness of controlled assessment as a solution to the problems with coursework.

1.49 Regulation was effective in ensuring comparable outcomes nationally – preventing students in 2012 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. We have identified that regulation could have been more effective:

- The lack of openness in the past about problems with coursework and the pressures it was creating in the old qualifications obscured significant risks here. We foresaw that these new qualifications would be difficult to award, and we were monitoring awarding, as we set out in this report. However, we were not looking especially at what happens in schools, because we do not regulate schools.

- Although exam board moderation of controlled assessment marking could not of itself have prevented the problem, the standard tolerance (6 per cent) applied to all GCSEs enabled it to some extent.

1.50 We have also identified that exam boards could have been more effective:

- Communications between exam boards and schools could have been better, and more focused, most especially on the dangers of assuming grade boundaries.
- Moderators' feedback to schools was not always sharp enough.
- Exam boards could have done more to identify and report emerging concerns to the regulators.

The future

1.51 Students in England starting their GCSE courses this year will not take these qualifications in a modular way unless they are planning to complete them next summer. In England we have already implemented changes that stop these and other GCSEs from being examined or otherwise assessed at intervals. That means that the current academic year is the last year that exam boards will be managing awarding in stages, unit by unit.

1.52 Our findings here are pertinent to the fundamental review of controlled assessment we started in April this year. We were concerned about the proportions of controlled assessment in some subjects and concerned about how valid the assessments were. As we said at the time, we want assessment in each subject to be valid, and to promote good learning and teaching. We put out a call for evidence (now closed) and expect to complete our review by the end of 2012.

1.53 We have been reviewing whether controlled assessment promotes good teaching in schools, whether it provides properly discriminating assessment, and whether it is consuming disproportionate teaching time and other resources. We expect to report on this review by the end of 2012 and to act on our findings as soon as is sensible and manageable. The issues set out in this report are relevant and significant to the review, but we do not assume that the answer is to move solely to written exams.

1.54 We will advise the Secretary of State for Education on the implications of our findings for the recently announced review of school accountability. We will also advise him on the implications of our findings in the context of his proposals for English Baccalaureate Certificates and other new qualifications.

- 1.55 In consultation with fellow regulators we are taking action to ensure standards for the current qualifications in 2012–13 and beyond, as set out in Section 2.
- 1.56 Finally we note that some of the problems identified in this report can be addressed by direct regulatory action, or through changes in the ways in which exam boards monitor and interact with schools. However, we regulate exam boards, not schools. Furthermore the forces of accountability act so strongly on schools that exam boards cannot reasonably be expected to counteract them entirely.
- 1.57 We would therefore draw the attention of all those with responsibility for schools to the importance of understanding the pressures on schools in relation to the teaching and administration of qualifications. Those who should understand this include governors, local authorities, academy operators and Ofsted as well as the Secretary of State for Education.

2. Findings and actions

2.1 Examining and awarding worked as it should, but with unpalatable consequences:

- Exam scripts were properly graded in January and June 2012 but the January boundaries were taken by schools to signal what the June boundaries would be.
- Early awarding was difficult, as anticipated. However, exam boards and examiners acted properly in awarding at unit and qualification level.
- In summer 2012, strong evidence of teacher over-marking in controlled assessment, not corrected by moderation, meant that grade boundaries needed to be set at a higher level than in January, to deal with over-marking.

2.2 The new qualifications are better in some respects than those they replaced:

- Early evaluation suggests that the new qualifications have more content than the old qualifications.
- The exams for the new qualifications are less predictable than those for the old qualifications.
- An implausible imbalance between students' results for coursework and exams in the old qualifications has been addressed.

2.3 However, the new qualifications' design was flawed:

- The combined effect of modularisation and control weaknesses in controlled assessment makes these exceptionally high-stake qualifications susceptible to pressures from the accountability system.
- The high proportion of controlled assessment, the freedom to submit controlled assessments at the end of the qualification and the ability (for the first time) for teachers to calculate the total marks individual students need weaken the qualifications still further, making them especially susceptible to those pressures.
- Curriculum considerations and a common approach for qualifications appear to have been given more weight than standards when the new qualifications were designed.

- Insufficient attention was given to the totality of the changes and the incentives on schools. The challenges of awarding modular qualifications in the context in which the qualifications would operate do not appear to have been considered.

2.4 The qualifications were put under pressure:

- Teachers relied unduly on grade boundaries published for January 2012 in preparing students for summer series assessments. Exam boards did publish their usual advice that boundaries might change in the summer but some schools did not take this sufficiently into account.
- Some teachers over-marked students' controlled assessment units in order to attain what they calculated to be the overall marks needed to get a particular grade for the unit or for the qualification as a whole
- The grade boundaries set in June meant that some students did not get the results they were predicted.

2.5 Regulation did not stop this happening:

- The risks to awarding and to standards should have been identified in the design stage.
- When the qualifications started to be taught in schools, radical suggestions to reduce the risks to standards were aired, but were discounted because they undermined the prevailing philosophies at the time. At that stage, regulation did not have an independent voice or statutory powers.
- Regulation ensured comparable outcomes nationally, and reduced unjustified discrepancies in results between one exam board and another.
- Exam board moderation of controlled assessment marking could not of itself have prevented the problem, but applying the normal tolerance to teachers' marking (6 per cent) enabled it to some extent.
- Communications between exam boards and schools could have been better, and more focused.

2.6 We are taking regulatory action immediately to reduce the risks to standards for the future:

- Starting with the controlled assessment units in the November 2012 re-sits, moderation will be tightened.

- No grades, or other information about aggregate performance, will be issued for January 2013 assessments (whether exam or controlled assessment) until June 2013 assessments have also been marked. Awards for both January and June assessments will be made at the same time.
- We have already announced that from September 2013, GCSEs will no longer be modular in England. To protect standards we will consider any other design changes needed.
- We will require exam boards to improve and sharpen communications with schools.
- We will review our own approach to regulation, to make sure that in regulating and planning qualifications reforms we take into account fully the context in which qualifications operate.

2.7 We are discussing these matters with fellow regulators in Wales and Northern Ireland.

3. Our approach

3.1 When we prepared our initial report in August our aims were to:

- confirm and explain what had happened to GCSE English results
- take swift action if problems were found
- maintain confidence in the exam system
- identify broader questions for further work.

3.2 Full details of the work that we did are described in Appendix 4 of our initial report, published on 31st August 2012. We found that each exam board had set standards appropriately. The June 2012 grade boundaries had been properly set and students' work had been properly graded.

3.3 We are acutely aware of concerns expressed by schools whose GCSE English pass rate has been significantly lower than predicted, and how significant GCSE English results are for students. In developing this report our aims have been to:

- listen and work with schools to understand fully the concerns
- find out and explain the issues this year
- identify any collective weaknesses in the model used for GCSE English in 2012, and understand and learn from these
- ensure standards for English qualifications and units in future.

3.4 To help us prepare this report, we:

- (a) met with school representative groups and teaching associations shortly after the publication of our initial report to share with them the evidence behind our initial report, and to give them the opportunity to clarify their concerns following the publication of the report.
- (b) asked each exam board to report to us on all aspects of the qualifications and how they were awarded. The reports provided by each exam board are available as annexes to this report.
- (c) interviewed each exam board's Chair of Examiners for GCSE English, to ensure that we heard evidence directly from them as key participants in the awarding process. We asked the Chairs of Examiners about their confidence in the examining, marking and awarding processes, their confidence in the awards

made, and about any concerns they had about the structure of the qualifications. We have taken into account what they told us in preparing this report.

- (d) used technical data from exam boards. We did further analysis, including school-level analysis not normally undertaken.
- (e) commissioned education specialists from the management consulting company Capgemini to interview headteachers, Heads of English and English teachers across 100 schools and colleges in England. These interviews have given us a far greater understanding of schools' concerns and enabled us to learn how the GCSE English qualifications were delivered "on the ground". Capgemini's summary of findings, which reflects the views of the schools interviewed, is provided as Appendix 1 to this report.
- (f) reviewed available documents from QCA archives to see how and why new qualifications were introduced.
- (g) with subject experts, compared the new qualifications with the old ones, to see how they differ.

4. The qualifications

Well established GCSEs in English and English Literature were replaced in 2010 by new and quite different GCSEs in English, English Language and English Literature.

On first analysis, the new qualifications and the assessments (exams and controlled assessment) are better in some respects than those they replaced. For most students they would have been more demanding.

However, the new qualifications are particularly susceptible to pressures, because of the flexibility inherent in them.

New qualifications

- 4.1 GCSEs in English were changed in 2010. The old qualifications were GCSE English and English Literature. They were unusually longstanding, by GCSE standards: GCSEs generally change every five years or so, but these were eight years old. There is nothing inherently wrong in qualifications being long living. If qualifications stay stable, teachers get increasingly familiar with the detail of what is to be taught, and experience shows that teachers are better at teaching qualifications they know well.
- 4.2 Most (80 per cent) students studied both English and English Literature. All students in maintained schools in England studied English.
- 4.3 Four exam boards each offered these GCSEs in England, with AQA having the largest market share. At the time it held 71 per cent of the market. WJEC had 18 per cent, OCR 6 per cent and Edexcel 5 per cent.
- 4.4 Students who would previously have studied English and English Literature will generally now study English Language and English Literature¹⁴. These tend to be the more able students. Students who would previously have studied English alone will generally still do so.

Why did the qualifications change?

- 4.5 Individual GCSEs change periodically, to keep up to date, or to implement changes in government policy or the curriculum, or changes in the way students are to be assessed. This time the changes were for three reasons:

¹⁴ The arrangements are different in Wales and in Northern Ireland. This report relates to England only.

- to reflect changes in government policy on what students should study, with a move to stress literacy and language skills within real life contexts
- to reflect changes that were happening to all GCSEs
- as part of a move to rationalise the way students are assessed in GCSEs; to make it more consistent across all GCSE subjects.

Changes in government policy on what students should study

- 4.6 In February 2005¹⁵ the government at the time signalled its intention to revise GCSEs in English, ICT and mathematics to ensure that students were competent in “functional skills”, as part of an initiative to improve the country's literacy, numeracy and ICT skills. Similar curriculum drivers were present in Wales and Northern Ireland.
- 4.7 For the subject of English, this meant an emphasis on speaking and listening, reading and writing, and English language skills within real-life contexts, including spelling and grammar and presentation skills. It was decided that GCSEs would change to reflect this new emphasis.
- 4.8 At the time, the content of two established GCSEs overlapped to some extent: GCSE English covered the whole of what is known as the “programme of study¹⁶”. English Literature covered and enhanced the literature requirements in the programme of study. The prospect of new qualifications provided the opportunity to deal with that¹⁷.
- 4.9 Following advice from QCA, Ministers endorsed the development of three GCSEs in English: English, English Language and English Literature. The idea was that students in England could take one of two routes:
- GCSE English: a single GCSE containing both language and literature and covering the entire programme of study for English in England
 - GCSE English Language and GCSE English Literature: the two GCSEs were to cover the programme of study in England when taken in

¹⁵ 14–19 Education and Skills, DfES 2005.

¹⁶ The National Curriculum Programme of Study for each core subject sets out the prescribed syllabus that pupils in maintained schools must be taught in that subject.

¹⁷ QCA Executive meeting decision paper: *GCSE 2011: proposed changes to English subjects, information and communication technology (ICT) and mathematics criteria*, 5th November 2008.

combination. Each GCSE includes content additional to the programme of study.

- 4.10 The rules for the combination of the three qualifications that students in England, Wales and Northern Ireland can study are different in each country, reflecting differences in government policy in each country.

Changes that were happening to all GCSEs

- 4.11 All GCSEs changed in two respects: they became modular, and controlled assessment was introduced to replace coursework.

Modularisation

- 4.12 In mid-2007, QCA was developing the Qualifications and Credit Framework, to embed a concept known as “credit accumulation and transfer” in qualifications at the same level¹⁸ as GCSEs. The idea was that students could build up a whole qualification by combining units (modules) from across a spectrum of suitable units. QCA was keen to promote modular GCSEs, with this underlying philosophy in mind¹⁹. At the time, the idea was that GCSEs would be studied in units, and GCSE units would sit alongside units in another qualification to be taught in schools called the Diploma, and units in other qualifications said to be equivalent.
- 4.13 Modularisation was not obligatory: exam boards could still design GCSEs as they used to, but they did not. All the indications were that schools would welcome and would prefer modular GCSEs, and that is what happened.
- 4.14 Most GCSE subjects became modular in 2009, with first teaching in September 2009. For a few subjects (English, ICT and mathematics) the changes did not happen until a year later, with first teaching in September 2010.

Controlled assessment

- 4.15 At the same time, QCA, with the agreement of Ministers, decided to change how internal assessment²⁰ worked for GCSEs, by replacing coursework with controlled assessment.

¹⁸ Qualifications are grouped in levels. GCSEs are at level 1/2. Some other academic qualifications and some vocational qualifications are at the same levels – level 1 and level 2.

¹⁹ QCA Executive meeting decision paper, 4th June 2007.

²⁰ Internal assessment is work done and assessed internally by the school. External assessment is work assessed externally by exam boards, usually by exam.

- 4.16 It has long been the case that GCSEs contained coursework. However, QCA had identified concerns about coursework²¹, mainly about ensuring that work submitted for qualifications was the student's own work. QCA reported uncertainty among parents and teachers about the kinds of help allowed. It also reported a lack of consistency in exam boards' rules about the ways in which teachers may help students in different subjects, and suggestions that in "a very small proportion of cases" there was deliberate malpractice.
- 4.17 Controlled assessment was designed to deal with those concerns. The arrangements changed so that student work for assessment by teachers was to be done under controlled conditions, under the supervision of teachers.

Introducing a more consistent approach to assessment

- 4.18 In 2007 QCA discussed with teachers, subject associations, higher education and employer representatives the balance to be struck between internal assessment and external assessment (exam) in all GCSEs²². At the time, the balance differed across subjects. In Economics for example, the rules²³ at the time allowed for a minimum of 80 per cent exam and no more than 20 per cent internal assessment. Design and Technology, on the other hand, had 40–60 per cent exam, and at least 40 per cent but no more than 60 per cent internal assessment. English had 40 per cent internal assessment; English Literature had 25 per cent.
- 4.19 QCA was concerned about the differences between subjects. In late 2007, as coursework was replaced with controlled assessment QCA decided to apply a more consistent approach across subjects and decided that the balance of assessment in any subject should be struck in any one of three, but only three, ways: no controlled assessment, 25 per cent controlled assessment, or 60 per cent controlled assessment.
- 4.20 QCA determined which of the three levels – 0, 25 or 60 per cent – would be most appropriate for each subject and set this out in new rules for GCSEs. In English Literature the proportion of assessment taking place in schools remained the same as in the old qualification, 25 per cent. For English

²¹ QCA (2005) *A review of GCE and GCSE coursework arrangements*.

²² QCA (2007) *Revisions to the GCSE qualification and subject criteria* draft paper.

²³ All GCSEs must comply with certain requirements set out in published GCSE qualifications criteria. GCSEs in some subjects also need to comply with requirements about subject content set out in published GCSE subject criteria.

Language the new rules directed 60 per cent controlled assessment, instead of 40 per cent as it had been. For English, the rules directed 60 per cent.

- 4.21 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.

How did the changes come into effect?

- 4.22 For most subjects, first teaching started in September 2009. GCSE qualifications became modular, and coursework was replaced with controlled assessment. The general approach was then followed through in the remaining subjects – ICT, mathematics and English.

- 4.23 In July 2008, QCA developed and consulted on new draft English and English literature subject criteria²⁴ to put into effect the changes agreed with government together with modularisation and controlled assessment. Parallel consultations took place in Wales and Northern Ireland. The intention was that the new qualifications would be introduced in schools by September 2010, for first teaching.

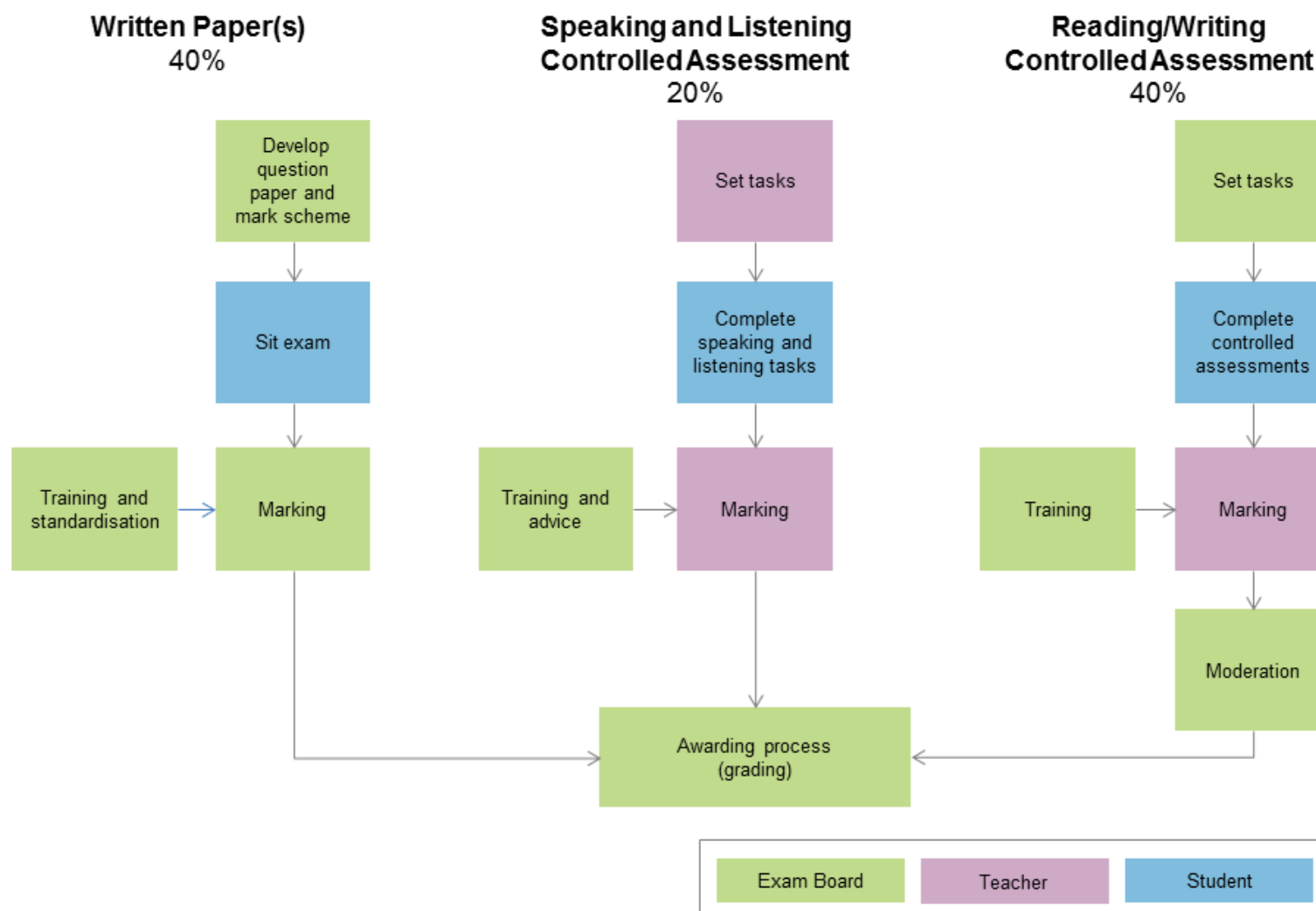
- 4.24 All three consultations generated similar outcomes. There was considerable support for the majority of the proposals. QCA issued final rules for English and English literature, and exam boards then developed qualifications in accordance with the subject criteria and submitted them for approval in late 2009.

- 4.25 The three new qualifications replaced the two old qualifications in schools in September 2010. During the academic year 2010/11, teachers were teaching the new qualifications and also the last year of the old qualifications. There is always an overlap like this when GCSEs change, because they are normally taught over two years. Summer 2012 was the first time that the new qualifications were awarded.

- 4.26 The following diagram illustrates the process followed by schools and exam boards for delivering the new GCSE English and English language qualifications this year.

²⁴ Subject criteria set out the knowledge, understanding, skills and assessment objectives for the subject. They provide the framework for an exam board to create the qualification.

English GCSE: Examining Process



Comparing the old and the new qualifications

4.27 We began what is known as a full scrutiny exercise for the new qualifications as teaching for them began. We will complete that in January 2013, and until then we cannot be definitive about the conclusions to be drawn about the differences between the old and new qualifications. However, to inform this report, we commissioned independent subject experts to review the new rules, the new qualifications and their internal and external assessments on a preliminary basis, to see how they compare with the old.

4.28 It appears that the new qualifications are in some respects better than those they replaced. They generally cover more ground. The exams also appear to be better for most students: less predictable and more thorough, and subject experts consider that most students were likely to have found the qualifications more demanding. On their analysis of the rules for the qualifications, they considered the new rules:

- do reduce the overlap between qualifications
- do show the drive to improve literacy, which was the policy intention. They allocate 45–55 per cent of the assessment to the functional elements of English, leading to a change in what students need to learn
- contain new elements, including understanding how variations in language relate to identity and cultural diversity, and understanding how literature from other cultures is influential
- are less prescriptive about the range of writing required, with the previous list of modes of writing replaced by a requirement to “write accurately and fluently choosing content and adapting style and language to suit a wide range of forms, media, contexts, audiences and purposes”
- are far less prescriptive about the content for the speaking and listening elements of the qualifications
- contain a new requirement to “participate in a range of real-life contexts in and beyond the classroom”.

4.29 Moreover, the rules now provide that the controlled assessment tasks are set largely by exam boards rather than by schools. The subject experts concluded that this is likely to have made the assessment more demanding for students, as teachers now have less scope to tailor tasks to the individual student.

4.30 Subject experts found the new qualifications from AQA likely to be more demanding for students than those they replaced, because:

- AQA controlled assessment units appear to be better than the coursework units they replaced. In the AQA old syllabus, students could produce reading coursework that could also be used for English literature coursework. In the new syllabus, if students are using the same books for English and English literature, they must produce separate tasks.
- two AQA exam papers in the old syllabus have been replaced with a single, longer paper worth 40 per cent of the overall GCSE: students now have to produce two pieces of writing in a single exam, rather than one piece of writing in each of two separate exams.
- the exams are less predictable, and less likely to narrow study of modes of writing.

4.31 Subject experts found Edexcel's exams likely to be more demanding for students than those they replaced. On a preliminary view of WJEC and OCR qualifications, subject experts have not identified any noticeable change in demand.

4.32 At first sight the new qualifications are in some respects better. However, they fall short in other respects, as we show in section 6 of this report.

5. Responding to concerns about awarding

- Grade boundaries were set properly in June 2012. They were not unduly harsh.
- Exam boards and regulators used statistical modelling properly and as intended.
- Results this year match those of last year, as one would expect, as standards were to stay the same when the qualifications changed.

5.1 In this section we deal with three related concerns.

5.2 First, it has been suggested that exam boards and regulators put too much emphasis on statistical modelling to ensure that standards matched last year, and that this led to harsh grade boundaries in June 2012 in order to ensure that overall, results were comparable to last year. Of the 34 units that were awarded in June 2012, concerns have centred on two in particular: the AQA foundation tier written paper (ENG1F) and the Edexcel controlled assessment unit for English (5EH03).

5.3 Second, concerns were expressed about the regulators' interventions to ensure exam boards' results were comparable with each other.

5.4 Third, there have been suggestions that the reduction in the results for all students is evidence that the June grade boundaries were harsh.

5.5 In this section we take each of these in turn.

5.6 The Education Select Committee²⁵ recently noted:

The issue of grading standards is complex and the awarding of grades a technical process, involving a blend of the professional judgement of examiners with sophisticated statistical data. It is a difficult area, which quickly becomes impenetrable to non-specialists. Therein lies much of the challenge for the exam boards and Ofqual when seeking to provide reassurance about grading standards.

5.7 The senior examiners awarding GCSE English in summer 2012 were experienced teachers of English who used their professional expertise and experience, guided by exam board staff, to make these finely balanced judgements about awarding grades. Very often there is no "right" answer – if

²⁵ Education Select Committee (2012) *The administration of examinations for 15–19 year olds in England*, First Report of Session 2012–13, HC141.

there were, there would be no need for awarding meetings. So we cannot say definitively that one grade boundary set is generous and another is not. It is a matter of professional judgement.

- 5.8 In this section we deal with concerns expressed about January and June 2012 grade boundaries. In our interim report on 31st August we reported that some January 2012 boundaries were generous. We have now done further analysis and reviewed more evidence. We conclude that some boundaries could have been set at different marks in January 2012. Nevertheless, those decisions were made properly and the senior examiners stand by them. We interviewed the Chairs of Examiners for all the exam boards, and they told us that they are confident in the awards they made in summer 2012 and in the earlier series.

Using statistical models

- 5.9 Senior examiners making judgements about where to set grade boundaries have always used statistics to help them. In the last century these statistics will have been based on results for the previous year. The statistical models used by exam boards are now more sophisticated and can take account of subtle changes in the overall ability of the students from one year to another.
- 5.10 Maintaining standards from one year to another is relatively straightforward in a qualification that has not changed. Senior examiners can look at student work in response to question papers, and they can refer back to work that students did in the previous year in response to similar question papers. But they will also use statistical predictions to help them.
- 5.11 Maintaining standards is much more difficult when qualifications change. The greater the change, the more difficult it will be for senior examiners to compare student work in the new qualification to student work in the previous year. So statistics will play a greater part in guiding senior examiners in the first years of a new qualification, as exam boards, teachers and students get used to the different requirements.
- 5.12 We know that students will do less well in the first year of a new qualification. There are often fewer study aids and materials available and their teachers are getting to grips with new and different requirements – new subject material, different question styles, and so on. If senior examiners are looking for an equivalent level of **performance** in the first year of a new qualification, this is likely to make it harder for students in the first year to do as well as they might have done in previous years.
- 5.13 So we require exam boards to look instead at **outcomes** – the results students are expected to get. They work on the assumption that, all other things being equal, the students taking an exam in one year ought to get outcomes (results)

that are comparable with the previous year. That is the basis of our comparable outcomes approach.

- 5.14 We know that all other things are rarely equal. The statistical modelling allows exam boards to take account of any differences in the ability of one year's students compared with another. So if the overall ability of the students is higher in one year, then outcomes would be predicted to be higher.
- 5.15 In the case of GCSE the statistical model is based on the relationship in previous years between Key Stage 2 prior attainment and GCSE performance. We used this approach for the new GCSEs that were awarded for the first time in summer 2011. We are not aware of any widespread concerns about those awards. We followed the same approach for the new GCSEs in English, mathematics and ICT to be awarded for the first time in summer 2012. A similar model is used for AS and A level based on GCSE prior attainment. There is further detail in Appendix 2.
- 5.16 Much attention has been focused on our stated aim of reducing grade inflation and using the comparable outcomes approach to do that. It has been suggested that this is "fixing" results and allowing only a quota of each grade to be awarded.
- 5.17 In fact, we used the comparable outcomes approach in the first year of these new qualifications with two aims in mind:
- first, to make sure that students taking the new qualifications were not disadvantaged
 - second, to make sure that standards between exam boards were in line, so that students entering with one board do not have an advantage over others.
- 5.18 The evidence we have suggests this was successful.
- 5.19 There have been suggestions that grade boundaries in June were higher than they needed to be, to compensate for generous boundaries in January, so that exam boards could meet the statistical predictions. That is not what happened.
- 5.20 In June 2012 there were a total of 34 units offered by the exam boards. Concerns have centred on two in particular: AQA foundation tier written paper (ENG1F) and Edexcel controlled assessment for English (5EH03).

AQA foundation tier written paper

- 5.21 The foundation tier unit is aimed at students working at grades C–G. The unit has been offered four times in total and is worth 40 per cent of the overall

qualification. The table below shows some of the data from the awarding meetings for each series.

Exam series	Entry²⁶	Mean mark	Statistically Recommended Boundary²⁷	Grade C boundary mark
January 2011	4,674	35.4	44	46
June 2011	31,186	36.3	40	44
January 2012	54,426	38.1	41	43
June 2012	140,743	39.6	52	53

5.22 In the first three exam series before June 2012, the statistical predictions would have been over-generous. This is because they were based on how the students were predicted to achieve on the qualification overall at the end of the two-year course. Students in 2011 and even in January 2012 were not yet at the end of their course, and this was only one unit out of their overall qualification.

5.23 Exam boards knew the statistical predictions were generous, but they could not know by how much, because they had no way of knowing how much better those students would be performing by the time they got to June 2012, or how well they would do on other units. So senior examiners knew that the statistical predictions in June 2012 were more reliable, and they therefore gave them more serious consideration.

5.24 It should also be noted that grade C is the highest boundary on the foundation paper. For a foundation tier paper to differentiate well between all students, the grade C boundary should ideally be set at between 60 per cent and 70 per cent of the maximum mark.

5.25 The grade C boundary mark on the AQA paper ENG1F in summer 2012 was set at 53 out of 80 (66 per cent of the maximum mark) and is a more appropriate boundary mark than those set in earlier exam series. In June 2011 and January 2012 the boundary marks for grade C on both the foundation and higher units were very similar, which meant students sitting the higher tier paper

²⁶ This is the entry on the distributions at the award (those students whose work had been marked and processed at the time of the award). Total entries were: January 2011 – 5,150; June 2011 – 41,391; January 2012 – 73,395; June 2012 – 159,916.

²⁷ The boundary mark suggested by the statistical modelling.

had to achieve a similar number of marks on a more demanding paper to achieve the same unit grade.

5.26 In our interim report, we said that some of the January grade boundaries were generous. It is certainly the case that senior examiners could have set the boundary at a higher mark on the basis of the increase in the mean mark for the paper. However, it is also clear from the reports made at the time of the award that senior examiners thought they were being severe.

Edexcel English controlled assessment Unit 3 (5EH03)

5.27 Concerns were also expressed about the grade C boundary on Edexcel's controlled assessment Unit 3 for English (5EH03). The unit is untiered, and therefore targeted at the full grade range (A*–G). It was offered for the first time in June 2011.

5.28 The unit is worth 40 per cent of the qualification and half of the unit (20 per cent of the qualification) assesses speaking and listening. The table below shows some of the data from the award meetings for each series.

	Entry²⁸	Grade C boundary mark
June 2011	1,661	55
January 2012	716	55
June 2012	23,522	65

5.29 It is clear from the table above that awards in June 2011 and January 2012 were made on unusually small entry numbers. It would have been difficult for Edexcel to set a different boundary mark in January 2012 on the basis of only 700 students. Because schools only send a sample of their controlled assessment work to be moderated, there was little evidence of student work available, and there will have been no evidence of speaking and listening performance for senior examiners to review²⁹.

5.30 The Edexcel awarding report from the June 2012 award notes that this unit “contains the Speaking & Listening component, which contains 50 per cent (48

²⁸ This is the entry on the distributions at the award (those students whose work had been marked and processed at the time of the award). Total entries were January 2011 – 5,150; June 2011 – 41,391; January 2012 – 73,395; June 2012 – 159,916.

²⁹ Speaking and Listening controlled assessment is not moderated.

of the 96) of the marks and which is likely to inflate outcomes”. The report also notes that setting the boundary at 65 would “bring C more in line with the C on 5EN03 (with the shared Speaking and Listening component), as 5EN03 is set at 64³⁰”.

- 5.31 So the boundary mark for C in June 2012 helped ensure better comparability across the two optional Edexcel units – 5EH03 was taken by students doing English and 5EN03 was taken by those doing English Language. With half of the unit comprising a common Speaking and Listening element, it is appropriate that the grade C boundary marks on both units are comparable.
- 5.32 The Chair of Examiners for Edexcel endorsed this award in the summer and reiterated his endorsement of the award when we interviewed him as part of our investigation.
- 5.33 We have concluded therefore that these boundaries were set appropriately and were not unduly harsh. Senior examiners were making the judgements expected of them, balancing the evidence they had at the time.

Regulatory interventions

- 5.34 One of the aims of the comparable outcomes approach is to make sure that the grade standards of the different exam boards are in line, so that it is not easier to get a grade C with one exam board than with another.
- 5.35 At the beginning of August we reviewed with the exam boards the emerging results for GCSE English and English Language. The following table shows the picture then.

³⁰ Edexcel Awarding Officer’s Report of Awarding, 23rd July 2012.

	AQA % A*–C	WJEC % A*–C	Edexcel % A*–C	OCR % A*–C
English Language				
Predicted	77.5	77.3	77.5	79.0
Actual	77.5	81.4	78.7	80.1
Difference	0	+4.1	+1.2	+1.1
English				
Predicted	31.3	36.1	34.7	33.6
Actual	32.5	38.8	42.7	34.9
Difference	+1.2	+2.7	+8.0	+1.3

(source: data supplied to Ofqual by AQA, Edexcel, OCR and WJEC, August 2012)

5.36 We had agreed with exam boards a tolerance of +/- 3 per cent on each qualification, with a combined tolerance of 1 per cent overall. There was no clear rationale to explain why the WJEC and Edexcel outcomes were so far outside tolerance. We therefore challenged WJEC over its English Language outcomes and Edexcel over its English outcomes (in bold in the table above) in early August.

5.37 The figures in shaded cells on the following page show the position after WJEC and Edexcel had made changes to their awards, to bring them closer into line with the other exam boards. It was entirely appropriate for us to intervene in this way, in line with our statutory objective to maintain standards.

	AQA % A*–C	WJEC % A*–C (revised)	Edexcel % A*–C (revised)	OCR % A*–C
English Language				
Predicted	77.5	77.3	77.5	79.0
Actual	77.5	80.9	78.7	80.1
Difference	0	+3.6	+1.2	+1.1
English				
Predicted	31.3	36.1	34.9 ³¹	33.6
Actual	32.5	35.3	41.5	34.9
Difference	+1.2	-0.8	+6.6	+1.3

(source: data supplied to Ofqual by AQA, Edexcel, OCR and WJEC, August 2012)

Results year-on-year

5.38 When the results were published on 23rd August, the proportion of UK students achieving A*–C in English was 63.9 per cent, compared with 65.4 per cent in 2011. There have been suggestions that the fall in overall English results of 1.5 percentage points is evidence that June 2012 boundaries were harsh.

5.39 In our interim report on 31st August we concluded that the standard in GCSE English had been maintained although changes in the mix of students meant that results were down overall by 1.5 percentage points. We have looked further and we found that like-for-like results (taking into account changes) went down by only 0.3 percentage points. This is well within the usual range of year-on-year variation for GCSEs.

5.40 The diagram below shows the difference in outcomes for all students and for two different sub-sets of all students – the matched candidates³², and what we have called “like-for-like” students.

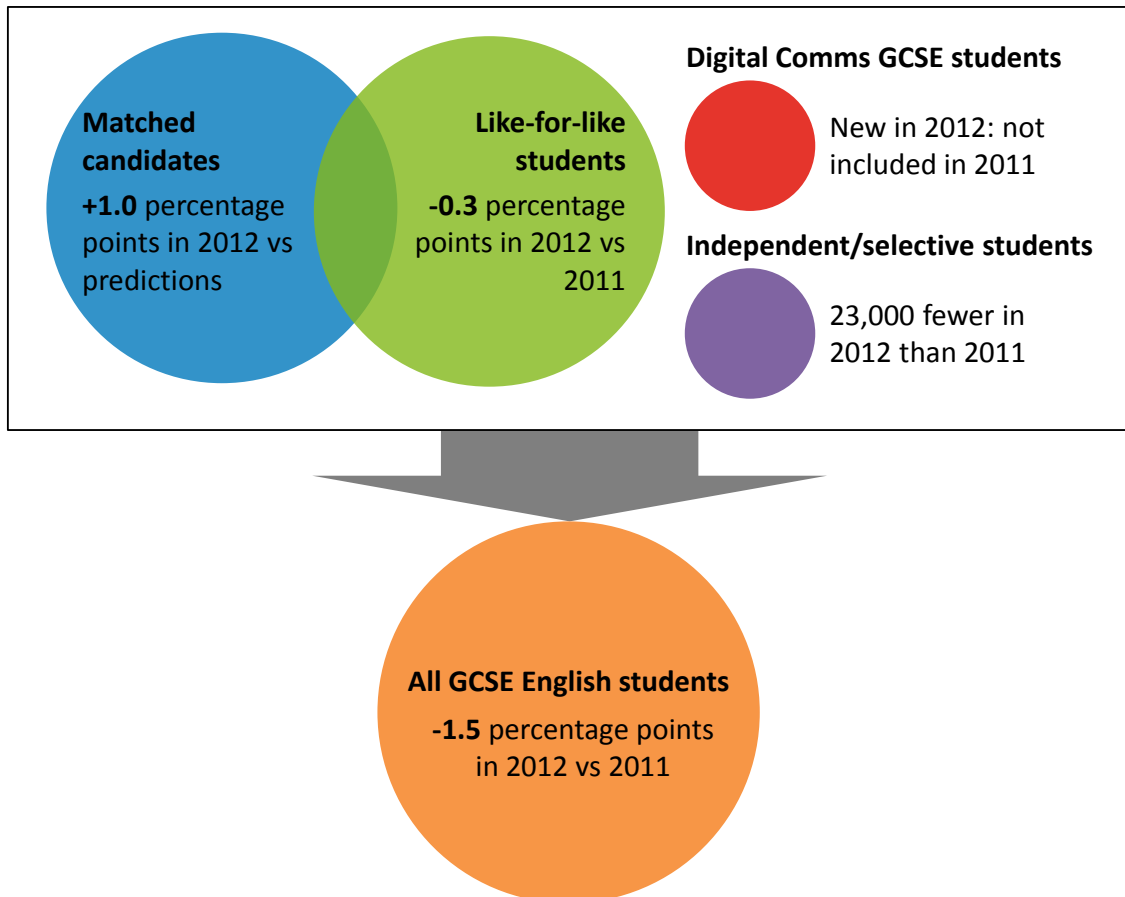
5.41 Results for matched candidates were higher than predictions. This increase does not carry through into the overall figures because of other differences in the overall mix of students in summer 2012. About 23,000 students from independent and selective schools, those who would generally be expected to

³¹ The prediction is slightly different because there were more matched candidates included in the revised figures.

³² Matched candidates are those students who can be matched to their attainment in Key Stage 2 National Curriculum Test in England, in this case from 2007. In summer 2012, over 80 per cent of the students were matched to their Key Stage 2 results.

get high grades, did not take GCSE in summer 2012. This had a negative effect on the overall figures.

5.42 The results reported in August 2012 for “English” also included a new qualification (digital communication) that was not available in 2011. Results for this qualification were lower than for English so it also had a negative effect on the overall figures. We explain this in more detail in Appendix 2. We therefore conclude that the like-for-like variation is 0.3 percentage points, well within normal variation.



6. Results for individual schools

- Schools are under extreme pressure to set high targets for their students and to meet them.
- The variation experienced in some schools' results in GCSE English this year is greater than would normally be expected in the first year of a new qualification.
- We have found evidence that marks given by schools for their students' controlled assessments were anticipating the grade boundaries that exam boards would set.

6.1 In this section we explain that GCSE results routinely vary, school by school, particularly when new qualifications come in, and if schools change from one exam board to another. We go on to explain the difficulties examiners faced in GCSE English when they were awarding some of the controlled assessment units in June 2012. Across the exam boards, there are 18 controlled assessment units, out of 34 units in total. They were faced with evidence of seemingly widespread over-marking by teachers, and they took this into account, properly, as they set grade boundaries in June. From interviews we have conducted with schools and from other sources we know that the way schools are held to account – the way they are measured and rated – influences the way controlled assessments are taught, administered and marked in schools.

6.2 We start by considering the way schools are held to account. We then explain how GCSE results do vary to some extent each year, and look at the variability of results in English this year. Finally we summarise the evidence we now have about results in controlled assessment units. It suggests very specific preparation of students by teachers, so that students would reach perceived grade boundaries, and grades for the whole qualification. It also shows the over-marking of controlled assessment units by some schools, as seen by examiners in June 2012.

Accountability measures

6.3 Grade C in English/English language (and also mathematics) is a critical qualification for students progressing to further education or employment. It is also a critical qualification in the measurement of state schools, academies and colleges as well as for assessing teachers' individual performance. We have found evidence that this can lead to undue pressure on schools in the way they mark controlled assessments. A recurring theme in our interviews with schools was the pressure exerted by the accountability arrangements, and the extent to which it drives teachers to predict and manage grade outcomes.

- 6.4 It was clear that most state-funded schools are strongly focused on maximising the percentage of C grades at GCSE. The government's three headline accountability measures (5+ A*–C GCSEs including English and mathematics, 5+ A*–C GCSEs and the English Baccalaureate³³) are all calculated from the proportion of students achieving at least five C grades. Several other measures are now published in performance tables that do not rely exclusively on C grades: for example, Best 8 Value Added, Progress in English, Progress in Mathematics and Average Point Score (but these were rarely mentioned by teachers in our interviews).
- 6.5 When inspecting schools, Ofsted looks at the whole picture including all of these measures. Nevertheless most schools appear to see the proportions of C (or better) grades achieved, particularly in English and mathematics, as by far the most important criteria against which they are judged.
- 6.6 These measures, together with the floor standards³⁴, have led many schools to focus their efforts and resources on intervention strategies to maximise grade Cs in both English and in mathematics. Schools tailor individual plans for each student, track the progress that each student is making and allocate specific extra support if needed, until they are confident that the key grade C will be achieved. It is also important for schools to maximise the overlap between C+ students in each of English and mathematics. A school with poor overlap might have 65 per cent of students achieve C or better in each of English and mathematics (and all of them achieve at least three Cs in other subjects), but still achieve only 50 per cent on the "5+ A*–C GCSE including English and mathematics" measure.
- 6.7 Many schools therefore run highly targeted cross-subject interventions to maximise the number of borderline students achieving Cs in both English and mathematics. This can lead to the intervention focus for a particular student shifting away from English to mathematics (or vice versa) once they are believed to be a "safe" C grade in that subject.
- 6.8 One teacher explained that he "used the controlled assessment and exam marks achieved in January to work out what students needed to do [in June]." Another Head of English said that their senior management wanted teachers to

³³ The English Baccalaureate is a school performance measure. See: www.education.gov.uk/schools/performance/secondary_11/s3.html (accessed on 31st October 2012).

³⁴ See: www.education.gov.uk/schools/performance/secondary_11/s8.html (accessed on 31st October 2012).

“secure” the English grade and then “take the student on to another subject, such as maths.”

- 6.9 In this context, predicting grades has become important in its own right. Schools report that Ofsted now expects them to predict grade outcomes, and to define and act on the steps required to improve students’ grades. Headteachers also want the reassurance of accurate predictions to ensure that their school will meet its targets. School-level predictions in English and mathematics therefore carry particularly heavy weight.
- 6.10 Many teachers considered the ability to predict grades accurately as a fundamental part of their role as a professional. As one Head of English commented: “the pressure on us is enormous and comes from pupils, parents, headteachers and the accountability system. [Teachers] have no choice but to use their best judgement to predict/translate grades from marks.”
- 6.11 Ofqual research into the impact of high-stakes GCSE exams in England³⁵ shows that teachers and headteachers feel pressure to improve GCSE results. Over three-quarters of headteachers felt under pressure from national government to improve students’ outcomes at GCSE – 46 per cent felt under “almost constant pressure” and 29 per cent felt under “a lot of pressure”.
- 6.12 In many schools the prediction process is supported by data analysis which is updated frequently. In a typical school marks from student class work and mock exams in Years 10 and 11 were fed into the tracking system every six weeks, and senior management met every two weeks to discuss them. When students were not making enough progress to achieve their target outcomes, interventions were arranged. This process of targeting teaching and learning support to secure target grades has in many schools been successful in previous years, which added to schools’ surprise when this year’s results were not as predicted. Teachers commonly said that they were used to their predictions being “spot-on”.
- 6.13 As a result of this pressure, any unexpected variation in English results is very keenly felt in schools.
- 6.14 Interviews also confirmed that some schools focus on getting students on the borderline of a C grade to a “safe C” level. Teachers often focus on where they think a C grade will be, which may not be where grade boundaries are ultimately set.

³⁵ Ofqual (forthcoming in 2013) *Impact of high-stakes National Curriculum tests and public examinations in England*.

- 6.15 Teachers told us that their understanding of a “safe C” comes from historic grade boundaries (as published by the exam boards), and their own professional judgement of the standard of work required to achieve a C grade.
- 6.16 Some schools told us that once students are seen to have reached a “safe C” standard, they then prioritise other subjects over English. However, when the controlled assessment grade boundaries in June 2012 were higher, some students who had been thought to be “safe C” grades received D grades.
- 6.17 We have also seen comments that teachers have posted on the TES Forum³⁶ that reflect the pressure they feel under the accountability system and how this impacts on their approach to the qualifications, in particular, to controlled assessment units. For example:

“I’m sure I’ll feel better about it when the coursework forms have all gone off, but will still be nervous about moderation and grade boundaries etc – feel it’s unfair to have intense pressure on English results on one hand, and a woolly, unreliable, loose assessment system. This is particularly chronic when the people under pressure are the same people who have to then administer and invigilate the exam – a recipe for stress, teaching to the test³⁷.”

“However, the real problem is that no matter what syllabus we teach, we will still be expected to get students up to grades that are unrealistic and we will still be expected to 'teach to the test' to get them through. The pressure on teachers to get results is preventing us from doing any real teaching³⁸.”

Some variation in schools’ results is normal

- 6.18 Stable results at national and exam board level can mask variations at school level. Even when national results are virtually unchanged, there will be a wide range of outcomes at school level: some will have similar results, some will do better and some will do worse. This happens in every subject every year and can be caused by a combination of different factors:

³⁶ www.tes.co.uk/forums.aspx

³⁷ TES Forum, posted March 2012.

³⁸ TES Forum, posted March 2012.

- a different mix of students from one year to the next
- different teaching staff or approach to teaching
- a different exam board being used.

6.19 New qualifications do lead to more variation in school-level results. Edexcel, in its report to us, notes that “it is not an unusual phenomenon for grade outcomes to fluctuate, sometimes quite markedly, from year to year, when a new specification is introduced³⁹”. OCR also found that there was “slightly more variability in English than in other subjects in year of specification change, but not considerably more⁴⁰”.

6.20 We plotted the variation in school results (the percentage of their students being awarded at least a grade C) in English and a sample of other subjects for:

- three years before a new qualification, and
- the year when the new qualification was introduced (2012 in English, 2011 in the other subjects).

6.21 The charts on the next two pages show the spread of year-on-year changes in outcomes at school level⁴¹. The greater the spread of the distribution, the greater the year-on-year variation of results. In the first year of the new qualifications in French, geography and history, the spread of year-on-year variation was greater than in previous years.

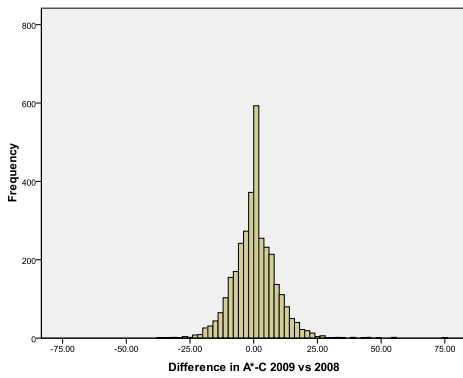
³⁹ Edexcel Report, October 2012 (Annex C, pp.13–14).

⁴⁰ OCR Report, October 2012 (Annex D, p.21).

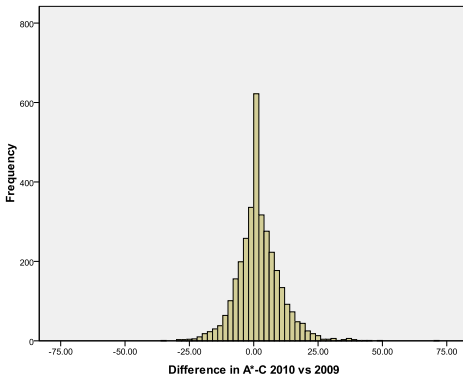
⁴¹ 2008/07 – 2011/10 data is from the National Pupil Database, subject is classified as the subject for discounting purposes. 2012/11 data for English is provided by awarding organisations and the subject is classified as English or English Language. Centres were only included if they had at least 30 results in both years.

English

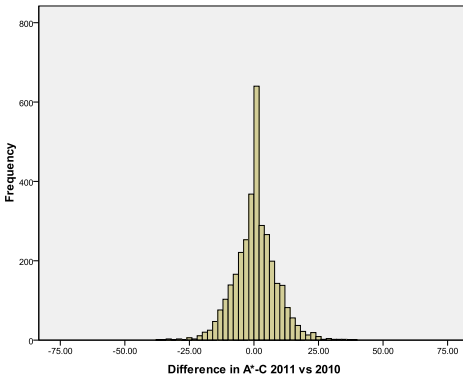
2009 vs 2008



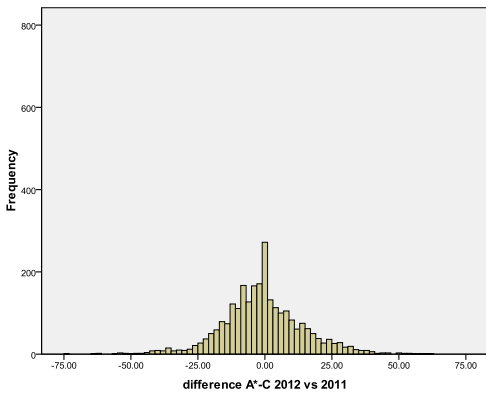
2010 vs 2009



2011 vs 2010

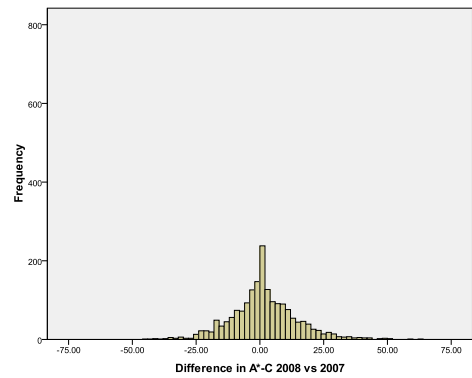


2012 vs 2011

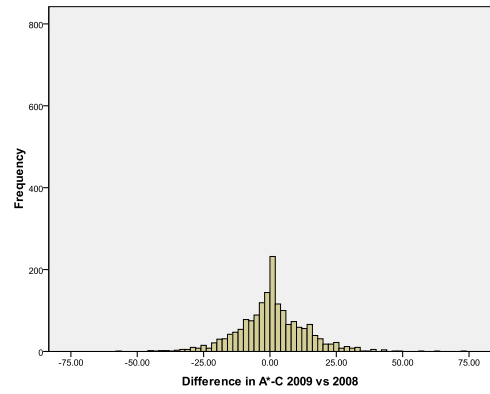


French

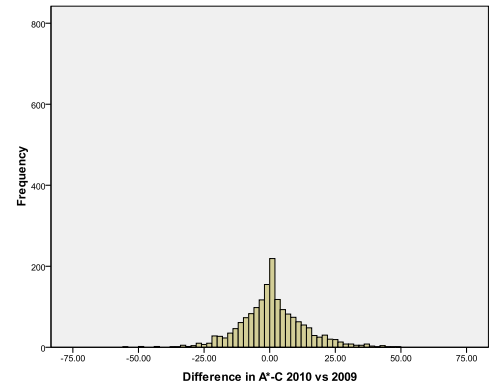
2008 vs 2007



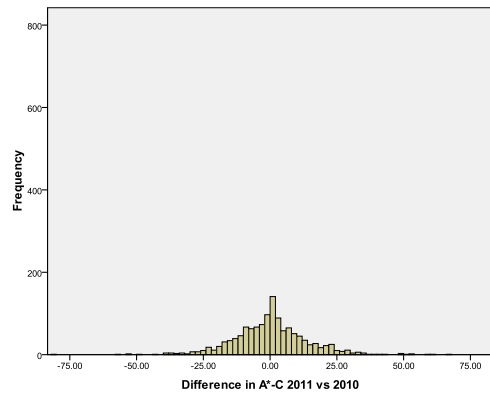
2009 vs 2008



2010 vs 2009

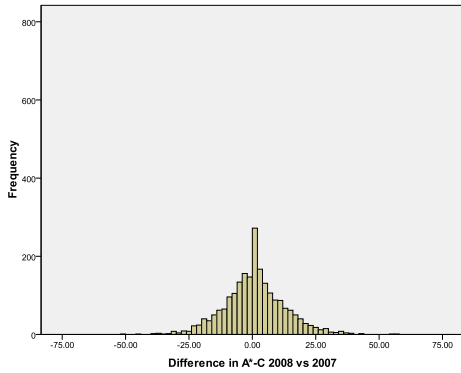


2011 vs 2010

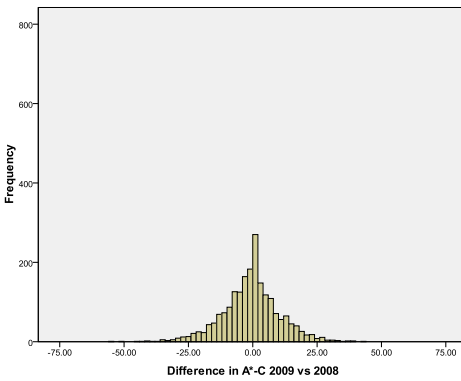


Geography

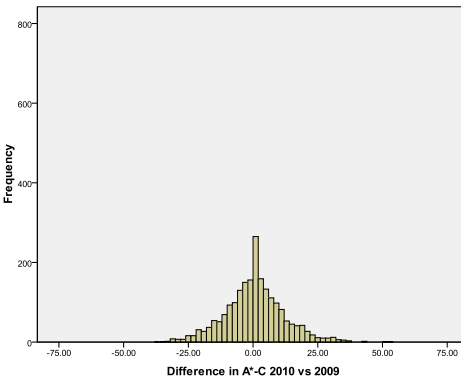
2008 vs 2007



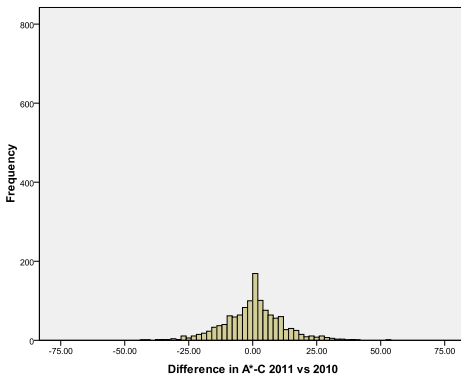
2009 vs 2008



2010 vs 2009

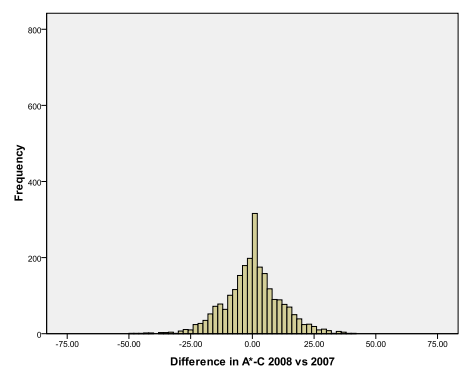


2011 vs 2010

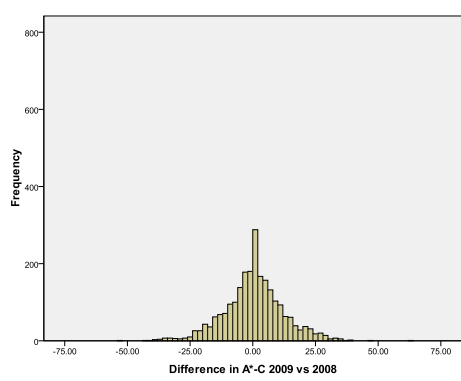


History

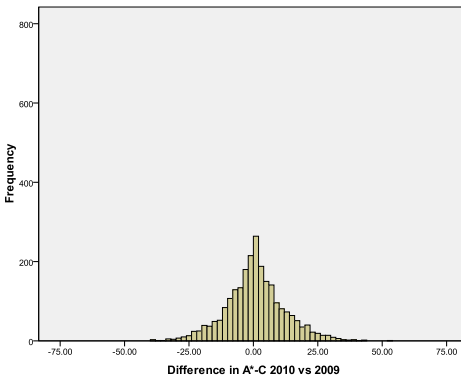
2008 vs 2007



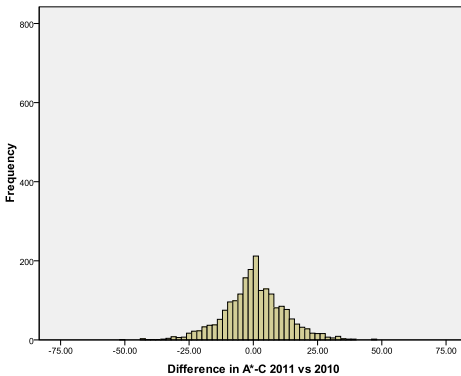
2009 vs 2008



2010 vs 2009



2011 vs 2010



6.22 **The new qualifications in English:** similarly, we found that there was more variation in English in 2012 than there had been in the three years before the new qualifications were introduced but the change in variability was greater than for French, geography and history. The extra variability in this year's results in GCSE English is caused by a combination of:

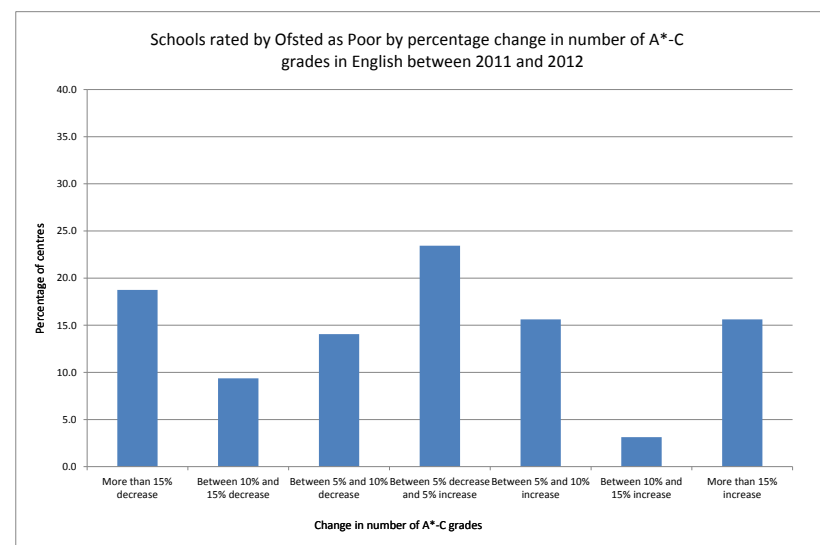
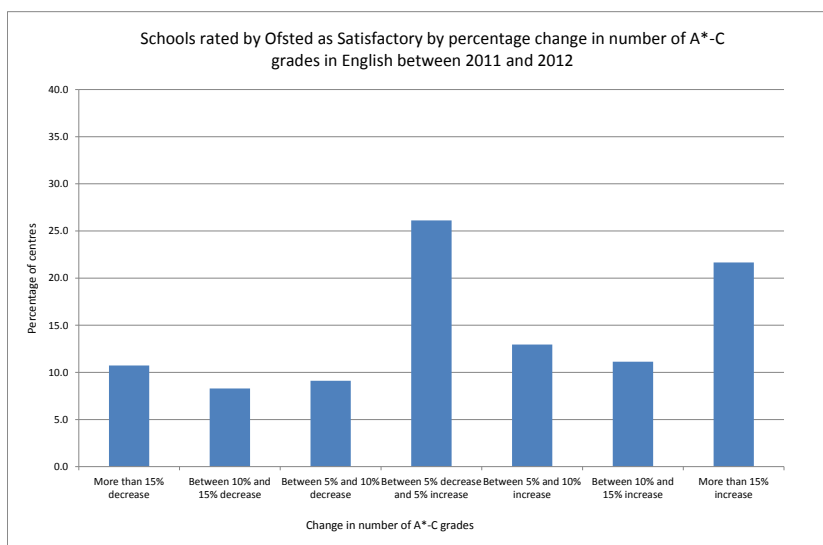
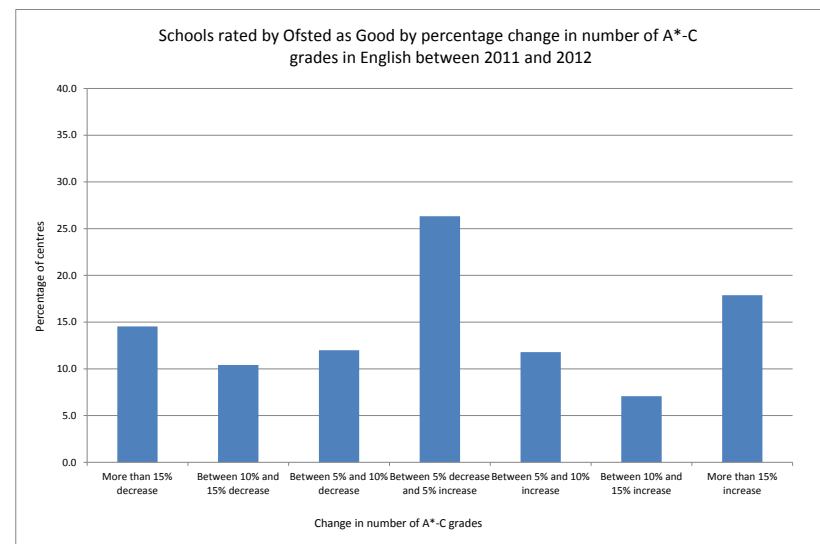
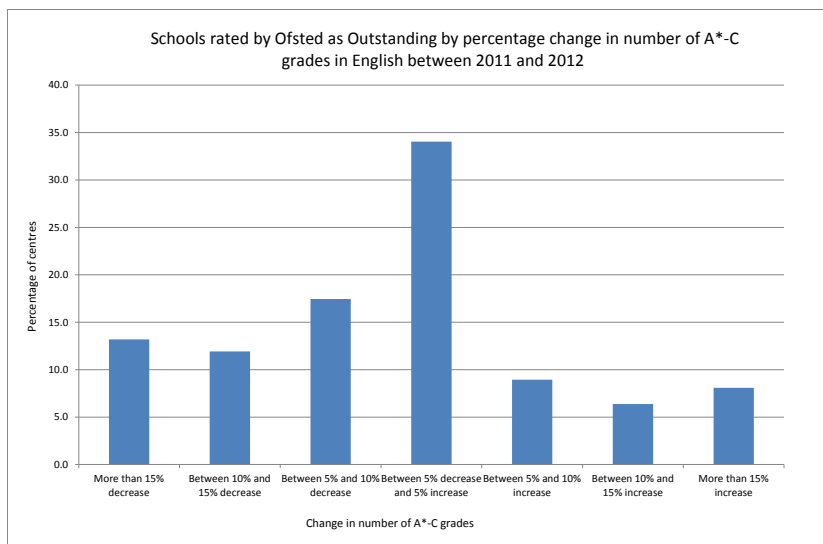
- the large number of options that are available through the qualifications from which each school has to select (the “route effect”):
 - between the single English qualification and the two language and literature qualifications
 - between foundation tier and higher tier written papers
 - choices of tasks for the controlled assessment units
 - when to enter students for each unit, and whether students should re-sit any unit;
- the evidence from exam boards, supported by our further analysis, that some schools over-marked some controlled assessment work. This has created a discrepancy in the standards set by schools and this could not be fully corrected through the moderation and awarding process. There is also evidence that schools have interpreted differently the rules that set out how controlled assessments should be carried out. This will have led to variability between schools of the standards of work from their students. We discuss this in more detail later in this section.

6.23 It is not possible to apportion the increased variability in this year's English results to each of the various factors.

6.24 **Variability by type of school:** We have analysed the variability across types of schools. We know that independent schools and selective state schools experienced least variability in results. Sixth form and further education colleges experienced a marked downturn in results. These differences are probably explained by the nature of the student mix in each type of school. Independent and selective schools typically have a much larger proportion of their students achieving well above the C grade boundary. Sixth form and further education colleges have a higher proportion of students that are re-sitting the qualification.

6.25 We have also seen some differences in the variability of results between schools in relation to their Ofsted rating. Those rated as “outstanding” are the most likely to have retained comparable results year-on-year but, if there has been a change of more than 5 percentage points, they are more likely to have had a decrease in the percentage of students that achieved at least a grade C

this year. Those rated “good” were equally likely to see their results improve or decline. Schools rated “satisfactory” were more likely to have improved their performance on 2011. The results are shown in the following charts.



- 6.26 The graphs on pp44-45 show that the old GCSE English qualifications, which were largely unchanged since 2002, were exceptionally stable. Schools will have become accustomed to greater stability in English than for other subjects.
- 6.27 Schools interviewed were clear that the historical stability in English made them confident about the grades they had predicted for their students. The variation in 2012 was not only unexpected but also seen as so abnormal as to indicate something wrong, with one teacher noting that this was “beyond what might be expected as normal volatility.”⁴²
- 6.28 To put this perception in context, we know from AQA analysis “the legacy GCSE English specification in summer 2010 and 2011 delivered less variable centre [school] results than any other specification modelled, old or new, meaning that the level of stability in the legacy specification was atypical⁴³.” Even with the introduction of the new GCSE, AQA English Language results were no more volatile this year than those in history, geography and French. Only the single English GCSE was more volatile.
- 6.29 One reason why the previous AQA qualification might have had very stable results was its highly predictable assessments. Getting the right level of predictability is always a judgement call: too little predictability and students and teachers cannot prepare for the sorts of questions on the paper so results do not reliably reflect students’ abilities; too much predictability and students and teachers can focus on a narrow range of content or skills.
- 6.30 We know that some of the previous AQA assessments were overly predictable and meant that students knew what sort of questions to expect when they sat their written papers.
- 6.31 The previous syllabus had two written papers. In each paper Section B assessed writing and was worth 50 per cent of the marks. Section B of the old paper 1 gave students a choice of writing to “argue”, to “persuade” or to “advise”. Section B of the old paper 2 gave students a choice of writing to “inform”, to “explain” or to “describe”.
- 6.32 The context changed in each paper, but the choice of “argue/persuade/advise” on paper 1 and “inform/explain/describe” on paper 2 did not. Schools could focus teaching on just two of these modes of writing in the knowledge that they would always appear, and students could choose to focus their preparation on,

⁴² Capgemini report, October 2012 (Appendix 1).

⁴³ AQA Report, October 2012 (Annex A, p.23).

for example, writing to persuade and writing to explain, knowing that there would be a question on each.

- 6.33 In contrast all questions on the new AQA written paper are compulsory. Students have no choice and they do not know what modes of writing will be tested. The writing questions are still worth 50 per cent of the marks but students may be required to write in any of five modes. The table shows the pattern of questions to date.

	Section B questions	
January 2011	explain (16 marks)	persuade (24 marks)
June 2011	describe (16 marks)	persuade (24 marks)
January 2012	explain (16 marks)	argue (24 marks)
June 2012	inform (16 marks)	persuade (24 marks)

- 6.34 This is one way in which these new qualifications were different. We have already discussed other ways in section 4. Some of these changes will have subtly increased the level of demand. Decisions made at awarding will aim to take that into account and the comparable outcomes approach will have helped to ensure that increased demand did not make it any harder this year for students to get a particular grade.
- 6.35 **Balance in the design of qualifications:** we also know that the old English GCSEs were not well balanced across the different components – students typically did much better on their coursework than on the written papers. Results from AQA in 2011 illustrate this⁴⁴:

⁴⁴ AQA Report, October 2012 (Annex A, table 22).

Unit level outcomes for AQA English/English Language units in June 2012

	Type of assessment	Weighting	Proportion of students getting at least a C on the paper/coursework
Foundation tier	written paper	30%	4.4%
	written paper	30%	4.5%
	coursework (speaking)	20%	70.9%
	coursework (written)	20%	56.2%
Higher tier	written paper	30%	66.6%
	written paper	30%	75.9%
	coursework (speaking)	20%	98.7%
	coursework (written)	20%	97.1%

6.36 This imbalance in the previous qualification was not apparent to teachers and students because grades were not issued for individual components.

6.37 In contrast, the table below shows the picture for the AQA units in June 2012.

Unit level outcomes for AQA English/English Language units in June 2012

Type of assessment	Weighting	Proportion of students getting at least a C on the paper/controlled assessment
Foundation tier written paper	40%	10.2%
Higher tier written paper	40%	66.3%
Speaking and Listening controlled assessment	20%	76.2%
English controlled assessment	40%	28.8%
English Language controlled assessment	40%	73.8%

6.38 The balance across these new units is more in line. Exam boards have told us that they were working to establish a better balance across the different units in the new modular qualifications.

6.39 Many of these changes ought to have meant the qualifications worked better. But there were weaknesses in them that made them particularly susceptible to pressures from the accountability system.

Controlled assessment

6.40 We have already noted the high proportion of controlled assessment that is used for GCSE English and English Language, 60 per cent of the total. One-third of this is the speaking and listening assessment, which is marked by teachers and evidence of performance is ephemeral as no records of actual work are available. The remaining two-thirds of the controlled assessment are written tasks, also marked by teachers. A sample of candidates' work is provided to the exam board for moderation.

6.41 When we interviewed teachers, some commented on how controlled assessment tasks were carried out. They were concerned that the exam boards' rules were ambiguous and therefore could be interpreted differently by schools. They were also aware that some schools sought to "stretch" their interpretation of the rules to such an extent that this created a significant advantage for those schools. Here are some comments that teachers made to Capgemini:

"Initially it was considered acceptable to collect student work during the controlled session and give some feedback, then not; initially any piece of paper could be used for the plan taken into the controlled session, subsequently only the board's official plan could be used; initially the plan was to be detailed, then succinct, and subsequently they were criticised for not having enough detail."

"The school felt that proper regulation of CA was well-nigh impossible and that controlled conditions were being interpreted very differently in different schools. This was creating suspicion and distrust between schools."

"Wide variety of methods for putting CA in place, such as students writing a first draft which was then given written comments by teacher. This being subsequently written up by a student. Doesn't feel like a level playing field."

"There was confusion in advance about how to operate CA – for example what exactly was to be included on 'top sheets' was never really clear."

[re. Speaking and Listening unit] "There was no clarity on how teachers were to maintain records. There was no format supplied so we had to rely on our own judgement. AQA published something but quite late."

"The mark scheme is so vague you can drive a coach and horses through it."

6.42 We have also looked at comments that teachers have posted on the TES Forum. For example:

“Just got on to this thread – and I feel I am being made to cheat. I've taught the kids and then let them do the tasks – we have to do them in the classrooms, except for those who need access arrangements, who are under the beady eye of external invigilators. I taught my kids, gave them the opportunity to make notes, and then did the damned things like an exam. Result? Lots of them underperformed against their targets. Not good enough. This work, I am told, is really coursework, and has to be at target grade, or they will not reach their targets at the end of the course. Others in the department have done marked drafts. I'm now feeling pressured to get some of mine to redo various pieces. I've voiced my objections, but have been told that the long and the short of it is that they have to be nannied through at every stage – there is disbelief when I say that some schools are doing the CAs as exams. I resent the implication that I am failing my kids, when actually what they produce is probably more accurate as an indication of their abilities than their target grades are. The sooner this nonsense is stopped and we go back to 100 per cent exams, the better⁴⁵.”

“I have only been teaching for four years but have become very concerned over the way we are carrying out our controlled assessment practice. We have always completed these assessments under strict conditions – timed and in silence etc. Now however SMT are in a panic about results. All year 11 are constantly in study sessions to get other subjects up so English has really suffered. We have year 11 students with missing and/or incomplete controlled assessments. We have been asked to rework controlled assessments, mark them and give them back for improvement. In some cases we are virtually writing them for the students. This to me is no different to coursework and raises the issue of why coursework was replaced. I know that we cannot be the only centre doing this but how do you feel about this⁴⁶?”

⁴⁵ TES Forum, posted in November 2011.

⁴⁶ TES Forum, posted in January 2012.

"We've had so many issues over what is acceptable and what is not acceptable. Have tried to play things by the letter of the spec and then are told that things still aren't right. No-one seems to be able to agree on what constitutes what mark and the board haven't helped the situation by being, as you so rightly say, woolly over everything⁴⁷."

"I suppose reflecting on all the contributions made here, as well as the experience of administrating and submitting the CA marks, my abiding view is that it just doesn't feel right as a way to manage our GCSE assessment. I think a fundamental problem is that the stakes are just too high – filling in the mark sheets, I felt really aware that so much was riding on it – obviously we are happy to take on the responsibility of helping the students to do as well as they can, but it also feels like the reputation of myself, my colleagues and the school as a whole is at stake. This is a problem because it creates all sorts of strange and unhealthy pressures, I think, making teachers simultaneously anxious about if CA marks are going in too high, or too low! I suppose my basic point is that, if we are going to put so much emphasis on these marks, created by league tables etc, then at least assess the work externally so that a level playing field might be achieved, and schools can concentrate on becoming experts in promoting learning, rather than working as exam markers. I'd feel better about it if it wasn't English, with the inevitable subjectivity of the marking, and if it was possible to get real consistency in the implementation of the CAs. It's clear that with formal, externally marked exams, fair uniformity can be pretty much achieved – the variables can be controlled so that the conditions of assessment in every school can be pretty much the same. The sheer number of variables mean that this can never ever be achieved through CA. I can see that the exam boards are trying to provide clear guidance, but what are we expected to do with advice such as: "Can CA sessions have starters and plenaries? No – there can be no teacher intervention once a CA session has started." Here's an attempt at clarity, at providing the uniformity of conditions needed for fairness. But hang on – in the real world, what does this mean? How do we decide when a CA session starts? Does this mean we can't speak to them before the lesson? In the dining hall? At the end of whatever lesson they happen to have before English? In the corridor? Although I can see the thinking behind CAs, it does not work in practice, and it's too important for students, teachers, schools, communities to tolerate this sort of looseness. In my

⁴⁷ TES Forum, posted March 2012.

view, the system is flawed, and it's also insidious and corrosive to morale to ask English departments to summatively mark 60 per cent of the GCSE⁴⁸.”

6.43 The result is that students have benefited to varying degrees between schools in the way they have carried out their controlled assessments and therefore the same mark reflects different standards of work. This will have contributed to the greater variation in schools' results this year, the first year in which controlled assessment has been used for English qualifications.

Controlled assessment – patterns of marking

6.44 We have also looked at the patterns of marking in controlled assessment. We know that in the old coursework units there was “bunching” of marks at grade boundaries. Several of the Chairs of Examiners we interviewed told us that there had been evidence of coursework mark “creep” in the old qualifications, with schools' marks tending to move upwards when boundary marks were moved upwards.

6.45 There is some evidence of a similar effect in the new qualifications. AQA found instances where schools appeared to be marking to grade boundaries set in previous series⁴⁹. This is shown in the following extracts from the raw mark distribution for five schools sitting English Language reading/writing controlled assessment in June 2012. The grade boundary set in January 2012 for this unit was 43 marks.

School	Size of school entry	Number of candidates scoring 42	Number of candidates scoring 43
A	184	0	20
B	154	0	16
C	109	0	13
D	175	0	11
E	145	0	9

6.46 There were also several references to over-marking in the reports of the award meetings. For example:

⁴⁸ TES Forum, posted May 2012.

⁴⁹ AQA Report, October 2012 (Annex A, table 30).

“In some cases the assessment indicated by annotations and summative comments was very accurate, but the numerical marks did not reflect these comments and in some cases were fixed just into a grade boundary from a previous series... There was some evidence that schools had marked accurately but then given numerical marks to push the candidates into what they thought were grade boundaries⁵⁰.”

Speaking and listening controlled assessment⁵¹

6.47 One third of GCSE English controlled assessment is ‘Speaking and Listening’. Work is marked by teachers but it is not recorded and only a third of schools are visited in any one academic year. Therefore any over-marking on speaking and listening is hard to spot and control. The Chairs of Examiners told us they had concerns about the accuracy of marking of speaking and listening work. This is also referenced in the exam boards’ reports to us. Edexcel states that “the moderation of the Speaking and Listening component [of Unit 3] creates particular challenges due to the ephemeral nature of the evidence available for moderation⁵².”

6.48 In its report, AQA notes that: “moderation processes are constructed to support teachers in applying a consistent standard across [schools]. Relying on a sample of work (in the case of Unit 3) and upon observation (in the case of Unit 2) the moderation process does not, and cannot, rigorously “police” teacher marking. It relies on the professionalism of the teacher community. Teachers are aware of the marking tolerances any assessment is subject to and they are obviously aware of the ephemeral nature of the assessment of speaking and listening⁵³.”

6.49 Concerns were expressed during award meetings about over-marking by teachers. For example: “there is much evidence from the field that, despite impressive efforts by the Principal Moderator and his team, teachers are

⁵⁰ Edexcel Principal Moderator’s Report to the Awarding Committee (unit 5EH01), June 2012.

⁵¹ See p27, GCSE English Examining Process.

⁵² Edexcel Report, Annex C (p.10).

⁵³ AQA Report, Annex A (p.31)

marking candidates higher than the standard exemplified on AQA [speaking and listening] DVDs⁵⁴.”

Reading/writing controlled assessment

6.50 Two-thirds of GCSE English controlled assessment is Reading/Writing. Schools mark their students' written controlled assessment work. Each school sends a sample⁵⁵ of work to the exam board moderator. The moderator re-marks a sub-sample⁵⁶ of the work and if the difference between the school's marks and the moderator's marks are less than the marking tolerance⁵⁷, the school's marks will be accepted.

6.51 If any of the differences in marking are greater than tolerance, the moderator will re-mark the rest of the sample (so the whole of the sample will have been re-marked) and submit those marks to the exam board, so that an adjustment⁵⁸ can be calculated. Where that adjustment means that all mark changes would be within the marking tolerance, the exam boards will revert to the school's marks rather than make an adjustment.

6.52 At our request, AQA, the largest provider, analysed a random sample⁵⁹ of its moderation records for schools that were in tolerance⁶⁰. For English Language 2,332 (95.2 per cent) of the 2,450 schools entered had their marking judged as 'within tolerance', and for English this was 2,008 (92.4 per cent) of the 2,172 schools entered. The results from this sample show clear evidence of over-marking in both English and English Language.

⁵⁴ AQA Chair of Examiners' report on the summer 2012 award.

⁵⁵ The sample is a minimum of ten pieces of work, including the highest and lowest marks. The sample size increases in line with the number of students entered, up to a maximum of 20 pieces of work.

⁵⁶ Usually half of the sample.

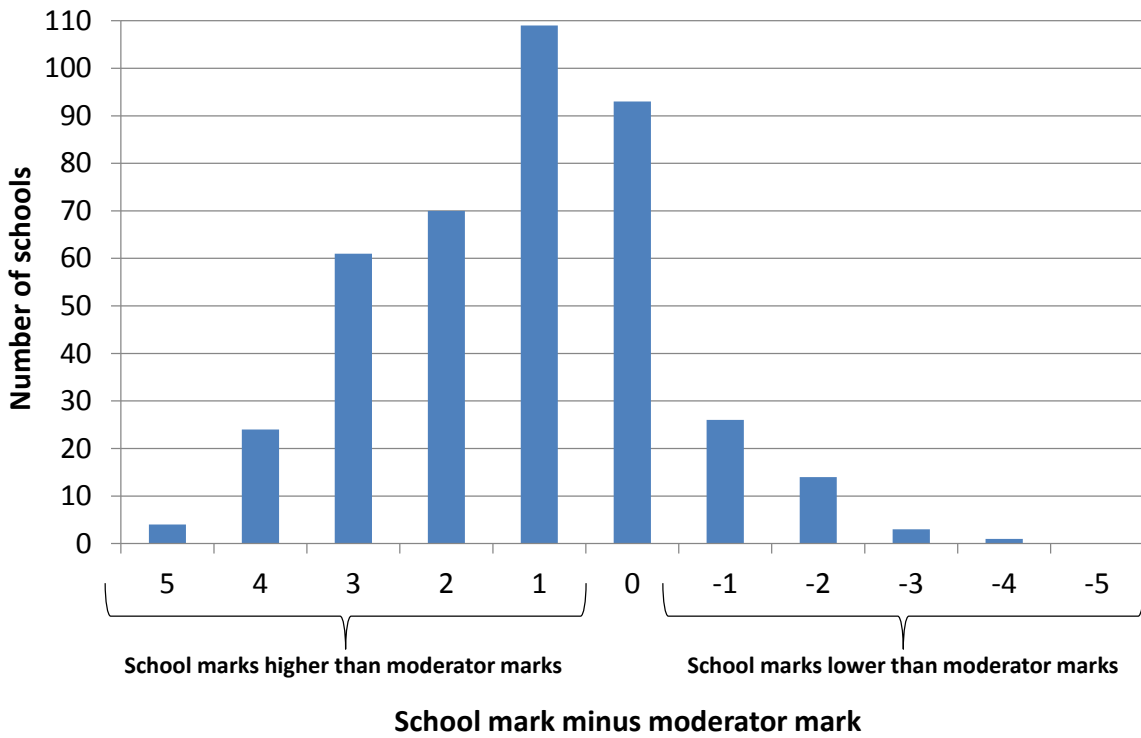
⁵⁷ This is 6 per cent of the maximum mark for the unit, which is usually rounded up where this is not a whole number. The same percentage tolerance is used for all GCSEs and A levels.

⁵⁸ The adjustment will be calculated using a regression program, intended to preserve the school's rank order and ensure adjusted marks are closer to the moderator's marks.

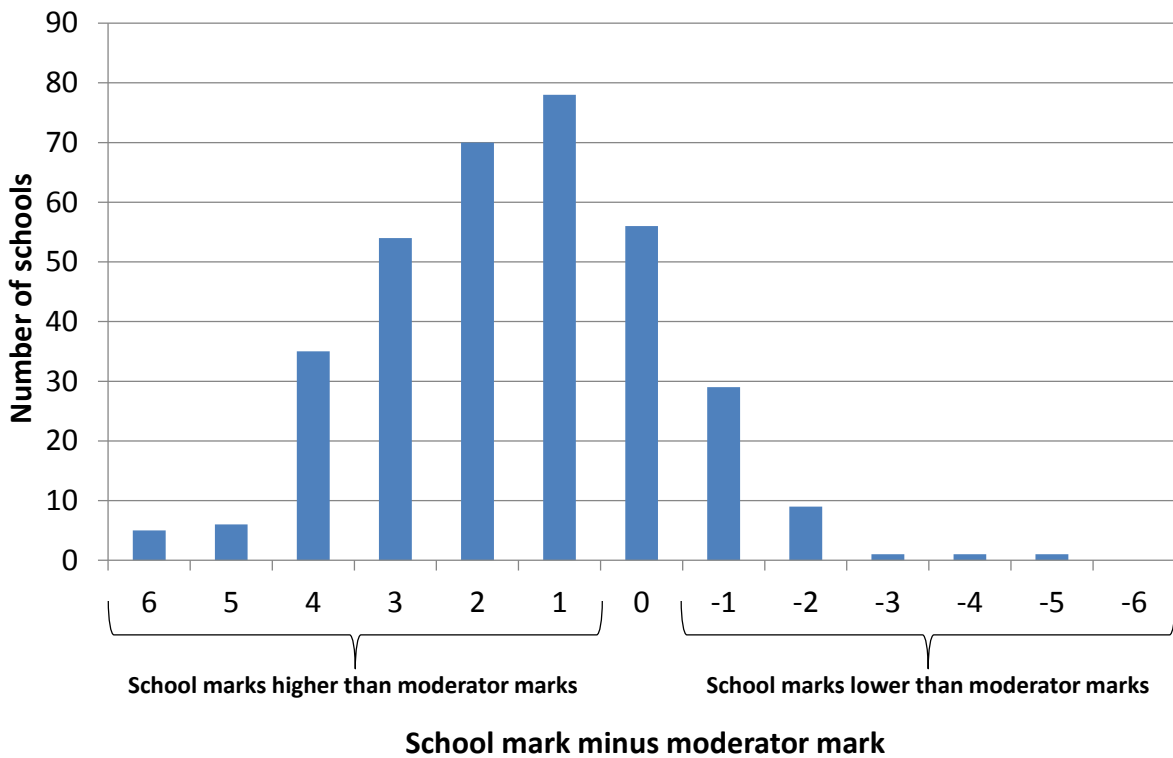
⁵⁹ Approximately every fifth centre.

⁶⁰ The records for those schools that are within tolerance are kept on paper only.

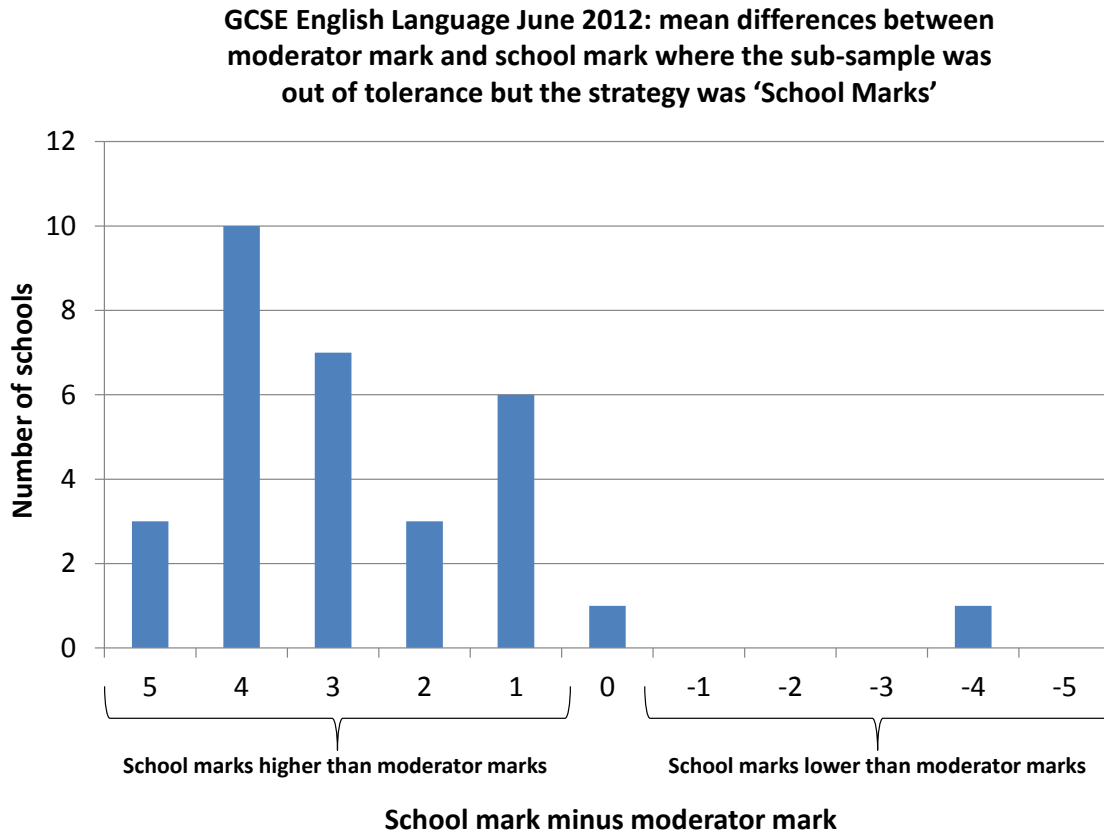
GCSE English Language June 2012: mean differences between moderator mark and school mark for schools in tolerance



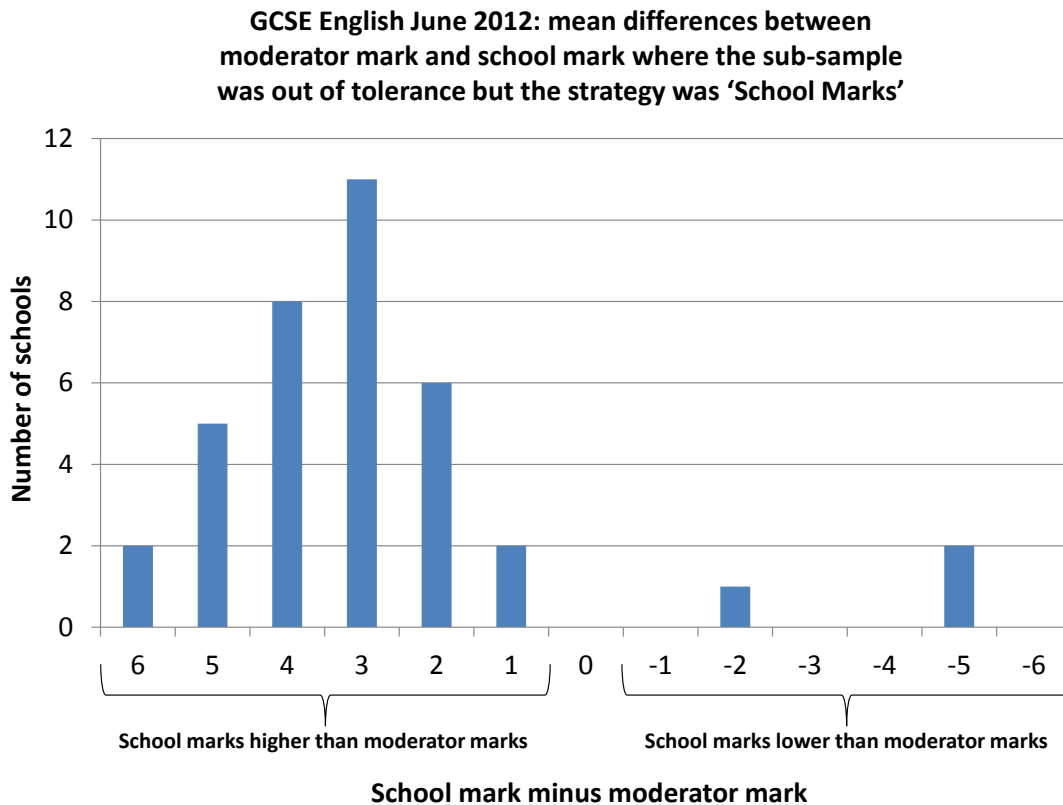
GCSE English June 2012: mean differences between moderator mark and school mark for schools in tolerance



6.53 AQA also analysed marking patterns for the relatively small number of schools (32 for English Language and 37 for English) where marking was not in tolerance but where the school's marks were used⁶¹. Again there was clear evidence of over-marking.



⁶¹ This is because the calculated adjustments were all within tolerance and so the school's marks were accepted, even though some of the original marking was found to be outside tolerance.



6.54 The marking tolerance of at least 6 per cent⁶² meant that schools could be over- or under-marking by up to 6 marks in some units⁶³ and still have their marks accepted. The tolerance level is long established, and applied in all GCSE subjects.

6.55 Moderation is not able to, nor was it designed to, counter this level of over-marking. Widespread over-marking means that when senior examiners review work, they are likely to set grade boundaries at higher marks than in the previous exam series for the same standard of work, in order to compensate.

6.56 It also appears that some teachers were warned that over-marking would lead to controlled assessment boundaries having to be moved. For example, the Edexcel Chair of Examiners told us that Edexcel had warned some schools at training meetings that over-marking would lead to grade boundaries being moved upwards⁶⁴.

⁶² See Appendix 2.

⁶³ AQA Unit ENG03, Edexcel Units 5EH03 and 5EN03.

⁶⁴ Interview with Edexcel Chair of Examiners, October 2012.

6.57 We have also seen comments posted on the TES Forum that indicate the practices that have contributed to over-marking. For example:

“I've just read my school e-mail to find the instructions for getting the CA folders together, and including the instruction ‘All folders must be at or above target grade.’ This is being done by either getting kids to rewrite CAs after they've been marked, or by fiddling the Speaking and Listening grades to make up for lost marks on the written work. When I've dared to suggest that the CAs should be done in exam conditions and that lots of schools are doing that, I'm told that that is rubbish, that CAs are really coursework, and that we have to cheat because other schools will be doing so, and we cannot afford to let our results slip at all. I'm sorry, but that is wrong, and the demand to have all folders at or above target grade is unrealistic, if the CAs are done in the required conditions. I thank God that I am taking early retirement at the end of this year. The drive to achieve targets is definitely corrupting and I loathe being made to feel that I am not doing right by my students because I am not making them stay behind after school week after week to rewrite the bloody things⁶⁵.”

6.58 Four out of five controlled assessments were submitted in June 2012. Controlled assessment boundaries in June 2012 were different in January in a number of units, in most cases by no more than the marking tolerance for the units. This is not because the boundaries set in January were too low but because the over-marking apparent in June led to inflated marks, which could not be controlled by moderation.

6.59 There were several references in the reports from award meetings to over-marking by schools. For example:

“The committee were unanimous in endorsing the controlled assessment boundaries as there was evidence of significant teacher over-marking⁶⁶.”

“Unit 2 assesses speaking and listening and is entirely ephemeral. While there is training of teachers in the standard, there is no

⁶⁵ TES Forum, posted February 2012.

⁶⁶ AQA Approver's report on the summer 2012 award.

moderation of their marking. This is the same across the awarding bodies and as in the legacy GCSE English specification. It is however unsatisfactory on at least two counts: 1) the lack of evidence upon which to base boundary decisions, 2) the very high marks awarded by teachers⁶⁷.”

6.60 There was also some evidence of dubious practice in controlled assessment. The WJEC examiners’ reports noted the following:

- “There were occasions when the notes were far too detailed and when this happened it was necessary to refer a number of such occasions to the Board for investigation.”
- “Some centres allowed students’ notes for the writing tasks. This is not permitted by the specification.”
- “The administrative rules for controlled assessment also require that students only have one attempt at a particular task. It is perfectly acceptable for them to redo a piece but it must be on a different task. Sadly we did find evidence that in some centres students had been allowed to make two attempts at the same task with the benefit of the teacher’s comments on the first attempt. Where there was proof of this, the matter had to be dealt with by the Board’s malpractice committee⁶⁸.”

Modular structure of the qualifications

6.61 The modular structure meant that students could take one unit at a time throughout their two-year course, or take all units at the end. That meant that exam boards had to set grade boundaries for one unit at a time, ahead of the first full qualification awards in summer 2012.

6.62 This made awarding particularly complex, particularly in early awards (in January and June 2011 and in January 2012). Exam boards had statistical predictions for those early units but those predictions were based on an expectation of how well students would do for the whole qualification at the end of a two-year course. Using those predictions for individual units when students were only a quarter, half or three-quarters of the way through the course was

⁶⁷ AQA Approver’s report on the summer 2011 award.

⁶⁸ WJEC GCSE Examiners’ reports, summer 2012.

difficult, especially when those entries were a relatively small sub-set of the overall cohort.






















- 6.63 The modular system means that at each awarding series the students entered for each unit are an overlapping mix of school years (Years 9, 10 and 11), those doing English and English Language (for the common units), and those sitting for the first time and those re-sitting (including some at further education colleges), and some students (but by no means all) will be entered for more than one unit in the same series.
- 6.64 Even when there are subject-level awards (as in June 2012), the students being graded are still a mix of years (Year 11s claiming their qualification and Year 10s taking early units), qualifications and first-time/re-sitters.
- 6.65 This makes it particularly difficult to separate the different sub-groups and establish or quantify any differential effect. While it might be tempting to make assumptions that might allow us to estimate or quantify any effects, we do not believe it is valid to do so.

Entry combinations

- 6.66 A large minority (38 per cent) of students took this modular qualification in a linear way, with all unit entries at the end of the course. Most others sat a written paper before June 2012 but left a large controlled assessment until June 2012. The “terminal rule⁶⁹” requires students to sit at least 40 per cent of the GCSE at the end of the course, and it is clear that schools overwhelmingly chose to use controlled assessment for this purpose.
- 6.67 Across the exam boards there are, in theory, over 2,500 different ways that a student could achieve a GCSE in English/English Language. This is because schools have a choice of tier (foundation or higher), of exam series (January or June in Year 10 or Year 11), of sequence (exam before controlled assessment, or vice versa) as well as decisions about re-sitting.
- 6.68 The most popular routes through the qualification are illustrated below.

⁶⁹ The terminal rule requires students to sit at least 40 per cent of the qualification at the end of the course.

All Exam Boards: Most Popular Entry Routes, GCSE English and GCSE English Language (UK)

Date of entry				Approx % of students taking this route
January 2011	June 2011	January 2012	June 2012	
			  	38%
			  	11%
			 	9%
			  	7%
			 	4%
			 	2%
All other routes				30%

* Re-sit of the examination.

100% = 677,423

Note that the total exceeds 100 per cent due to rounding.

A “whole-qualification approach” to marking controlled assessment?

6.69 The aggregation of units in the new GCSEs is much more transparent than in the old GCSEs, with UMS marks⁷⁰ simply added together to determine the overall qualification grade. For students who had been entered early for one or more units, schools could then calculate how many more UMS marks each student needed to reach their target grade.

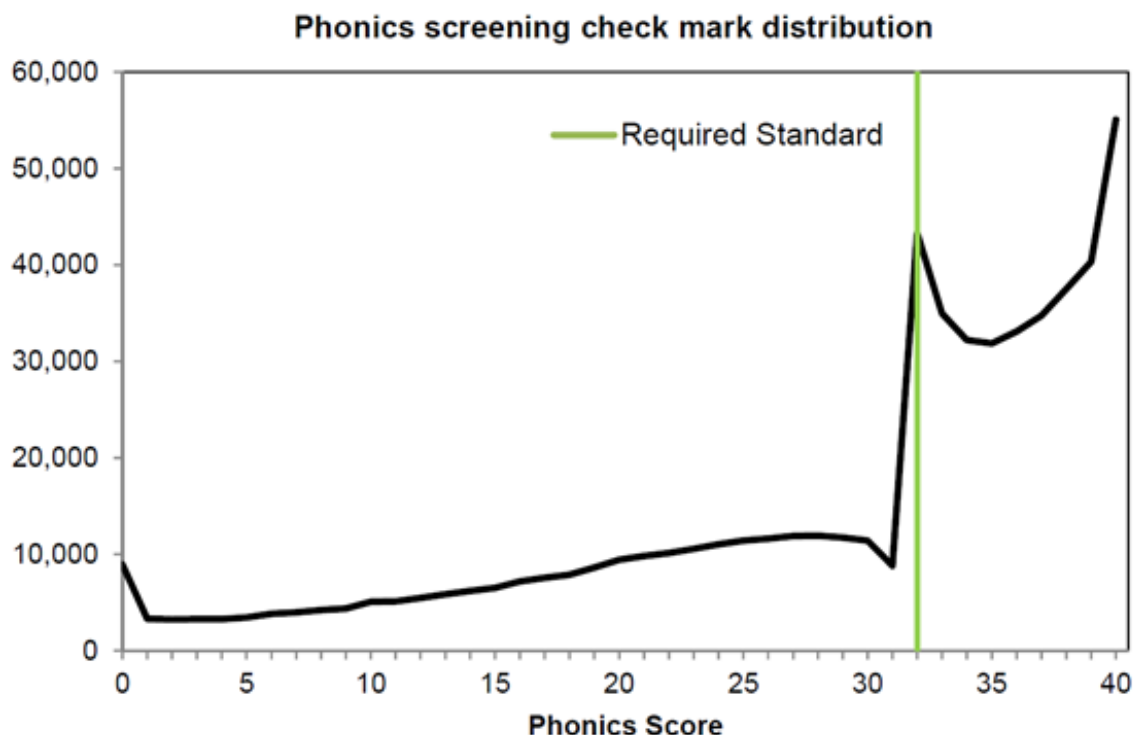
6.70 We know that schools made use of grade boundary information from earlier series (in 2011 and January 2012). Teachers used the grade boundaries to check their predictions and to check that students were on track to reach their target grade. One school described its tracker which showed teachers how many marks were needed for a student to get a target grade, based on the January grade boundaries⁷¹.

⁷⁰ Uniform Mark Scale. See Appendix 2.

⁷¹ Interviews with schools conducted in September and October 2012.

- 6.71 In this way the January 2012 grade boundaries, especially for controlled assessment, were relevant to many students, not just the minority who took units early, because so much subsequent preparation was based on an assumption of what those boundaries would be, and the belief that any changes in June 2012 would be small. Teachers told us that they took note of the warnings that grade boundaries might change, but they thought any changes would be only a small number of marks. Exam boards did make clear that grade boundaries could differ, but it appears that this was often not heeded. Indeed, one of the schools we interviewed told us that it did not occur to them that the boundaries would change.
- 6.72 The evidence suggests a more sophisticated approach to marking controlled assessment for students who had already sat their written paper. For these students, schools knew how many UMS marks they had achieved on the written paper and they could calculate how many more UMS marks each student needed to achieve a particular grade, assuming the June grade boundaries were the same as in January.
- 6.73 We have recently seen that knowledge of the grade boundary influences marking, even in relatively low-stakes assessments. For example, the following graph relates to the outcomes of the first national phonics screening check for Year 1 pupils⁷². The “pass” mark of 32 was made available to teachers before they administered the check. The graph shows how teacher scoring was affected by knowledge of the threshold.
- 6.74 The graph below shows how many pupils were given each score: fewer than 10,000 were given a score of 31 compared with over 40,000 being given a score of 32.

⁷² DfE Statistical First Release, 27th September 2012.



6.75 The evidence⁷³ we have suggests that teacher knowledge of UMS mark requirements had a particularly strong effect on the marking of GCSE controlled assessment.

6.76 Schools told us they would target a UMS score for each student that would get them “safely within the score range for their predicted grade”⁷⁴. When the boundaries were set at higher marks, these students did not achieve their predicted grades.

6.77 The charts below include data for students who sat the written paper early (before June 2012) and were sitting one or more controlled assessment units in June 2012. Students who followed that entry sequence represent nearly half of the entry for each exam board.

6.78 The charts show the difference between the UMS actually achieved by candidates and those that were required for a grade C.

6.79 There are four for each exam board – two for English Language and two for English. In each pair, the first chart is based on the January 2012 controlled assessment boundaries; the second is based on the controlled assessment

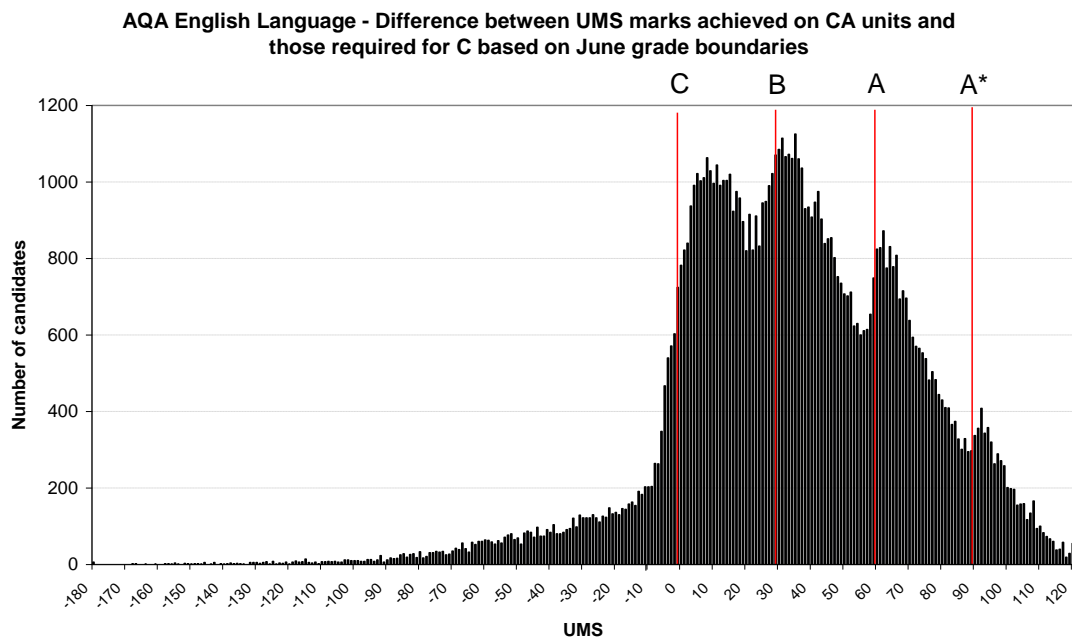
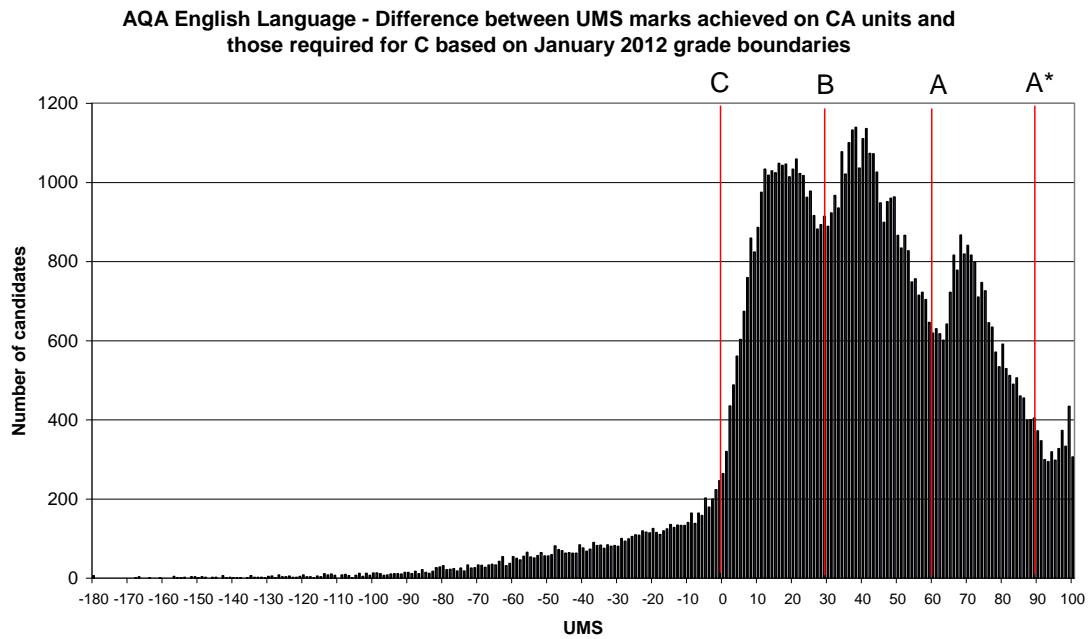
⁷³ AQA Report, October 2012 (Annex A, Figure 9).

⁷⁴ Interviews with schools conducted in September and October 2012.

boundaries that were set in June 2012. (We have not analysed WJEC entry data in this way because the first opportunity available to enter the controlled assessment units was in June 2012. Therefore there were no grade boundaries from earlier series to influence schools' marking.)

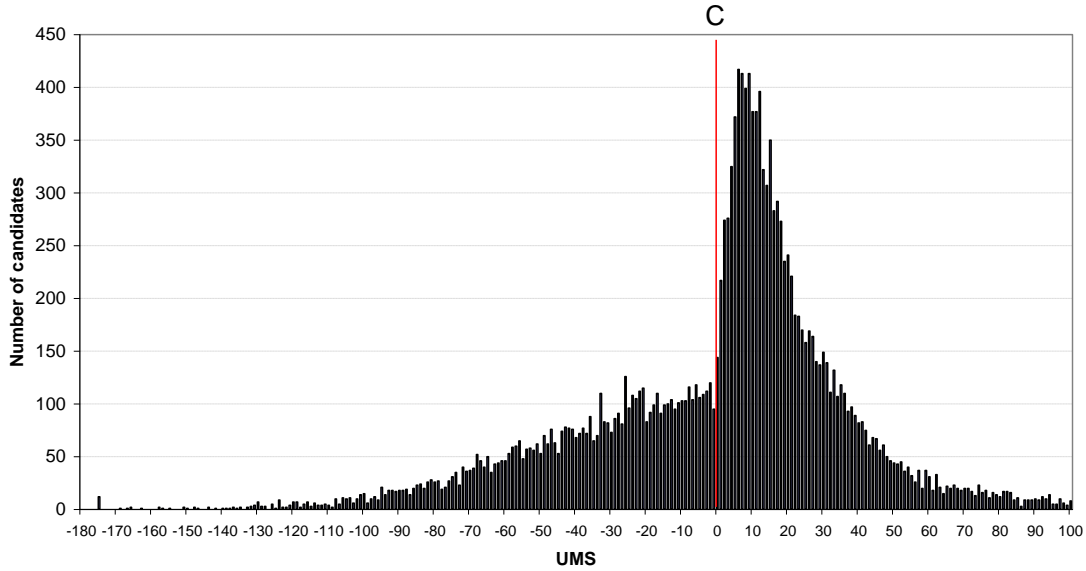
AQA

6.80 AQA has 62 per cent of the entry for English/English Language. The following charts show the analysis for English Language, which has the larger entry. The charts show multiple spikes, reflecting the impact of teacher knowledge of the UMS marks required by individual students to meet target grades. In the first chart, the “peaks” are comfortably into the boundary. In the second chart, which is based on the actual June boundaries, many more students just miss a boundary.

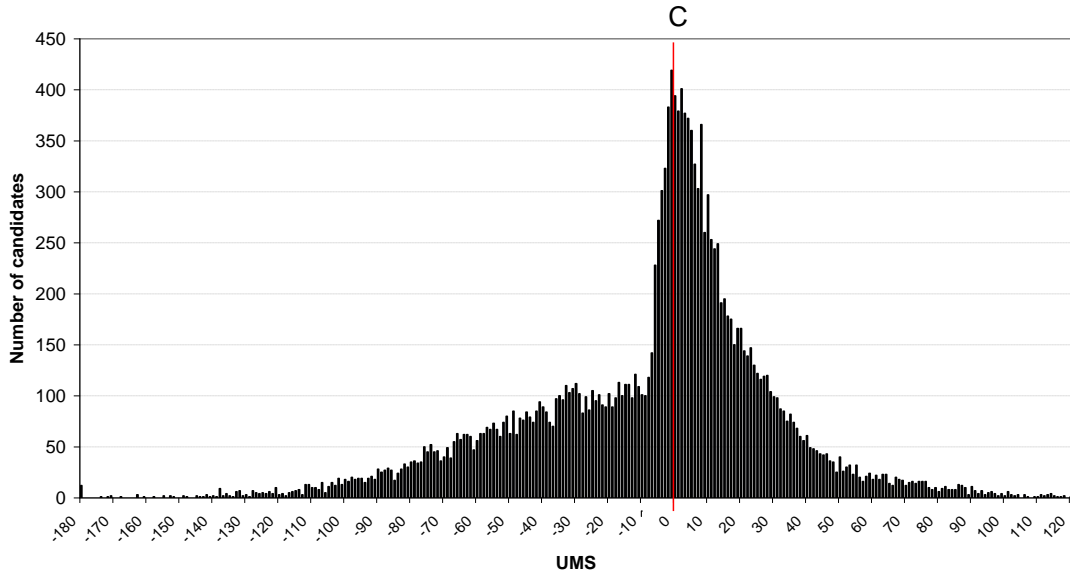


6.81 There is a similar pattern in English. Due to the different entry for English (tending to be lower ability students) the peak is focused on the grade C boundary only.

AQA English - Difference between UMS marks achieved on CA units and those required for C based on January 2012 grade boundaries



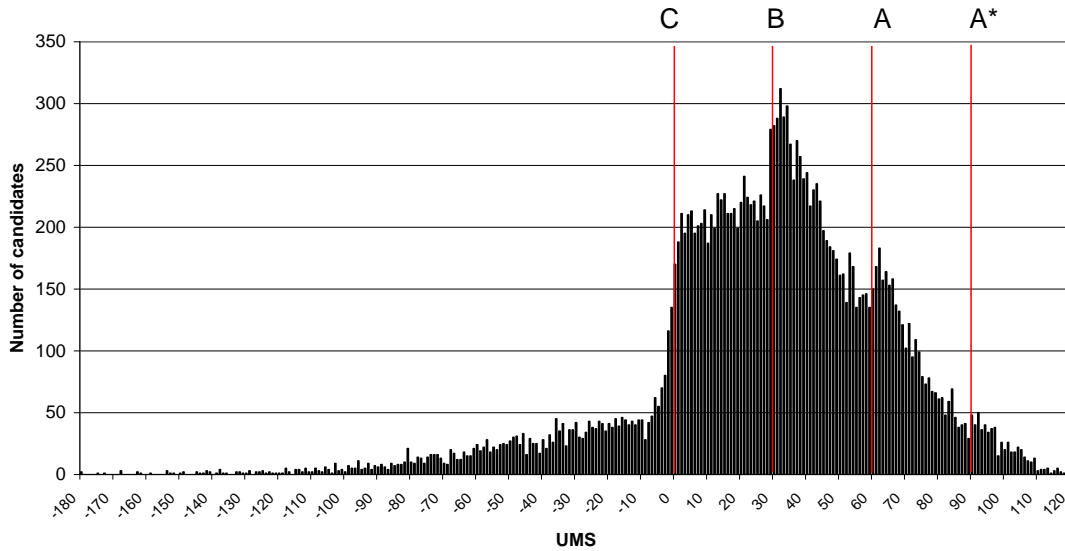
AQA English - Difference between UMS marks achieved on CA units and those required for C based on June grade boundaries



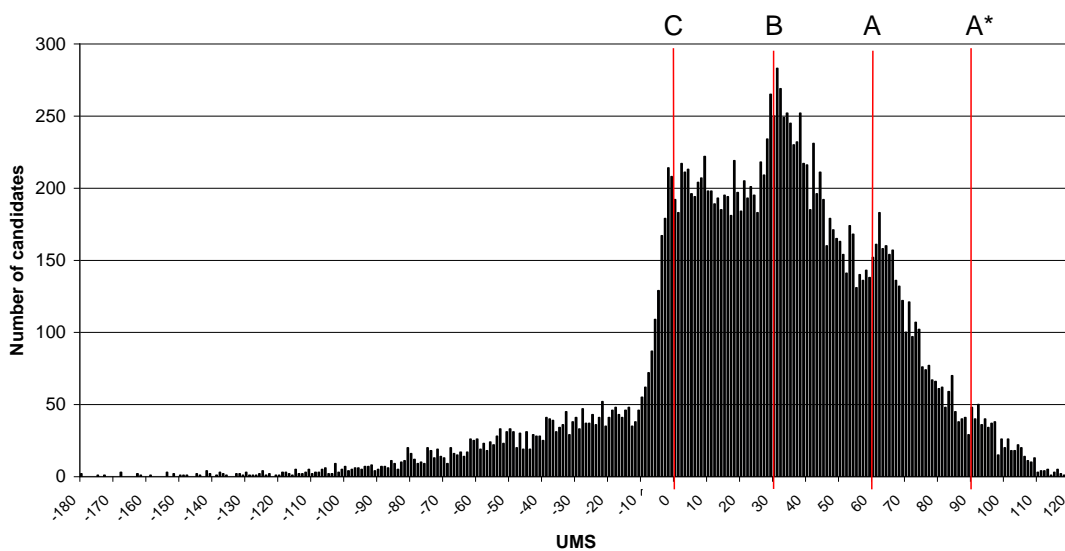
Edexcel

6.82 Edexcel has 11 per cent of the market. The pattern for Edexcel is not so clear-cut as for AQA, because it has a smaller entry and so the numbers of candidates in the histograms are much smaller. But there are still clear spikes at C and B in English Language and at C in English.

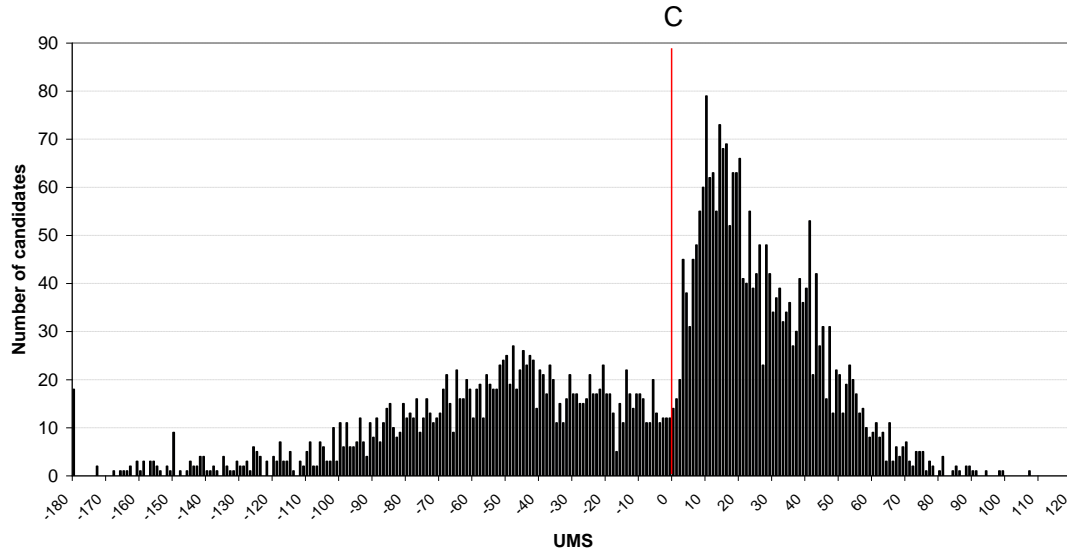
Edexcel English Language - Difference between UMS marks achieved on unit 1 (controlled assessment) and unit 2 (written paper) and those required for C based on January 2012 grade boundaries



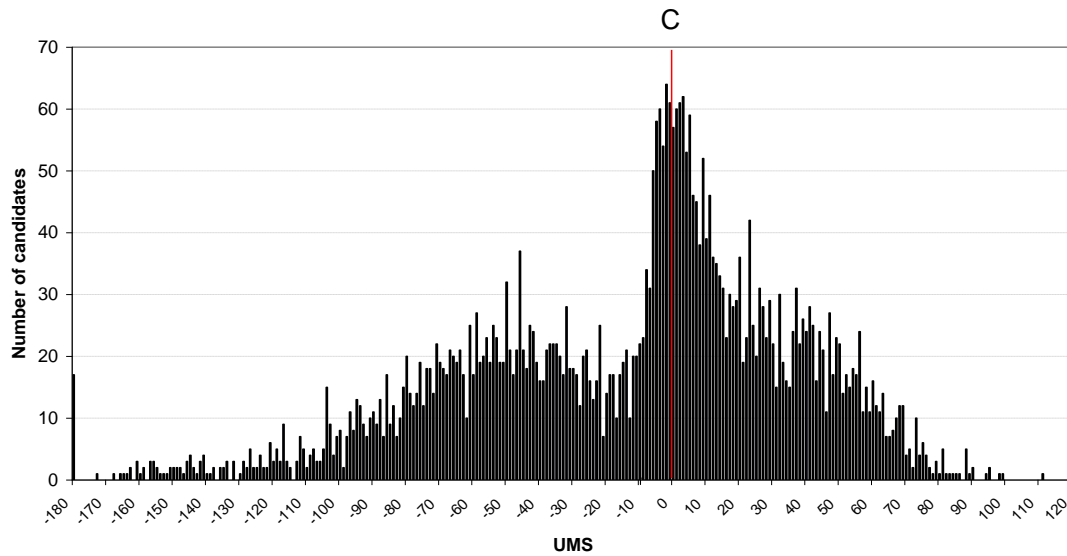
Edexcel English Language - Difference between UMS marks achieved on unit 1 (controlled assessment) and unit 2 (written paper) and those required for C based on June grade boundaries



Edexcel English - Difference between UMS marks achieved on unit 1 (controlled assessment) and unit 2 (written paper) and those required for C based on January 2012 grade boundaries

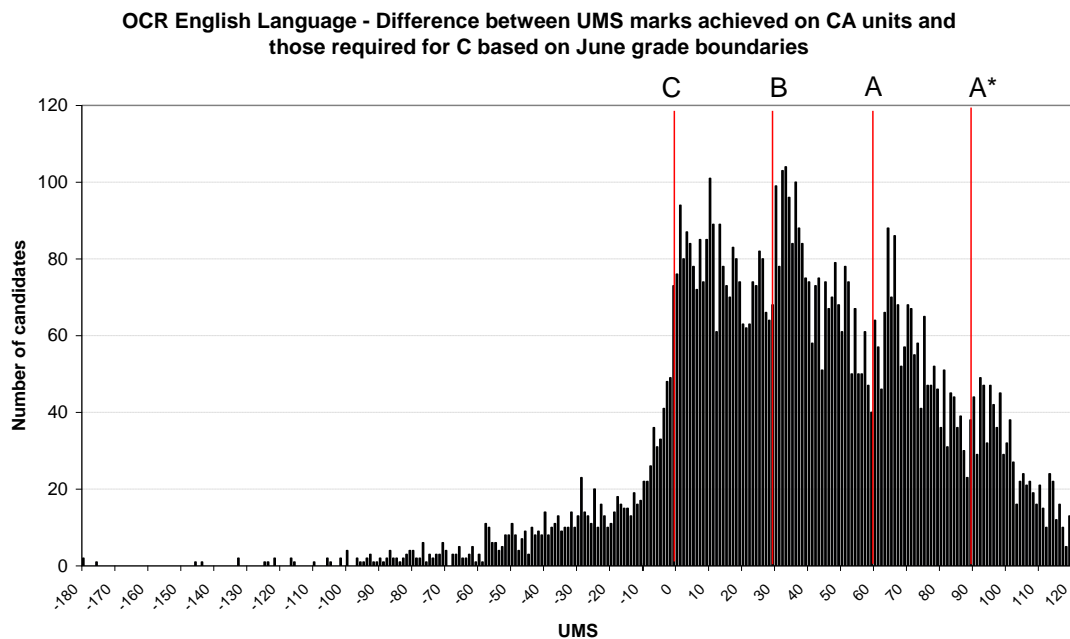
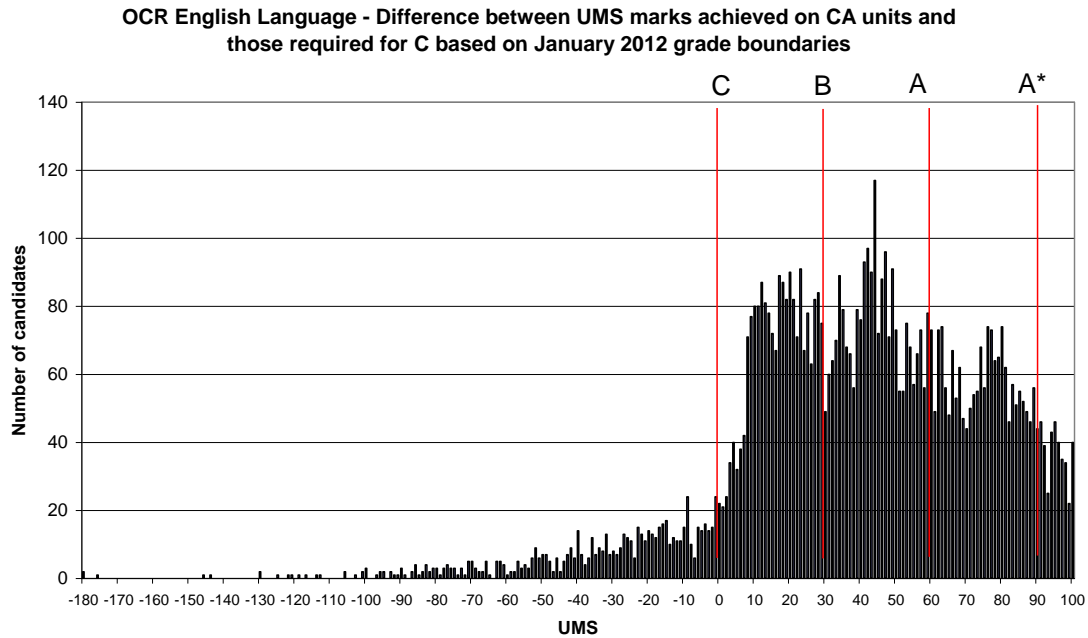


Edexcel English - Difference between UMS marks achieved on unit 1 (controlled assessment) and unit 2 (written paper) and those required for C based on June grade boundaries

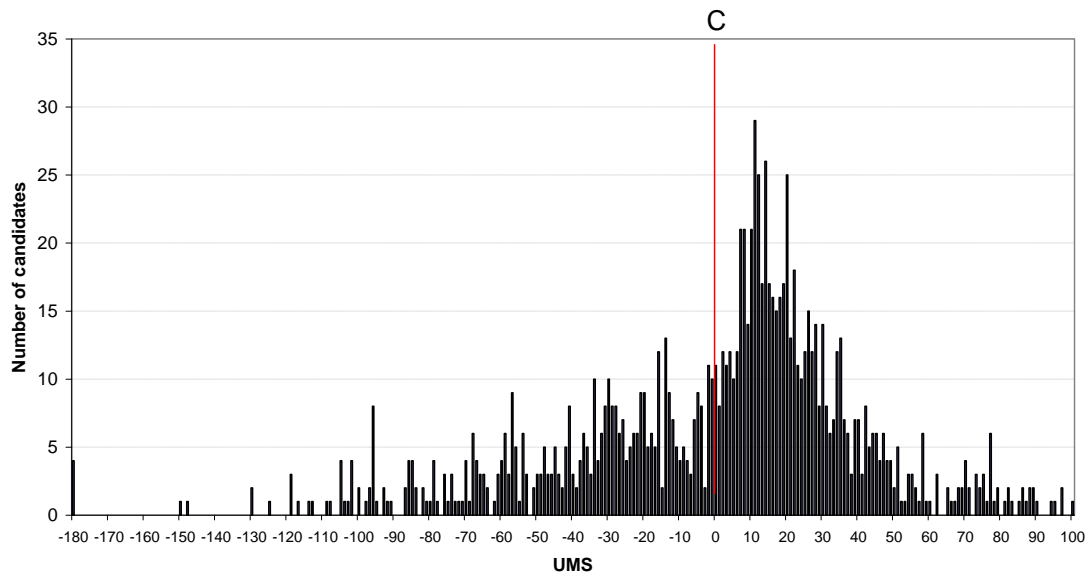


OCR

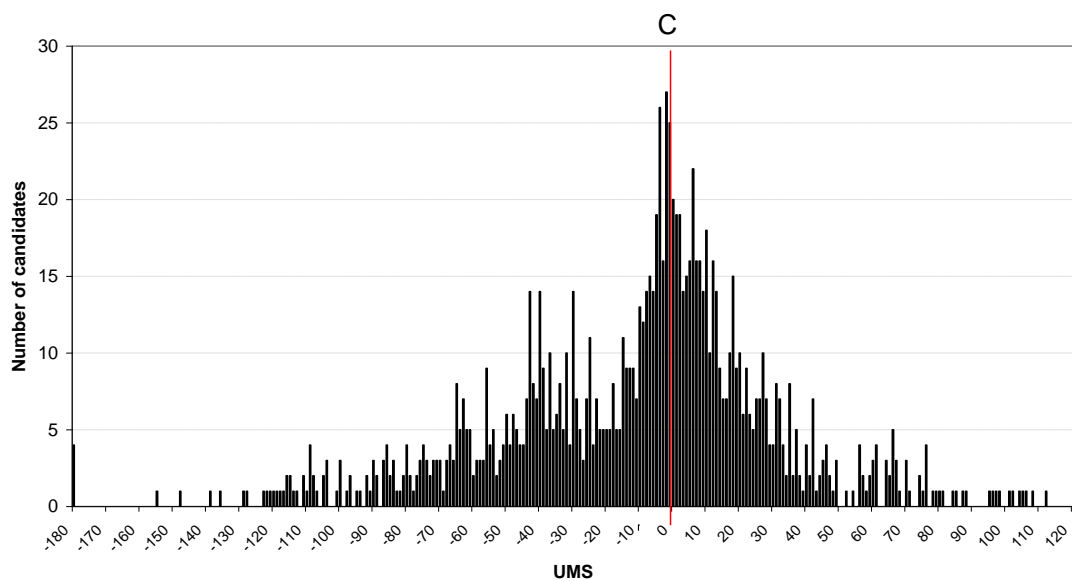
6.83 OCR has 7 per cent of the market. The data for OCR shows a similar, but less clear pattern, possibly due to the relatively small numbers in the analysis.



OCR English - Difference between UMS marks achieved on CA units and those required for C based on January 2012 grade boundaries



OCR English - Difference between UMS marks achieved on CA units and those required for C based on June grade boundaries



6.84 It is clear that wherever the controlled assessment grade boundaries were set in January, the very specific preparation of each candidate for the marks needed combined with widespread over-marking would have meant the boundaries needed to be moved in June. So if in January they had been set where the June boundaries were set, the June boundaries would have been even higher.

7. Regulating GCSE English and English Language

- The combined impact of the changes to GCSE English qualifications – in particular the high proportion of controlled assessment and the rule enabling students to submit controlled assessment to complete the qualification – makes them susceptible to pressures from the accountability system.
- Proposals to redress these weaknesses were rejected as they would undermine modularisation and student choice.
- Regulators gave extra attention to how exam boards were managing the particular risks of awarding, but did not actively review how the qualifications played out in schools.
- Moderation of controlled assessment could have worked better in schools, and communications between exam boards and schools could have been much clearer.
- Regulators allowed the common GCSE controlled assessment marking moderation tolerance (6 per cent) to apply. That enabled widespread over-marking, albeit within tolerance.
- Exam board and school communications were clear about the risks of reliance on past grade boundaries but reliance was placed on them anyway.

7.1 In this section we look at how regulators regulated the qualifications. The regulatory arrangements changed in 2010, and we deal with that first of all.

The regulators

7.2 Ofqual regulates qualifications offered in England. The Welsh Government regulates those offered in Wales, and CCEA regulates those offered in Northern Ireland.

7.3 Before 1st April 2010 QCA was responsible for regulating qualifications in England, alongside its other, wider responsibilities. In 2007, Ministers decided to separate out the regulatory functions of QCA and create an independent regulator for qualification standards. Legislation was needed in order to implement the decision.

7.4 In 2008 and in preparation for the change, what was known as an “interim regulator” was established within QCA. The interim regulator did not have statutory powers, and was not established as an independent body. Rather, the Secretary of State authorised a QCA committee, to be chaired by a member of the QCA Board, to exercise certain regulatory functions. The arrangements

required the committee to operate within the overall strategic direction of QCA and to report to the Board of QCA.

- 7.5 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. They allowed the interim regulator to use the name “Ofqual” pending creation of Ofqual itself.
- 7.6 In 2009 exam boards developed new English GCSEs in accordance with QCA rules published in 2008. The new qualifications were accredited by the interim regulator in December 2009.
- 7.7 In 2009 the legislation to create a non-ministerial department as an independent regulator for standards was passed⁷⁵, and in 2010 Ofqual came into being. We are now responsible for developing the rules for the qualifications and for approving individual qualifications such as GCSEs developed by each exam board.

Qualifications development

- 7.8 QCA no longer exists. Given the problems experienced with the new qualifications, we have reviewed the archive records available to see what was considered when the changes were mooted and agreed.

Modularisation

- 7.9 We can find no record to show that the particular difficulties of maintaining standards and awarding the qualifications were considered. Discussions were focused on the curriculum, underpinned by a belief in a common, unitised approach across all qualifications at similar levels. However, QCA recognised the incentives created by the accountability system⁷⁶, and saw that modularisation might not be trouble-free:

“The potential downside of unitisation is that assessment becomes fragmented, atomised and that students will be encouraged to re-sit

⁷⁵ The Apprenticeship, Skills, Children and Learning Act 2009.

⁷⁶ At the time the changes were being considered, the main performance measure for schools already included GCSE English as an essential component.

modules in order to gain the desired Grade C, especially for schools wanting to push their league table positions to the maximum⁷⁷.”

To counteract such tendencies QCA imposed what is known as “the terminal rule”: a certain percentage of any GCSE has to be sat or submitted at the end of the course of study. QCA considered that this would “retain the original aim, in ensuring that assessment does not become fragmented, while enabling schools to deliver applied qualifications in a sufficiently flexible way to meet the needs of all users⁷⁸.”

7.10 QCA decided after consultation on its original proposal – that generally, 50 per cent of the final assessment for GCSEs would have to be done by exam at the end of the course⁷⁹ – that 40 per cent would be done by written exam or controlled assessment. Consultees had expressed concerns about the effect of a 50 per cent rule in practical subjects such as Design and Technology.

7.11 QCA took the view that this would “retain the original aim, in ensuring that assessment does not become fragmented, while enabling schools to deliver applied qualifications in a sufficiently flexible way to meet the needs of all users⁸⁰.” This enabled modularisation in practice, to the fullest extent.

Controlled assessment

7.12 QCA’s 2005 review of coursework led to the introduction of controlled assessment. The arrangements for controlled assessment limit the opportunities for individual student malpractice, and they have been broadly welcomed by schools.

7.13 The new arrangements do not prevent school malpractice. It is questionable whether any arrangement could, but we cannot see that known concerns about it were aired at the time.

⁷⁷ QCA Executive meeting decision paper, 4th June 2007.

⁷⁸ QCA Executive meeting decision paper, 26th September 2007.

⁷⁹ QCA Executive meeting, decision paper, 4th June 2007.

⁸⁰ QCA Executive meeting decision paper, “Headlines from the GCSE Consultation and further work on Controlled Assessment in GCSE Qualifications”.

A more consistent approach to assessment

- 7.14 In late 2007 QCA decided to apply a more consistent approach across subjects and decided that the balance of assessment in any subject should be struck in any one of three, but only three, ways: no controlled assessment, 25 per cent controlled assessment, or 60 per cent controlled assessment.
- 7.15 Different methods of assessment are more or less valid, depending on what is to be assessed. The most effective proportion of controlled assessment is likely to be greater in more practical subjects, such as Design and Technology, as opposed to more academic subjects.
- 7.16 GCSE English Language had had 40 per cent coursework, and English Literature 25 per cent. 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: English and English Language would contain 60 per cent controlled assessment, while English Literature stayed as it had been, at 25 per cent.

Assessing risks

- 7.17 We have sought to identify whether the impacts and the risks of these changes were identified and managed.
- 7.18 At the time the changes for English were being considered, QCA conducted an impact assessment. We have not been able to locate the completed impact assessment from QCA archives, but we have found reference to it in a QCA Executive meeting paper⁸¹. The paper refers to potential disability implications, and also mentions what is described as “an important risk” in respect of the introduction of controlled assessment in place of coursework. The noted concerns at that stage were about a possible resistance to controlled assessment and a possible lack of understanding of the change.
- 7.19 The English teachers we have spoken to have told us that they understand controlled assessment well, and that generally they like it. In our review of the implementation of controlled assessment last year⁸² we found that the principles of controlled assessment are well received and that teachers were broadly supportive of it. Two-thirds of English teachers appeared to find it easy to implement. We conclude that QCA’s concerns about resistance to controlled

⁸¹ QCA Executive meeting decision paper, 5th November 2008: “GCSE 2010: proposed changes to English subjects, information and communications technology (ICT) and mathematics criteria”.

⁸² Ofqual, *Evaluation of the introduction of controlled assessment*, October 2011.

assessment or a lack of understanding of it have not been substantially borne out in practice.

- 7.20 In practice, the combined impact of the changes to GCSE English qualifications – in particular the high proportion of controlled assessment – makes these exceptionally high-stake qualifications susceptible to pressures from the accountability system. The terminal rule weakens the qualifications still further. We have found evidence to show the risks to standards were considered at the time the changes were mooted and agreed.
- 7.21 In an October 2009 paper⁸³, shortly after new modular GCSEs in most subjects had begun to be taught in schools, the interim regulator discussed awarding in the context of all modular qualifications. At the same time, technical staff in exam boards discussed with staff in the three regulators the same challenges. Options were identified for making awarding more robust, but, as they involved limiting when a student could sit a unit or else receive a mark and grade for it, they were rejected on the basis that they either undermined the philosophy of modularisation, or undermined public confidence. The interim regulator was obliged to operate within the overall strategic direction of QCA⁸⁴ and to report to the Board of QCA. No actions were taken.
- 7.22 Ofqual commenced operation in April 2010. By then, all GCSEs except those in English, ICT and mathematics were already being taught in schools. We did not consider reviewing the propositions made in 2009. Modularisation was already in full swing. Instead, in 2010 we identified particular high-stake GCSE qualifications for attention: GCSEs in sciences, mathematics and English. We decided to instigate exceptional monitoring arrangements for GCSE English qualifications. As teaching for the qualifications began, we commenced what is known as a scrutiny study. The Welsh and Northern Ireland regulators oversaw these matters in Wales and Northern Ireland for the exam boards WJEC and CCEA respectively.
- 7.23 In scrutinising GCSE English, we aimed to make sure that key processes operated properly and in particular that exam boards acted properly and reasonably when awarding the qualifications in England. We adopted a risk-based approach, looking particularly at AQA GCSE English and Edexcel GCSE English Language. We observed others as well (as shown below) to a lesser extent. AQA GCSE English includes Unit 1 (ENG1F), a unit where concerns

⁸³ Ofqual *The “ping” factor: what does it mean to maintain standards in composite qualifications?*, October 2009.

⁸⁴ Memorandum of Understanding between Ofqual and QCA.

have been expressed about the differing grade boundaries in January and June 2012. We attended selected exam board training, standardising and awarding activities in England as an observer, as set out in the tables below. In each instance our observers were collecting evidence of compliance with expected processes to secure standards. They will report on this at the end of the series, as is usual, to inform the following awarding season.

AQA GCSE English (4700)

Activity	Date(s)
Pre-standardisation: Unit 1 (ENG1F/H)	13th January 2011
Awarding: Units 1–3, winter 2011 series	21st–22nd February 2011
Centre Assessor (teacher) Guidance and Training	12th May 2011
Standardisation: Unit 1 (ENG1F/H)	13th–14th June 2011
Awarding: Units 1–3, summer 2011 series	27th–28th July 2011
Review of assessments and candidates' work	August–October 2011
Advisory centre visit: ENG02	18th April 2012
Advisory centre visit: ENG02	20th April 2012
Moderator pre-standardisation: ENG03	2nd May 2012
Moderator standardisation: ENG03	4th–8th May 2012
Question Paper Evaluation Committee: ENG1F/H 2013 papers	21st May 2012
Examiner pre-standardisation: ENG01F/H	2nd–3rd June 2012
Examiner standardisation: ENG01	6th–7th June 2012
Awarding summer 2012 series	20th–21st June 2012

Edexcel GCSE English Language (2EN01)

Activity	Date(s)
Pre-standardisation and standardisation: Unit 2 (5EN2F/H)	24th–25th June 2011
Pre-standardisation and standardisation: Unit 3 (5EN03)	30th–31st May 2011
Awarding: Units 1–3	18th–19th July 2011
Review of assessments and candidates' work	August–October 2011
Moderator pre-standardisation and standardisation: English Language Unit 1	15th May 2012
Moderator pre-standardisation and standardisation: English Language Unit 3	16th–17th May 2012
Examiner pre-standardisation and standardisation: English Language Unit 2	7th–9th June 2012
Awarding	9th–12th July 2012

OCR GCSE English Literature (J360)

Activity	Date(s)
Moderator standardisation: Unit 1 (A661)	15th December 2010
Standardisation: Unit 3 (A663)	15th January 2011
Standardisation: Units 2 and 4 (A662 & A664)	19th January 2011
Awarding meeting: Units 1–4	16th February 2011
Moderator standardisation: Unit 1 (A661)	5th May 2011
Standardisation: Units 2 and 3 (A662 and A663)	15th June 2011
Standardisation: Unit 4 (A664)	21st–22nd June 2011
Pre-award	19th July 2011
Awarding meeting: Units 1–4	19th–20th July 2011
Review of assessments and candidates' work	August–October 2011

AQA GCSE English Literature (4710)

Activity	Date(s)
Awarding	24th–25th February 2011
Moderator standardisation: 47103/5	4th May 2011
Awarding	29th–30th July 2011

AQA GCSE English Language (4705)

Activity	Date(s)
Awarding	24th–25th February 2011
Moderator standardisation: 47103/5	4th May 2011

Edexcel GCSE English (2EH01)

Activity	Date(s)
Moderator standardisation: 5EH03	21st May 2011
Awarding	20th July 2011

OCR GCSE English (J350)

Activity	Date(s)
Awarding	7th July 2011

OCR GCSE English Language (J355)

Activity	Date(s)
Awarding	19th February 2011
Standardisation: A680	19th–20th June 2011
Moderator standardisation: A651	6th May 2011
Awarding	7th July 2011

7.24 In 2011 we noted good practice in examiner preparation (AQA and Edexcel); good practice on task-setting in controlled assessment (AQA) and good moderator standardisation (Edexcel). Data supplied to us at the time evidenced successful teacher training on controlled assessment marking (AQA).

7.25 In observing awarding in 2011 we observed a genuine emphasis on standards over time (AQA ENG02) and noted with approval the establishment of a single awarding committee for both English and English Language, with crossover membership with the English literature committee (AQA). We noted due attention paid to standards over time and an emphasis on the vital importance of examiners' professional judgement (at C and E grades) in establishing standards for a new specification (Edexcel).

7.26 We observed an unusual proportion of new markers in 2011 (AQA).

7.27 From AQA's review of candidate scripts we observed that some foundation tier candidates had not completed the longer task on the paper and there were some brief answers and unfinished papers. Some schools had written to AQA to say that some candidates found the length of the paper (two hours 15 minutes) problematic.

- 7.28 Examiners noted the high instance of early entry – 15-year-olds rather than 16-year-olds (Edexcel). Examiners were reminded to judge the work on standards alone without taking into account age (Edexcel).
- 7.29 There were some signs of problems with controlled assessment, for example instances of schools appearing to use a “frame” of assessment, where students produce work in the same format (AQA). However, examiners also felt that schools had generally responded very well to the changes (Edexcel).
- 7.30 **In January 2012** we observed that steps AQA had put in place to develop their new markers were proving effective.
- 7.31 Examiners commented that the writing tasks on AQA ENG1F had worked well and had produced some interesting and effective writing. Examiners observed that when there were few candidates (for example in AQA English 03) awarding was difficult and that the tendency to enter early was still a complicating factor.
- 7.32 Examiners noted that, with school management teams placing ever-increasing pressure on departments to improve exam results (especially at grade C) schools were correspondingly putting pressure on exam boards to declare the correlation between marks and grades, and were tending to inflate marks so that candidates were placed within certain perceived grade boundaries (Edexcel).
- 7.33 **In June 2012** we observed that the AQA and Edexcel awards were based on an appropriate range of different types of evidence, including candidates’ work and statistical information. There was also an appropriate emphasis on aligning grading standards across different exam boards and over time.
- 7.34 Overall, the awards suggested that candidates’ performance was consistent – or improved – from previous exam series, which may have been the result of schools’ increasing familiarity with the requirements. However, we did also find evidence of some issues. Examiners noted examples of infringements in relation to the notes that candidates are permitted to use for written controlled assessment tasks.

Awarding at qualification level

- 7.35 The three regulators oversee awarding of GCSEs for all subjects. We review preliminary results from all exam boards. Most qualification awarding in summer 2012 was completed mainly without intervention by the three regulators, although our exchanges with Edexcel and with WJEC (jointly with the Welsh regulator) led to amendments to their English awards to bring them more in line

with other boards. We can now see⁸⁵ that although our interventions were necessary they may not have been fully effective, as final results of both these exam boards are still somewhat out of line with those of their competitors – but not as much as they would have been had the regulators not intervened.

Regulating controlled assessment

- 7.36 In 2011 we reviewed the implementation of controlled assessment⁸⁶. We were looking at that stage at school understanding of the new rules, and found widespread concerns among teachers about the logistics and the rules for controlled assessment. We required exam boards to be more consistent and to make more clear what is required. Arrangements have improved.
- 7.37 In April 2012 we began a fundamental review of controlled assessments on a subject-by-subject basis. We were concerned about the proportions of controlled assessment in some subjects and concerned about how valid the assessments were. As we said at the time, we want assessment in each subject to be valid, and to promote good learning and teaching. We put out a call for evidence, now closed, and expect to complete our review by the end of 2012.
- 7.38 Examiners rely on the marking of controlled assessments by teachers, with Reading/Writing moderated by exam board moderators. Our expectation, and theirs, is that the moderation process, as set out in the Code of Practice, will identify instances of marking that are overly harsh or over-generous, leading to an adjustment of all the marks from that school accordingly.
- 7.39 Exam boards must comply with the regulators' Code of Practice in the way that they work. The Code specifies the requirements for moderation, with exam boards able to determine the tolerance limits or sample sizes. It has long been exam board practice to set tolerance at 6 per cent, and this was applied to both the old and the new qualifications.
- 7.40 It has been suggested that the tolerance should have been tighter. Moderation is not designed as a second marking exercise but as a check on the overall marking of individual controlled assessment scripts. A tolerance will always be necessary and in English the case for a reasonable tolerance is made by the nature of the subject itself. Given what we know now, however, we think there are compelling standards reasons to tighten the tolerance.

⁸⁵ JCQ screening for GCSE English/English Language summer 2012.

⁸⁶ Ofqual, *Evaluation of the introduction of controlled assessment*, October 2011.

Exam boards

- 7.41 From our scrutiny study and from all the additional evidence we have seen, we are satisfied that exam boards applied themselves properly to awarding, and followed expected process.
- 7.42 So far as we are able to judge, they took proper steps to train controlled assessment moderators and to provide training to schools on the new arrangements. They applied established moderating tolerances, and acted when necessary.
- 7.43 However it is clear from what we have heard 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) or other problems with controlled assessment. Rather, some moderators took a helpful, supportive stance.
- 7.44 Exam boards did not identify that school intervention strategies were unduly reliant on published grade boundaries and that in some cases they led to sequential over-marking. Moderation is not designed to identify sequential over-marking.
- 7.45 Exam board communications with schools should have been sharper. We mention earlier in this report the confusion of some teachers about marks and grades that are now apparent. Despite the training provided by exam boards for teachers it is clear that more, and more universal statements are needed for the future.
- 7.46 Exam boards produce a large amount of material, and this can make it harder for teachers and school leaders to identify key messages. Early series grade boundaries were published quite properly as usual, with the confirmation that they should not be relied on for units still to be sat. We now know that some schools still used the published grade boundaries to inform their predictions. Exam boards should therefore consider how they can further reinforce the message that previous grade boundaries should not be relied on.
- 7.47 Finally, from the interviews with teachers that we have conducted, we know that some teachers received confusing and even misleading communications from moderators on occasions. Messages should have been consistent and firm, given the risks to standards.

8. Ensuring standards

2012 awards

- 8.1 In the light of what we now know (and set out in this report) we have reviewed the decisions we made in August not to direct exam boards to change grade boundaries set in January or June 2012. We have concluded that this was the correct decision, that awarding happened as it should, grade boundaries were set properly, and our use of the comparable outcomes approach allowed us to be confident that the standard was carried forward from the previous version of GCSE English.
- 8.2 We have considered the issues carefully, including the equalities implications.
- 8.3 Adjusting the June grade boundaries to be in line with January boundaries would create a disparity over time and ignore the evidence we have of over-marking – the fact that average marks for similar pieces of work were higher in June than January.
- 8.4 If the June controlled assessment units were to be re-graded on January grade boundaries, the overall results could not be justified by reference to educational achievement, and it would introduce substantial new unfairness. Consider for example that almost one in three of all students sat a particular written exam in January⁸⁷ and submitted controlled assessment in June. If the June controlled assessment boundary was moved to match January for these students, then the proportion receiving a GCSE grade C would be 85 per cent, compared to 64 per cent who actually achieved a grade C in the new qualifications and 65 per cent who achieved grade C in English last year.
- 8.5 It has been suggested that students who sat all modules in June 2012 were penalised when compared to those that took some modules earlier. School choices about when students take units and in what order, and when they are re-taken, may affect results, and this is so for all modular qualifications. It is not possible to know whether any individual candidate would have done better, the same or worse had they entered in January. It is clear, however, that a large proportion of students would be significantly over-graded if June boundaries were changed to match January boundaries.
- 8.6 It is impossible to adjust controlled assessment grades to reflect what students might have got if their schools had ignored or acted differently on information about January grade boundaries: there is no possible way of identifying those

⁸⁷ AQA Unit 1 (ENG1F) Understanding and Producing Non-Fiction Tests.

students, or determining how those students would have performed in different circumstances.

- 8.7 We have considered whether it might be possible to allow schools to appeal their controlled assessment moderation, where they believe the marking was relatively harsh. However, even if the logistical difficulties with this could be resolved, we do not believe it could be done fairly and consistently for all schools and all candidates.
- 8.8 For both controlled assessments and exams, we believe the early re-sit option already offered by exam boards offers the best balance between, on the one hand, giving students who believe they have been affected the chance to improve their grades quickly – and large numbers of students have entered for re-sits – and on the other hand, introducing significant new unfairness into the results.

Future years' awards

- 8.9 In the light of what we now know, we have considered immediate and longer-term actions to protect standards, for discussion with our fellow regulators in Wales and Northern Ireland.

Immediate actions

- 8.10 We want to make sure that students sitting GCSE English in 2013 can have confidence in the results they receive. For this reason we have decided that exam boards should:
- tighten moderation arrangements for all controlled assessments
 - not issue grades, or other information about aggregate performance, for any assessments undertaken in January 2013; and
 - grade January and June 2013 written papers and controlled assessments at the same time.
- 8.11 We are requiring exam boards to implement tighter controls over controlled assessment, with immediate effect. So this will include the November re-sits. This will improve the consistency of marking across schools.
- 8.12 By requiring exam boards to postpone grading January 2013 written exams and controlled assessments until they are ready to grade the written exams and controlled assessments entered in June 2013 we will assure ourselves and others of a consistent approach.
- 8.13 So there will be no English or English Language unit or qualification awards made in January 2013. Students will still be able to take exams, and submit

controlled assessments in January 2013. Exam boards will provide raw marks for exams, and will moderate controlled assessments, but they will not award any grades, and will not allow students to cash in their units to get a qualification.

8.14 We will also require exam boards to improve and sharpen communications with schools.

8.15 For the June 2013 awards, we will require exam boards to identify and manage specific risks to standards in these qualifications, and we will monitor closely how they are managing those risks.

Longer term

8.16 We have already announced that from September 2013, GCSEs will no longer be modular in England. To protect standards we will consider any other design changes needed. In doing so we will act on the findings of our review of controlled assessment, already well underway, including considering with CCEA any changes to the proportion of controlled assessment in GCSE English.

8.17 We will review with the regulators in Wales and Northern Ireland how standards will be secured in each country over the coming years, given experience this year. We note here the concerns raised by the Education Select Committee in its report in July⁸⁸.

8.18 We will advise the Secretary of State for Education on the implications of our findings for the recently announced review of school accountability. We will also advise him on the implications of our findings in the context of his proposals for English Baccalaureate Certificates and other new qualifications.

8.19 Finally, we will review our own approach to regulation, to make sure that in regulating and planning qualifications reforms we take into account fully the context in which qualifications operate.

⁸⁸ Education Select Committee (July 2012) *The administration of examinations for 15–19 year olds in England*, paragraph 107. See: www.publications.parliament.uk/pa/cm201213/cmselect/cmeduc/141/141.pdf (accessed on 31st October 2012).

Appendix 1 - GCSE English Awards 2012: Report of Findings from Centre Interviews, September – October 2012

Produced for Ofqual by Capgemini

1. Executive Summary

Capgemini were commissioned by Ofqual to investigate 'why some centres' results differed significantly from their predictions'. Capgemini interviewed 100 schools and colleges across England during September and October 2012, and our findings are summarised below.

Grade boundaries varied between January 2012 and June 2012

English teachers relied on January 2012 grade boundaries to inform predictions for their students' final grades in June 2012, and did not think the boundaries would vary significantly between January and June. The extent of the variation in grade boundaries between January and June contributed to centres' sense of surprise when their results were not as predicted.

Considerable effort went into making accurate predictions

GCSE English is a key accountability measure for centres in England. English teachers are under pressure to predict the grades that their students will achieve. Our interviews highlighted that accurate prediction of grades is seen as a key part of the English teacher's role.

The emphasis on accurate predictions and the historical accuracy of those predictions meant that teachers continued to be confident in their grade predictions in 2012.

In many schools this prediction process is supported by data analysis, which uses Key Stage 2 results, Fischer Family Trust data and mid-year reports. In many instances predictions are updated on a six-weekly basis during Years 10 and 11 to reflect progress being made. Where students deviate from the 'target outcome expected of them', interventions are made to ensure that they achieve to the best of their ability. The fact that teaching is so targeted to secure the predicted grade has added to the 'surprise' that was felt when the grades were not as predicted.

Many teachers we spoke to also considered that the grade given by the Awarding Organisation is a confirmation of what they already know about the student's ability. The results for some centres in 2012 were unexpected because they differed from teachers' own predictions. Teachers felt that these results did not reflect the abilities of some of their students.

The historical stability of the qualifications strengthened confidence

Prior to the 2012 results, English GCSE qualifications had been characterised by stable, predictable results. Centres told us that their results tended not to vary much in relation to their predictions, or from one year to the next. This strengthened

teachers' confidence in their ability to make accurate forecasts. The variation in results this year meant that teachers' predictions were not realised, and this was unexpected.

Impact of controlled assessments

Controlled assessment made up 60% of the English and English Language qualifications, whereas the previous GCSE English specification had 40% coursework. Teachers had marked students' controlled assessments and knew the number of marks they had given each student for this element of the qualification. They used information about previous grade boundaries to judge the likely grades those marks would achieve. They therefore had a high degree of confidence in their anticipated grades for at least 60% of the qualification and it was unexpected when those grades did not materialise.

Impact of moderation

Most centres that we interviewed did not have their marks for the controlled assessment changed through moderation. The majority of centres took this to mean that their marking had been accurate. On the whole, teachers had received letters from the Awarding Organisations saying that their marks had been accepted without adjustment. Teachers interpreted this message as confirmation of their students' grades, based on their understanding of the previous grade boundaries. Therefore it was unexpected when their predictions were not met.

In a number of cases, teachers were passionate that the moderation process did not work and some were sceptical whether it had been carried out at all. They thought this because they believed, following conversations with teachers at other centres, that they were 'harsh markers' and had thought their marks would be moderated up.

The confirmation of controlled assessment marks also led to some centres questioning the quality of marking for the written exam, where a student's UMS score between the two was significantly different. This exacerbated the concerns that centres had about their grades.

There was a lack of clarity about the relationship between marks, bands and grades

Centres believed that there was a lack of clarity from awarding organisations regarding marks, bands and grades. Specifically, they believed that advice given by awarding organisations about grade boundaries was imprecise, and this led to misplaced confidence about future grade boundaries among centres. Centres thought that the move to marking 'bands' in the new specification for the controlled

assessment units and the unclear way in which those bands aligned to grades further complicated the situation.

Teachers felt they were prepared for the new specification

Teachers we interviewed had invested time and effort in understanding the changes to the GCSE English qualifications and preparing themselves for these changes. The Awarding Organisations had run conferences, which teachers attended; teachers had reviewed the new specifications, taken advice from their local authority and from consultant experts (many of whom work or had worked for Awarding Organisations) and therefore felt prepared. They were aware that there would be some challenge as the qualifications were new, but they felt well prepared.

Having seen the actual examinations, with the exception of one written paper containing elements that a small number of centres did not expect (a radio script question), the majority of centres were confident that their preparation had been successful. It was therefore unexpected when the results were not as predicted because they thought they had done as much as was possible to be ready.

GCSE English Literature results were as expected by centres

It is clear that the surprise at the results was also driven by the fact that centres' GCSE English Literature results were much closer to their own predictions than the results for GCSE English and English Language. They believed this highlighted the issue as it was typical that many of the affected students achieved their predicted grade in Literature but achieved at least a grade lower than predicted in English Language. Teachers were unable to reconcile this difference, and this contributed to their surprise.

Teachers feel that the 'route' they took through the qualification had no impact in 2012, except for the way in which the grade boundaries were applied

The GCSE English qualifications became modular in 2010, and 2012 was the first year that students could finish the course in this way. The modular nature of these qualifications means that students can submit controlled assessments and take examinations throughout the two-year course. Teachers did not believe this change directly led to the unexpected results. Centres that entered students into earlier exams were frequently as surprised as those that entered their students at the end of the qualification. However, centres felt that the way in which the grade boundaries varied between January and June 2012 did have an impact on their results.

2. Introduction

GCSE results were announced on 23 August 2012. Following concerns raised by schools and colleges into outcomes in the GCSE English subject qualifications, Capgemini on behalf of Ofqual interviewed 100 centres to identify why some centres' results differed significantly from their predictions.

The study focused on qualitative insights from a selection of centres and teachers across England. It was not a statistical study. We interviewed Heads of English and English teachers, with some headteachers also being interviewed. Interviews were mainly face to face, although telephone interviews were also conducted.

To secure the interviews we contacted more than 300 centres, using a sample of:

- centres randomly selected to provide a representative sample of centres with GCSE English results in 2012 that were higher, lower or very similar to their results in 2011 (69 of the interviewed centres came through this route)
- centres that had submitted a written complaint to Ofqual about their GCSE English results (19 of the interviewed centres came through this route)
- centres nominated by representative bodies, such as the Association of Colleges (AoC) and Association of School and College Leaders (ASCL), following a request to these bodies by Ofqual (12 of the interviewed centres came through this route).

We experienced a higher response rate from centres with lower pass rates in 2012, so the majority of interviews were with centres with lower English GCSE A*–C pass rates in 2012 than in 2011.

Of the 100 centres we interviewed, the distribution by Awarding Organisation was as follows:

- 65 with AQA
- 18 with WJEC
- 9 with Edexcel
- 8 with OCR

The distribution by centre type was as follows:

- 33 local authority maintained schools
- 47 academies

- 13 further education colleges
- 6 independent schools
- 1 special school

We worked with Ofqual to develop a number of hypotheses for why some centres' results differed significantly from their predictions. These hypotheses were based on findings from Ofqual's initial investigation in August and other discussions. We then tested those hypotheses in interviews over a four-week period. At the end of each week, we reviewed the interview evidence, and then refined and updated the hypotheses.

3. Why did some centres not expect the GCSE English results they received?

3.1 Grade boundaries varied between January 2012 and June 2012

Hypothesis

Centres used the January 2012 grade boundaries as information on which to base their predictions. They assumed that grade boundaries would not change significantly between January and June. The variation in the grade boundaries between January and June 2012 was greater than centres anticipated, meaning that students' final grades varied from their predicted grades.

Findings

The centres that we interviewed did not anticipate that June 2012 grade boundaries would differ significantly from the January 2012 grade boundaries. In one case, a teacher explained how, at the time, the January results appeared to confirm his predictions for June and he felt confident that everything was on track for June 2012.

We found that centres also used earlier June 2011 grade boundary information to support their predictions. The fact that there was minimal variation between the June 2011 and January 2012 grade boundaries led some centres to believe there would be minimal variation between January 2012 and June 2012 grade boundaries as well – thereby strengthening their confidence in their predictions.

Centres we spoke to had applied the grade boundaries from earlier exam series to their predictions in the belief that these boundaries would not change significantly. Some centres felt that an acceptable boundary movement would consist of no more than two marks, whilst others felt that the maximum shift should not be more than five marks. This view was based on the fact that, historically, grade boundaries have tended not to vary significantly. Centres felt that variations in the grade boundaries by

10 marks in an AQA written exam and an Edexcel controlled assessment unit represented a much larger change than they had anticipated.

The variation in the boundaries between January and June did mean that these predictions, based on the January 2012 grade boundary information, were not realised and this was not what teachers thought would happen.

3.2 Considerable effort went into making accurate predictions

Hypothesis One

English is a ‘high stakes’ subject included in a range of government accountability targets. Because of this, teachers are required to predict students’ final grades with a high degree of accuracy in order to check that the centre is on course to meet its targets. Over the years, teachers have become increasingly confident in their predictions and expect students’ final grades to match their predictions.

Findings for Hypothesis One

The centres we interviewed confirmed that they are under pressure to provide accurate predictions for the English subjects. This pressure stems from the importance of English qualifications to students’ futures and to centres’ own accountability measures. Pressure comes from Ofsted, from centres’ leadership teams and from parents.

Schools told us that Ofsted requires them to predict students’ grade outcomes, and to articulate the steps required to improve students’ grades. As one teacher remarked, “I am confident in my ability to predict results and [this year’s results] were approximately 15% out from what we expected. There is pressure to accurately predict early in the year where the results may come out, and therefore this is an exercise that I take seriously and systematically.” Another centre described the “need to work to grades” and explained that “the pressure on [English teachers] is enormous and comes from pupils, parents, headteachers and the accountability system. [We] have no choice but to use [our] judgement to translate grades from marks.”

In response to this need for accuracy, centres have developed highly sophisticated grade prediction and tracking mechanisms, and prior to 2012 these have proved to be highly accurate in predicting grades for GCSE English. These mechanisms use different sources, including Fischer Family Trust data, Key Stage 2 and 3 results, mid-year reports on student progress and historical grade boundary information. Centres update these predictions on a regular basis. In one instance, a teacher

described how marks from students' class work and mock examinations were fed into their tracking system every six weeks, and that the senior management team met every two weeks to discuss this topic. So the thorough way in which centres approach predictions, the reliability of these predictions prior to 2012, and the perceived stability in grade boundaries year-on-year ultimately strengthened centres' confidence in the results they were predicting.

A common phrase among teachers we interviewed was that their predictions were 'spot on'. Teachers' strong track record in predicting students' GCSE English grades heightened the surprise felt when, in many cases, the grades that students achieved in June 2012 fell short of their predicted grades.

Amongst those centres we interviewed, the pressure to make accurate grade predictions (and the historical reliability of these predictions) *did* contribute to centres' surprise about their GCSE English results in June 2012.

Hypothesis Two

Centres are under pressure to achieve five A*–C GCSE grades (including English and Maths) for as many of their students as possible. Some centres focus on getting a student to a 'safe C' grade and prioritise their teaching accordingly. Teachers base their prediction of what constitutes a 'safe C' on their knowledge of historical grade boundaries. The variation in grade boundaries between January and June 2012 meant that some students who were predicted to be a 'safe C' for GCSE English achieved a lower grade.

Findings for Hypothesis Two

Interviews confirmed that the emphasis on achieving five A*–C GCSE passes means that some centres focus on getting borderline students to a 'safe C' level. Teachers often focus on where they anticipate a C grade to be (which may not always be reflected in the way boundaries are ultimately set – see 3.3 and 3.5).

Teachers told us that their understanding of a 'safe C' is based upon knowledge of historical grade boundaries (as published by the awarding organisations), combined with their professional judgement of the standard of work required to achieve a C grade.

Some centres described how they then prioritise other subjects over English once a C grade in GCSE English is thought to be 'safe'. One teacher explained that their Senior Management wants to "secure" the English grade and then "take the student on to another subject, such as maths." When the grade boundaries in June 2012 were higher than the grade boundaries in January 2012 and June 2011, some students who were predicted to achieve 'safe C' grades, actually achieved D grades.

The pressure on centres to achieve five A*–C grades **did** mean that teachers focused on obtaining ‘safe C’ grades in GCSE English, based to a large extent on earlier published grade boundaries. This **did** lead to unexpected results.

3.3 The historical stability of the qualification results strengthened confidence

Hypothesis

Historically, GCSE English qualification results have been stable: results for individual centres have varied very little from one year to the next. Teachers have grown accustomed to this stability in the English subject awards. The greater variation experienced by centres this year under the new GCSE English specifications was not expected by some centres.

Findings

Centres we spoke to expected their GCSE English results to be relatively stable from year to year, and they also expected their predictions to be highly accurate. Teachers felt that the turbulence in 2012 had undermined their predictions and contributed to the unexpected results.

The historical stability of results for GCSE English meant centres felt increasingly confident about their predictions. Centres acknowledged that the variance in 2012 was unexpected, with one teacher noting that this was “beyond what might be expected as normal volatility”. Some teachers explained how they allowed some latitude in their predictions. However, the increased volatility in the 2012 results meant that, for a number of students, this latitude was not sufficient, and centres’ predictions were not realised in the same way as in previous years.

Teachers told us that the unexpected results in June 2012 mean they are no longer confident in their predictions for 2013, and feel uncertain about “what a C grade looks like”.

When asked whether they had experienced similar year-on-year variation in other GCSE subjects, the perception of headteachers was that this year’s volatility in GCSE English was unprecedented.

With specific regard to the above hypothesis, the historical stability of GCSE English results and the increased variance in 2012 did contribute to the unexpected results for some centres.

3.4 Controlled assessments had an impact

Hypothesis One

Students completing the GCSE English qualification this year were the first cohort to undertake controlled assessment instead of coursework. The introduction of controlled assessments contributed to the unexpected results.

Findings for Hypothesis One

Whilst some centres said that they found controlled assessment a logistical challenge, on the whole centres did not feel that the format of controlled assessment affected the outcome for their students. One teacher explained how she “understood what was required and there were some benefits to students from this more focused and structured approach”. Teachers told us that they “spent a good deal of time preparing students for the controlled assessment to ensure [they] were confident in controlled conditions and knew exactly what was required of them”.

The only centres that did not conform to this view were the further education colleges that we interviewed. These centres found controlled assessment harder to deliver logistically than traditional coursework and, in their view, this contributed to their unexpected results. The specific challenges cited by the further education colleges included large cohorts, a high proportion of candidates on the C/D boundary, and the need to deliver the whole course in one academic year. One further education college explained how they “found it very difficult to fit in [the controlled assessment] ... and this left only five weeks for exam preparation”.

Most centres did not believe that the structure of controlled assessment contributed to their unexpected results. However, the further education colleges we spoke to believed that controlled assessment format did have an impact on their results.

Hypothesis Two

Controlled assessments accounted for 60% of the marks in GCSE English and English Language. Teachers were responsible for marking controlled assessment, and used students’ controlled assessment marks to inform their predictions. The high proportion of marks accounted for by controlled assessment gave teachers a high degree of confidence in their predictions, heightening the sense of surprise when these predictions were not realised.

Findings for Hypothesis Two

Centres frequently stated that they used controlled assessment marks to make predictions for students’ final grades. Teachers were confident that a student

achieving what the teacher believed to be a 'solid C' in the controlled assessment tasks would replicate that achievement in the examination. As one interviewee stated, "By the time the students took the exam in January 2012 the school knew their controlled assessment marks. [Teachers] therefore used the controlled assessment marks and the exam marks achieved in January 2012 to work out what the students needed to do to achieve their target grades."

Very few of the centres we interviewed had their marks changed through moderation, and this validation of the accuracy of teachers' marking strengthened teachers' confidence in their predictions (see section 3.5).

Centres did use controlled assessment marks as a basis for their predictions for students' final grades. When these predictions were not realised, it was unexpected.

3.5 Impact of Moderation

Hypothesis

Teachers mark their students' controlled assessment tasks. Their marking is then moderated by the awarding organisation. When awarding organisations accepted centres' controlled assessment marks without adjustments, centres interpreted this acceptance as implicit confirmation of the grades that teachers predicted for their students in the controlled assessment tasks.

Findings

The moderation process is a key factor in the level of surprise felt by centres we interviewed. Teachers told us that once they had marked their students' controlled assessment tasks, they then assigned grades to these marks. These grades were assigned based on knowledge of historical grade boundaries and teachers' professional judgement.

For the majority of centres we spoke to, moderators accepted their controlled assessment marks without adjustment. Teachers interpreted the Awarding organisation's acceptance of their marks as validation of the grades that the teacher had predicted. Centres that we interviewed found it difficult to reconcile the awarding organisation's acceptance of the centre's marking with the fact that in some cases their students' final controlled assessment grade did not match their prediction.

Central to the way teachers thought marks converted to grades is the way that the mark scheme works. Controlled assessment mark schemes divide the available marks into bands, and teachers are required to mark within these bands. Officially,

teachers are told by awarding organisations that bands do not directly equate to grades. Whilst in many instances this advice appears to have been given properly, in a large number of cases, centres appear to have been told informally that certain bands equated to certain grades (see 3.6). As one teacher stated, “When it has been clearly stated a certain mark corresponds with a particular grade and then that doesn’t materialise, it is assumed that there has been some ‘tinkering’ with the system.”

Some centres believed that the moderation process had not been effective and fair, and one centre doubted that its sample of marking had been moderated at all. This centre had identified itself as a ‘harsh marker’ after informally comparing its marking with another centre. The centre expected its marks to be adjusted upwards following the moderation process, but it received confirmation from the awarding organisation that its marks had been accepted without adjustment.

Finally, in some cases, centres did receive their expected results for the controlled assessment tasks but they were surprised when their students did not replicate this performance in the written examination. In an extreme case, one centre cited a student who received an A* in their controlled assessment and a U in their examination for English Language. This was improved to an E when remarked.

Where centres’ controlled assessment marking was accepted by the awarding organisation, some centres interpreted this as approval-by-proxy of the way in which they marked their students’ mock exams. When students’ results in the final examinations differed so significantly from their predictions, this contributed to the unexpected results. One teacher explained that he had looked at a group of students’ work in both the examination and the controlled assessment, and believed that it was of the same standard despite the two components receiving different results. Although some centres attributed the discrepancy between grades awarded for the controlled assessment and UMS score awarded for the examination to relatively harsh marking of the examination papers, some felt that there was also a possibility that the moderation process might not have worked effectively.

Most of the centres raising concerns about the accuracy of exam marking used the awarding organisation WJEC. However, it is important to note that, at the time of writing, many centres were still waiting to receive candidates’ written work for the examination, making it difficult to fully explore this issue during interviews.

The centres that we interviewed did equate moderated marks with validated grade predictions and this meant that the results they received were unexpected.

3.6 There was a lack of clarity about the relationship between marks, bands and grades

Hypothesis One

Awarding Organisations failed to clearly communicate to centres that grade boundaries could move significantly within a year.

Findings for Hypothesis One

Most centres interviewed stated that they thought examiners and Awarding Organisation employees gave imprecise information regarding the extent to which grade boundaries could move within a single year.

Some centres reported that, when they pressed awarding organisations for guidance on what the grade boundaries in June 2012 were likely to be, the Awarding Organisations told them to “refer to the January 2012 boundaries and use their professional judgment”. The imprecise advice given by awarding organisations was frequently mentioned by the centres we interviewed, with one teacher feeling that “AQA did not want teachers to work out where they were” in relation to their students’ final grades. This in turn “prevented me doing the job that my headteacher, governors and HMI were expecting me to do in respect of accurately predicting results”. Another centre remarked that their relationship with the awarding organisation felt like “the blind leading the blind”.

The centres that we interviewed suggested that, although advice given formally by awarding organisations (via written communications and specification documents) was cautious about recommending the use of grade boundary information, advice given verbally was less constrained. For example, one teacher explained how a WJEC representative had confirmed during a visit to the centre that ‘Band 3’ in the controlled assessment mark scheme equated to a C grade. Another centre said that AQA had told centres at a regional meeting in November 2011 “that 51 marks for controlled assessment would be a C”. In the event, 54 marks were required for a C grade. The teacher has this written down in his notes. These types of comments were common among centres that we interviewed.

The advice to refer to previous grade boundaries when making predictions appears to have been given to centres without sufficient warning. This is particularly the case for English Language, considering that this was the first year that there had been a standalone Language qualification. Based on interview evidence, it seems that, in some cases, centres also received similar advice from examiners and external consultants.

Centres understood that grade boundaries could change across different exam papers within a year, but teachers had grown accustomed to the scale of such changes being a very small number of marks.

One teacher remarked that she “understood that marks did not equal grades”, but at the same time reflected that “the small print does say that, but years of teaching gives you a feel for it”. Centres 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. A significant proportion of teachers interviewed were also examiners and may have felt that they had an ‘inside view’ into the marking and grading process, from their experience in marking written papers. This then reinforced their confidence in their predictions.

Centres did feel that the advice given to them by awarding organisations lacked precision. Centres understood that grade boundaries could change within a year, but did not anticipate the size of shift that occurred. This all contributed to the unexpected nature of their results.

Hypothesis Two

When marking controlled assessment tasks, teachers expected transparency from awarding organisations about how marking bands mapped onto grades. Centres thought there was a lack of clear guidance that made it difficult for them to understand how marks mapped onto grades, contributing to the unexpected results.

Findings for Hypothesis Two

Teachers we spoke to complained that the descriptors for each ‘band’ were poor and did not allow them to allocate marks properly. The teachers we spoke to frequently explained how the bands were poorly worded, and lacked sufficient detail.

Some centres we interviewed also felt that there was a lack of consistency in the descriptors compared to similar marking schemes from previous specifications. For example, one centre explained how the descriptor ‘Sound use of writing’ has always been used to describe C-grade work. This appeared against the ‘Band 3’ requirement, but once the grade boundary was set, it resulted in a D grade.

A number of centres equated bands with grades, believing that, for example, Band 3 was equal to a grade C. This connection was partly made based on advice from awarding organisations. The change in the grade boundaries meant that a student needed to achieve a mark within Band 4 (the next band up) in order to achieve a C grade. A common observation across centres was that “a C is now at the bottom of Band 4 instead at the top of Band 3”.

Centres feel that they know “what good looks like”, but found it difficult to qualify this within the awarding organisations’ mark schemes for controlled assessment.

3.7 Teachers felt they were prepared for the new specifications

Hypothesis

The GCSE English qualifications were new, both in terms of the syllabus and the way in which they were assessed. The way that some centres adapted to these changes may have contributed to their unexpected results.

Findings

The majority of centres we interviewed felt that they prepared thoroughly for the changes to the content and structure of the GCSE English qualifications. One centre explained how it had allowed “additional teaching time for English subjects”, and had brought in additional teachers to boost capacity. They had attended moderation meetings and INSET training days with AQA, and had worked with other schools to test and compare approaches and standards. Despite these preparations, they were surprised by their results. Preparations made by centres we spoke to included:

- attending awarding organisation conferences and training sessions
- employing external consultants
- co-operating with other centres
- making changes to teaching groups and ways of working.

Some teachers did highlight subtle changes that they had not anticipated. Specifically, the AQA Foundation written paper for English and English Language asked candidates to write their own radio script. Some teachers we spoke to felt that the specification had not prepared them for the possibility of this style of question. One teacher described the question as “a bolt from the blue” and another explained that “students found it difficult to write a coherent and cohesive response”.

Some centres were able to articulate their preparations for the new specification more clearly than others, but very few centres stated that the examination had been significantly different – or more difficult – than compared to previous years.

The change in specification did have a more significant impact on further education colleges. Colleges typically had less time to prepare their students for the new specification (taking the qualification over one year instead of two) and some felt that

this contributed directly to their unexpected results. This does not mean that these centres did not take appropriate measures, only that it was more of a challenge to teach this specification to their students.

The perception among centres we interviewed was that the changes to the content and structure of the assessment did not have a material impact on their results. However, some further education colleges may have been affected by the change in specification.

3.8 Teachers feel that the ‘route’ they took through the qualification had no impact in 2012, except for the way in which the grade boundaries were applied

Hypothesis

GCSE English subjects were structured in a way that allowed schools to choose from a number of possible ‘routes’ to certification. The approach taken by each centre towards the following choices affected whether or not centres received the results they expected:

- **Taking units before the end of the final year (allowing resits if required) versus taking all units at the end of the final year**
- **A single English qualification versus separate English Language and English Literature qualifications**
- **Foundation versus Higher tier papers for exams**

Findings

Based upon the sample of centres that we interviewed, the route chosen by a centre did not appear to affect the extent to which the centre received unexpected results in GCSE English. Centres that entered students into units throughout the two-year course had results that were just as unexpected as those that waited until the end.

For example, one centre entered all its Year 10 students into the AQA Unit 1 examination for English or English Language. It then entered the same group of students into the same examination in Year 11, to give them an opportunity to improve their grade. Whilst they felt that this approach would normally work well for them, this year students tended not to improve on their Year 10 grades in Year 11. The Head of Department “felt fortunate” that the centre had banked marks in 2011 when the grade boundaries were lower.

In addition, June 2011 and January 2012 grade boundary information was often used in predicting results in the same way as centres that either did or did not enter their students into those exams. The use of this grade boundary information was made in the belief that there would not be a significant change in June 2012. When this change occurred, this led to the unexpected results for many centres. As one centre stated, “[We] used January results to help establish a view of the required standard. [We] felt secure in doing this as January seemed to be in line with the previous June [2011].” Any impact from the differing ways that centres chose to enter students into the qualifications was the result of the grade boundary shift; it was not the result of the extra opportunity to re-sit examinations.

Amongst the centres that we interviewed, no pattern emerged around whether students taking GCSE English were more or less likely to achieve their predicted results than students taking GCSE English Language. Nor was there any pattern around students choosing to sit the Foundation examination paper being more or less likely to achieve their predicted grades than students sitting the Higher paper.

The centres that we interviewed did not believe that their entry strategy contributed to the fact that their results were not in line with predictions. But some centres felt that their students would have achieved higher grades had they cashed-in some or all of their units in June 2011 or January 2012.

Appendix 2 – Supporting information

Summer 2012 GCSE English and English Language results at A*–C

In our interim report of 31st August 2012, we concluded that the standard in GCSE English had been maintained, although changes in the mix of students mean that results were down overall by 1.5 percentage points. We have now looked further and have found that like-for-like results went down by only 0.3 percentage points, and that this is within the usual range of year-on-year variation for GCSEs. In this appendix, we explain the analysis that has led us to that conclusion.

The GCSE results reported by JCQ⁸⁹ in August include all candidates who have been entered. The predictions used by exam boards and regulators to review the awards are only available for a subset of candidates – those 16-year-olds who have been matched to their Key Stage 2 results (in this summer's case, from 2007). We call those 'matched candidates'. Because the overall entry can vary from one year to the next, often in ways that exam boards cannot know, exam boards base their decisions on the matched candidates – those they know something about.

We know that the results for **all candidates** reported by JCQ were down 1.5 percentage points. We also know that the results for **matched candidates** were up on predictions. In this appendix we explain the differences between these groups of candidates, their impact on awarding, and the changes in results for the two groups.

The differences between the two groups of candidates

Matched candidates are those who sat Key Stage 2 tests⁹⁰. In a GCSE subject, exam boards can typically match about 75–80 per cent of their entry to Key Stage 2 attainment. In GCSE English and English Language this year, the overall match rate was 82.1 per cent⁹¹.

Unmatched candidates are those who haven't taken Key Stage 2 tests. They tend to be students in independent schools⁹², learners who were educated in another country at Key Stage 2 age, and adult learners. Exam boards do not know the background of each student.

⁸⁹ The Joint Council for Qualifications, which acts on behalf of AQA, City & Guilds, CCEA, Edexcel, OCR, SQA and WJEC (www.jcq.org.uk).

⁹⁰ Key Stage 2 tests are sat in English and mathematics by all 11-year-olds in maintained schools.

⁹¹ Exam boards had Key Stage 2 data for 550,000 students out of the total entry of 670,000.

⁹² Independent schools are not required to use Key Stage 2 tests.

The impact of matched and unmatched candidates on awarding

Exam boards have always used statistical predictions to guide their awards. In the past they were based on the assumption that the same sorts of schools and colleges would perform in much the same way from one year to the next. So an exam board with a very stable entry could expect its results to be stable year-on-year and could predict one year's results by using the previous year's results.

Using that approach was more difficult when the entry from one year to the next was not stable, and there was no way of quantifying the effect of a difference.

As exam boards have become more sophisticated in their collection and analysis of data, predictions have become more sophisticated. Nowadays they are based not only on an assumption of stability from one year to the next, but also on data about the students who are entered. As a result, the exam boards are better able to manage subjects that are not stable from one year to the next. Since 2011⁹³, exam boards have used Key Stage 2 prior attainment data to predict likely GCSE results.

These predictions are used at cohort level (not individual school or student level). Of course, they can only ever be applicable to a subset of the overall entry – those students that can be matched to their Key Stage 2 attainment.

This year, predictions were based on the relationship between Key Stage 2 attainment in 2005 and GCSE attainment in 2010. Exam boards used 2010 as the reference year because 2011 was the last year of the old syllabuses and was likely to have an atypical entry. It is normal practice to use the last stable year of a qualification rather than the last year when it was offered.

Predictions are calculated for matched candidates, and the award is reviewed (within the exam board and by the three regulators) in relation to matched candidate outcomes compared with predictions. Exam boards have no means of predicting the likely results for unmatched candidates, and they have no way of knowing in any detail how the unmatched candidates might be similar or different from one year to the next. It is the matched candidate predictions that guide the awards.

This means that where results are in line with the matched candidate predictions, if the unmatched candidates are similar to those in the previous year, then the overall results for all candidates will be very similar to the previous year. However, if the unmatched candidates are in some way different from the previous year's unmatched candidates – because they come from different types of schools, for example – then the overall results for all candidates will be different from the previous year. In both

⁹³ Prior to 2011, exam boards used Key Stage 3 prior attainment data. This is no longer available because Key Stage 3 tests were discontinued after the 2008 tests.

cases, the senior examiners will have set the grade standard using their professional judgement and guided by the predictions for matched candidates.

What happened this year?

This summer, the results for GCSE English and English Language matched candidates were slightly higher than the predictions.

The results for unmatched candidates were lower. The unmatched candidates this year were a different mix of students compared to the 2011 unmatched candidates, and this meant that their results were different, year on year.

In what way were the unmatched candidates different this year?

At the time that awarding happens, we cannot know the detail of how the unmatched candidates differ, year on year. We have now been able to review the data and information that has become available to see how they differ.

We know that a number of selective and independent schools did not enter for GCSE English and English Language in summer 2012. There were 23,000 fewer students from independent/selective schools.

Students in independent or selective schools generally do much better than other students, as the table below shows. In 2010, for example, well over 90 per cent obtained A*–C in GCSE English, compared to just over 60 per cent elsewhere.

We have estimated the effect of the reduction in students from selective and independent schools on the overall results. We have done this by assuming that students from all centre types in 2012 would have achieved similar results to their peers in 2010, and then calculated the effect that the change in the overall cohort would have had, if everything else had stayed the same.

We calculate that the reduction in the number of students from selective and independent schools – all other things being equal – meant a drop of 1 percentage point at C and 1.2 percentage points at A.

Table 1: GCSE English entries and pass rates in 2010 (actual) and 2012 (assumed) by centre type

	Centre type	Entry	A*–A %	A*–C %
2010	Sel-2	19681	49.2	97.1
	Sel-5	12293	59.0	98.2
	Inde	42671	47.3	90.5
	All Other	634552	12.1	61.6
	Total	709197	16.1	65.0
2012	Sel-2	15333	49.2	97.1
	Sel-5	5468	59.0	98.2
	Inde	31290	47.3	90.5
	All Other	628126	12.1	61.6
	Total	680217	14.9	64.0

Notes:

For 2010, data are from the inter-awarding body statistics for 2010.
For 2012, the entry data are from the 'all awarding bodies' data used for the JCQ analyses.

Sel-2 = Sel (Community/Voluntary-Aided/Controlled); Sel-5 = Sel (Foundation)

(source: data submitted to Ofqual on 30 July 2012)

Other relevant considerations

We estimate that the change in the mix of students – the loss of 23,000 students from independent and selective schools – resulted in a reduction in overall results of about 1 percentage point, although we cannot be precise or certain. Results in GCSE English and English Language fell nationally by 1.5 percentage points.

Some of the fall in year-on-year figures is explained by the inclusion in the JCQ 'English' results of a new Edexcel GCSE in digital communication. This was not available in 2011 and so was not included in that year's results. It is normal for a number of similar qualifications to be grouped under a single title for JCQ results reporting, but the digital communication figures was included and its inclusion explains some of the fall in year-on-year figures. It should not be included when we are comparing like with like.

If we exclude the digital communication figures (just over 3,000 candidates), the proportion of students achieving A*–C in English/English Language dropped by only 1.3 percentage points. Table 2 shows JCQ figures for 'English' excluding digital communication.

Table 2: Summer 2012 English results excluding digital communication

	Total entry	A* and A	A*–C	A*–F
2011	649,557	16.8%	65.4%	97.2%
2012	666,383	15.1%	64.1%	98.1%
Difference		-1.7	-1.3	+0.9

(source: analysis of JCQ data for all candidates, October 2012)

Taken with the data above on the impact of the change in the overall cohort, these data suggest that the like-for-like drop in the proportion of students achieving A*–C in summer 2012 is 0.3 percentage points.

This is in line with expectations and indeed with patterns in other subjects this year. Table 3 shows the variation in other GCSEs with large entries.

Table 3: Variations between 2011 and 2012 for GCSEs in a selection of subjects with more than 150,000 entries

Subject	2011 A*–C	2012 A*–C
Additional Science	66.2%	66.4%
Biology	93.1%	92.6%
Chemistry	93.1%	93.0
Design and Technology	62.4%	62.7%
English Literature	78.4%	76.3%
French	72.5%	71.7%
Geography	70.1%	69.9%
History	70.4%	69.9%
Mathematics	58.8%	58.4%
Physics	93.7%	93.2%
Religious Studies	73.3%	73.7%

(source: JCQ results data for all UK candidates, August 2012)

The performance of matched candidates

Matched candidates did better than the predictions (based on Key Stage 2 data) suggested they would, and better than matched candidates did last year.

The percentage of matched candidates achieving A*–C grades was at or above predictions. Table 4 shows the differences between predictions and actual outcomes for AQA, Edexcel, OCR and WJEC.

Table 4: Predicted and actual outcomes by exam board

	AQA % A*–C	Edexcel % A*–C	OCR % A*–C	WJEC % A*–C
English Language				
Predicted	77.5	77.5	79.0	77.3
Actual	77.5	78.7	80.1	80.9
Difference	0	+1.2	+1.1	+3.6
English				
Predicted	31.3	34.9	33.6	36.1
Actual	32.5	41.5	34.9	35.3
Difference	+1.2	+6.6	+1.3	-0.8
Combined				
Predicted	67.3	65.0	67.8	65.2
Actual	67.6	67.8	68.9	67.5
Difference	+0.3	+2.8	+1.1	+1.3

(source: data supplied to Ofqual by AQA, Edexcel, OCR and WJEC, 31st July 2012)

Predictions based on Key Stage 2

Table 5 below shows the overall predicted GCSE results for 2010, 2011 and 2012, based on the candidates' Key Stage 2 prior attainment. The number of candidates that can be matched is also shown, with the match rate – the proportion of the total entry that the matched candidates represent.

The overall prediction for 2012 (English and English Language) is shown in order to make year-on-year comparisons, but the 2012 predictions are also shown for each of the qualifications separately.

Table 5: GCSE predictions for 2010, 2011 and 2012 based on prior attainment at Key Stage 2

Year	No.	Match rate (%)	Predicted outcomes at each grade (cumulative %)						
			A*	A	B	C	D	E	F
2010 English	392,878	81.5	3.33	14.15	35.79	65.67	86.29	95.25	98.72
2011 English	367,024	81.0	3.85	15.51	37.22	66.80	87.22	95.67	98.80
2012 English Language	430,871	82.1	3.18	15.25	37.33	67.86	87.35	95.81	99.05
2012 English	106,690	83.3	0.28	1.97	8.07	34.14	67.98	88.45	97.22
2012 English Language	324,181	81.7	4.14	19.62	46.96	78.96	93.73	98.23	99.66

Table 5 shows that the matched candidates in 2012 were similar to those in 2011 and 2010. The prediction for 2012 for A*–C is slightly higher than 2011. These predictions were generally met and in most cases exceeded.

Conclusion

The analysis above shows that like-for-like results in summer 2012 went down by only 0.3 percentage points, and this is well within the usual variation for GCSEs. Most of the 1.5 percentage point difference in 2012 results compared with 2011 results is explained by differences in the mix of students without Key Stage 2 results (unmatched candidates) in 2012.

Moderation

What is moderation?

A number of GCSE qualifications contain a combination of internal (controlled assessment) and external (written examinations) assessment. Controlled assessment is used to assess a different mix of skills than can be assessed through written exams alone. In modular GCSEs, the level of controlled assessment varies between subjects. Some subjects, such as economics, law and religious studies, do not have any controlled assessment – that is, all units are externally assessed and marked by the exam board. Some, such as business studies, geography and history, have 25 per cent controlled assessment. Others, such as design and technology, English and English Language, consist of 60 per cent controlled assessment⁹⁴.

In GCSE English and English Language, controlled assessment tasks are set by exam boards, although teachers are allowed to contextualise the task. They are prepared by students in the classroom and are marked by the students' own teachers. Teachers assess the students' answers against a mark scheme provided by the exam board, in the same way as an external marker would mark an exam response. So, just as external markers have their work quality checked, so do teachers when they are marking controlled assessment. This process is called moderation.

The school conducts a process of internal checks to make sure that all assessments within the centre have been marked consistently. The exam board identifies a sample of students' work to be submitted to the exam board for moderation to make sure that schools and colleges are marking controlled assessment tasks accurately and consistently.

What are the rules for moderation?

The *GCSE, GCE, Principal Learning and Project Code of Practice May 2011*⁹⁵ defines moderation as the process through which internal assessment is monitored to make sure it meets required standards and through which adjustments to results are made where required to ensure results are based on the required standard.

The code of practice requires exam boards to moderate the marks submitted by centres against the specified assessment criteria and to adjust the marks submitted by centres as necessary to bring each centre's judgements into line with the required standard. Where the difference between the centre's un-moderated marks and the

⁹⁴ www.ofqual.gov.uk/qualifications-assessments/89-articles/21-gcse-controlled-assessments

⁹⁵ www.ofqual.gov.uk/downloads/category/93-codes-of-practice?download=680%3Agcse-gce-principal-learning-and-project-code-of-practice-2011

exam board's moderated marks exceeds the exam board's specified tolerance limits, marks will normally be adjusted. Unless the marks are demonstrably inconsistent, these adjustments will not change the centre's rank order of students.

For assessments for which no written outcome arises – for example, in some speaking and listening components – exam boards are required to take all necessary steps to ensure consistency of standards, which may include reviewing samples of students' work by visit or by post.

How does moderation work?

Each exam board has in place its own specific process for moderation and is required to provide clear and comprehensive instructions and guidance for internal assessors. In addition to the syllabuses, exam boards produce specific guidance on controlled assessment. This is made available to centres and it covers details of the specific processes, including details such as the sample size and submission dates for each qualification. JCQ also provides guidance⁹⁶ for centres on conducting controlled assessment, including information about moderation.

Teachers mark controlled assessment using the exam board's mark scheme. Where more than one teacher marks a controlled assessment task within a single school, the school must carry out a process of internal standardisation to ensure that marking has been consistent across teachers within the same school.

The exam board will request a sample of the controlled assessment written tasks from each school. The sample size varies depending on the number of students taking the qualification and will be specified by the exam board. The sample will include work from across the range of attainment in the centre.

The purpose of moderation is to make sure that the mark scheme has been applied consistently and correctly, not to re-mark the work of the individual students being sampled. Therefore, so that the moderator can see how the mark scheme has been applied by the centre, the exam board requires teachers to show clearly how credit has been assigned in relation to the criteria defined in the specification. Exam boards provide guidance for centres on how this should be indicated, but it may include, for example, comments in the margins of work indicating where credit has been awarded by the teacher.

The moderator considers a sub-sample of work initially (this is half of the sample submitted by the school) and uses the comments and annotations provided by the

⁹⁶ *Instructions for conducting controlled assessments (GCSE qualifications)*

www.jcq.org.uk/Download/exams-office/controlled-assessments/instructions-for-conducting-controlled-assessments-2011-2012

teacher to compare the marks awarded by the school with those that he/she would give to that piece of work according to the marking criteria. Moderation marks are then compared with the school marks for the sub-sample sample of work. For each assessment, exam boards set a level of 'tolerance' to the marking. This is the number of marks by which the internal assessor's marks may vary from the moderator's marks without being adjusted by the exam board.

The rules for tolerances are set out in the *Common Principles and Practices for moderation*⁹⁷ which are owned and agreed by exam boards. This contains the inter-board rules that are used in determining the level of tolerance allowed in a subject. These rules have applied to GCSE, AS and A level coursework/controlled assessment since the introduction of AS levels in 2000.

This formula for calculating the tolerance is used across all subjects in which moderation is applied and by all GCSE and GCE exam boards. It states that:

- (a) a tolerance limit is established for each unit/component
- (b) the tolerance limit is normally no higher than 6 per cent of the total raw mark for the unit/component, rounded to the next whole number above (for example, 4.1 and 4.8 are both rounded to 5).

Exam boards are free to use tighter tolerances.

Exam board guidance documents, which are available to schools, are clear that tolerances are applied. Although the specific level of the tolerance is not detailed in these guidance documents, it is likely that centres will have an idea of the tolerance based on their experience of moderation and adjustments made by exam boards in previous years.

If the differences between moderator and teacher marks for the sub-sample pieces of work are all within tolerance, then no adjustment is made to the teacher's marks for all students within the school. If the difference for one or more pieces of work is outside of this tolerance, the moderator will re-mark the rest of the sub-sample and an adjustment may be made to the marks of all students within the school. The level of the adjustment is calculated by the exam board based on rules agreed across all of the boards and is applied to all students in the school. The adjustment does not alter the rank order of students as determined by the school.

⁹⁷ JCQ, Moderation of GCE, GCSE and FSMQ Centre-Assessed Units/Components – Common principles and practices, June 2010 (unpublished).

If the moderator significantly disagrees with the centre's rank of students, the centre may be asked to re-consider its marks. If the exam board is not able to determine the appropriate adjustments to make based on the initial sample, it may request additional samples of work from the centre. In some instances, where there are only a small number of students, the moderator's marks may be applied to all students in the centre.

Who is involved in moderation?

Within each exam board and for each subject, a team of moderators is responsible for checking that schools and colleges are marking controlled assessment tasks accurately and consistently. The number of moderators for a qualification must be sufficient to ensure that standards in a given specification can be aligned effectively within and across centres to the timescale required. The moderator roles that are employed by exam boards are:

- Principal moderator – reports to the chief examiner, who supervises the examining process for the whole qualification. They are responsible for setting the controlled assessment tasks (in GCSE English, tasks are set by the exam boards, but in some other GCSEs teachers may set their own tasks, within certain criteria) and supervising all moderation in that subject. The principal moderator must have substantial relevant experience of teaching and assessment in the subject area concerned.
- Team leaders – responsible for supervising and monitoring a group of moderators, to ensure the quality of their moderation meets the required standard. They also carry out moderation themselves.
- Moderators – moderate the marking of controlled assessment across a number of schools and colleges. Most moderators are current teachers.

Standardisation/training of moderators and guidance

Exam boards are required by the code of practice to provide training and follow-up guidance for moderators in all relevant aspects of the internal assessment units and moderation procedures. All moderators must take part in the training to ensure common understanding of procedures and standards before moderation begins, and all principal moderators have to undertake some first-hand moderation of centres' assessments in each examination series. The standards set by each moderator must be checked by a more senior moderator who repeats the assessment of students' work on a specified sample. The senior moderator completes a report for the exam board on the work of each moderator. Moderators whose standards are judged to be unsatisfactory may not continue their work, and all work from their centres must be re-moderated.

Exam boards provide guidance to their moderators on the moderation of controlled assessment. This includes information on both standardisation and moderation, including how to sample, what to do if any issues are encountered and how to feed back to centres.

Speaking and listening

Speaking and listening makes up 20 per cent of the assessment in GCSE English, but there is no evidence produced for moderation. Exam boards therefore adopt a different approach for the moderation of these assessments.

To make sure that schools' marking of these activities is consistent and that the mark scheme is being applied accurately, exam board moderators visit a number of schools each year to observe a sample of GCSE English students taking part in speaking and listening tasks. Typically, a third of schools are visited each year and all schools are visited on a three-year cycle.

Moderators will usually observe a sample of six to eight students at the centre performing each of the required speaking and listening tasks to determine whether the centre's marking is consistent with the mark scheme for the specification. The moderator may provide feedback to the centre, but will not make adjustments to schools' marks on this visit. If the moderator has concerns about the school's marking, then this will be referred back to the exam board and it is likely that a second visit will be undertaken by a senior moderator who will decide whether any adjustments are required. Where adjustments are required, these will be calculated by the exam board and applied to all students at the school. If a school does not receive a moderation visit from the exam board, their marks will not be adjusted.

Awarding, UMS marks and grade boundaries

How awarding works

Awarding is the process by which exam boards convert students' marks from across their various GCSE units (exams and controlled assessment) into a single grade from A* to G.⁹⁸

Awarding can't take place until the students' work has been marked and moderated, so it takes place at the end of the process. Each exam board has an awarding committee for each syllabus being awarded, chaired by the chair of examiners for that subject. The awarding committee includes the chief examiner for each subject and any principal examiners (those who write any question papers), any principal moderators (responsible for any controlled assessment units), the reviser (responsible for 'revising' the question papers to ensure consistency) and an exam board officer.

The awarding committee meets face to face or remotely to consider the evidence and recommend where the grade boundaries should be set. The Code of Practice requires them to consider student work as well as available statistical data. The main types of evidence they use to reach their judgements are:

- samples of students' work from that exam series, and samples from previous exams where they are a valid reference point
- reports from senior examiners and moderators about how well assessments have worked in practice
- grade descriptions – these define the typical performance levels expected at the judgemental grades
- statistical information – this includes the predictions about how well this cohort of students will do based on their attainment in Key Stage 2 assessments, results from previous exam series, and teachers' estimated grades for their students.

Grade boundaries are set on each unit in turn. Boundaries for three of the eight GCSE grades – A, C and F – are set in this way. These are called judgmental grade boundaries. The other boundaries are calculated arithmetically, relative to the judgemental grade boundaries.

⁹⁸ Detailed guidance on the awarding process is provided in the *GCSE, GCE, Principal Learning and Project Code of Practice* www.ofqual.gov.uk/downloads/category/93-codes-of-practice?download=680%3Agcse-gce-principal-learning-and-project-code-of-practice-2011

The grade boundaries are set on the raw marks – the marks the candidates actually scored on the paper. These raw marks are then converted into uniform marks, or UMS (uniform mark scale) marks, and the UMS marks are added together to give the overall grade for the qualification.

No account is taken during awarding of the protected characteristics of any of the students, nor of any other factors other than the achievement of all students in the exams or other assessments being graded. Ofqual's policy is that any factors that may disproportionately affect students with protected characteristics should be removed at the design stage (when the relevant regulatory rules are devised and the qualifications are accredited). Minimising bias is one of the five criteria used by Ofqual to determine whether a qualification is fit for purpose.

Why not just use pre-set grade boundaries?

Examiners writing question papers and controlled assessment tasks are aiming to ensure the level of demand is similar from one exam series to the next. However, assessment experts recognise that it is not possible to judge the level of demand for students until their responses have been completed and marked. Sometimes a question, or a combination of questions, can turn out to be more or less demanding than expected.

So, the number of raw marks needed for a given grade will vary between different exam series to ensure that the standard remains the same.

This is particularly important in modular qualifications, where students sit their units at different times in the course. We know that GCSE English students who cashed in their GCSE this summer may have entered units in one or more of the four exam series that have taken place. So, GCSE students this summer will have sat the same units, but they will have taken different papers, depending on whether they entered in January or June, this year or last year.

Since the papers and tasks are different, the level of demand will vary. Awarders will take that into account and set raw mark grade boundaries at different marks – so the same mark on a harder paper represents better performance. It wouldn't be fair to candidates to add up the raw marks to give the overall result because some students will have taken more demanding papers.

The uniform mark scale (UMS) puts all these raw marks on the same scale. The raw mark is converted to a UMS mark, which reflects the position of that mark relative to the grade boundaries set in that unit.

So, in a unit in January the boundary (minimum) mark for a grade C might have been 48, and in June the boundary mark for that same unit was 46. A candidate who scored 47 in January would have been just below the number of marks required for a

grade C. However, a candidate who scored 47 raw marks in June would have achieved a grade C (see table 6 below). The UMS mark awarded to these candidates would reflect that difference.

Table 6: The unit grades resulting in changes to grade boundaries

	Grade boundary	Raw mark	Unit grade
January	48	47	D
June	46	47	C

The uniform mark scale is used in all modular qualifications, including modular GCSEs and A levels. The UMS boundaries are fixed but the raw mark boundaries will vary between series.

The new English syllabuses use UMS marks, but the old syllabuses did not need to and did not. Under the old syllabus, schools and colleges would receive component marks⁹⁹ for students, but not component grades. In the new syllabuses, schools and colleges can work out a grade for each unit, based on the UMS grade boundaries. However, the overall qualification grade is calculated by adding together the UMS marks for the qualification. A grade C on a unit does not necessarily mean a student will achieve a grade C for the whole qualification.

Tables 7–11: Raw mark grade boundaries for all exam boards, January 2011–June 2012

The following tables show the raw mark grade boundaries for each exam board in all four exam series. Where units were not available, cells are shaded. Where there were no entries, and therefore there was no need to set a boundary, there is a dash.

AQA

Unit	Subject	%	Type ¹⁰⁰	Max mark	Jan 2011	Jun 2011	Jan 2012	Jun 2012
ENG1F	Common	40%	WP	80	46	44	43	53
ENG1H	Common	40%	WP	80	40	43	41	44
ENG02	Common	20%	CA	45		25	25	28
ENG03	Eng	40%	CA	90	-	51	51	54
ENL03	Lang	40%	CA	80	44	43	43	46

⁹⁹ A component is a written paper or coursework task/portfolio.

¹⁰⁰ WP = written paper; CA = controlled assessment

CCEA

Unit	Subject	%	Type	Max mark	Jan 2011	Jun 2011	Jan 2012	Jun 2012
GEG11F	Common	20	WP	48		34		33
GEG12H	Common	20	WP	48		23		22
GEG21F	Common	20	WP	48				35
GEG22H	Common	20	WP	48				22
GEG31	Common	20	CA	75				47
GEG41	Lang	40	CA	80				46
GEG51	Eng	40	CA	120				71

Edexcel

Unit	Subject	%	Type	Max mark	Jan 2011	Jun 2011	Jan 2012	Jun 2012
5EH01	Common	20%	CA	40	23	23	24	24
5EH2F	Eng	40%	WP	96		65	66	74
5EH2H	Eng	40%	WP	96		33	34	42
5EN2F	Lang	40%	WP	64		41	41	43
5EN2H	Lang	40%	WP	64		21	23	23
5EH03	Eng	40%	CA	96		55	55	65
5EN03	Lang	40%	CA	96		58	60	64

OCR

Unit	Subject	%	Type	Max mark	Jan 2011	Jun 2011	Jan 2012	Jun 2012
A641	Eng	20%	CA	40	21	21	22	25
A642	Eng	20%	CA	40	-	21	23	26
A643	Eng	20%	CA	40	23	23	23	26
A651	Lang	30%	CA	60	32	31	31	35
A652	Lang	30%	CA	60	-	34	34	38
A680F	Common	40%	WP	80	53	55	53	56
A680H	Common	40%	WP	80	43	37	37	39

WJEC

Unit	Subject	%	Type	Max mark	Jan 2011	Jun 2011	Jan 2012	Jun 2012
4171F	Common	20%	WP	40		27	31	27
4171H	Common	20%	WP	40		17	16	18
4172F	Common	20%	WP	40		27	27	26
4172H	Common	20%	WP	40		17	15	19
4193	Eng	40%	CA	80				50
4194	Eng	20%	CA	40				26
4173	Lang	30%	CA	80				48
4174	Lang	30%	CA	60				37

Awarding and grading: the comparable outcomes approach

In this section, we set out a brief history of grade awarding in A levels, O levels and then in GCSEs. We explain how we have regulated grade standards since the introduction of the new AS levels in summer 2009. We then consider the question of whether all the exam boards interpreted and implemented our instructions in the same way.

A brief history of grade awarding

There is a widely held belief within education that A level grades (first introduced in 1960) were set by a system of norm-referencing – a fixed proportion of students getting each grade every year. This continued until 1987 when grades began to be set by criterion-referencing – awarding grades to students who have reached a particular level of performance. This is discussed further in a 2011 paper by Paul Newton¹⁰¹, a member of the Ofqual Standards Advisory Group.

Newton argues that the awarding of public examinations has always used what he calls the ‘similar cohort adage’: the idea that if the cohort hasn’t changed much then we wouldn’t expect the proportions at each grade to change much either. He argues that this “is a rule-of-thumb that the examining boards in England have taken to heart and have integrated with their methodologies for maintaining standards”.

When GCSEs were introduced, the similar cohort adage was followed. Exam boards continued to balance the evidence from students’ work with statistical expectations. This has not changed since the introduction of GCSEs.

However, many other things about GCSEs have changed. GCSEs are now modular, so students can take (and re-sit) units at various points during their course. And exam boards have developed more sophisticated ways of predicting achievement and judging the comparability of grade standards. But the way in which they use these statistical predictions is really no different from the way in which statistics were used 50 years ago.

We know that the current system can lead to grade inflation, for various reasons. Awarding committees may apply the benefit of the doubt when making decisions about whether to set a grade boundary mark at, say, 67 or 68: they will often recommend the lower mark, which will inflate rather than deflate overall outcomes.

These decisions are understandable and are made with the best intentions. But if they are made each time grade boundaries are set, the cumulative effect of these

¹⁰¹ Newton, P. *A level pass rates and the enduring myth of norm referencing*. In *Research Matters* (October 2011), Cambridge Assessment

small changes is that results creep slowly upwards, without there being any real evidence of improvement in the quality of students' work.

Ofqual's approach

In 2001/02, discussions between exam boards and regulators about maintaining standards in the new 'Curriculum 2000' AS/A levels led to an agreement to prioritise 'comparable outcomes':

The comparable outcomes perspective implies that grade boundaries should be fixed so as to take account of any deficits in ... examination performance which are unique to the first cohort of candidates. On the other hand, the comparable performance perspective entails an acceptance that candidates' results in [the first year of a new syllabus] should suffer because for this reason they did not produce performances comparable to those which would have been achieved by candidates [in the previous year].¹⁰²

By 'comparable outcomes' we mean the proportion of students achieving each grade. So if we aim for comparable outcomes, and all other things are equal, then roughly the same proportion of students will achieve each grade as in the previous year. We would still expect to see year-on-year variation, but we would also expect to see variation in both directions – some results to be up on the previous year and some subjects to be down. Now that we're in the third year of monitoring the grade standards at A level, that is what we have seen.

In addition to comparable outcomes, we also talk about 'comparable performance'. This means the way that students perform in the exam. We know that performance 'dips' in the first year of a new syllabus. So if we aim for comparable performance in the first year of a new syllabus, it will make it harder for students to get each grade, compared with students in the previous year.

There are good reasons to aim for comparable outcomes in the first year of a new syllabus. Students taking their A levels or GCSEs in any particular year will be competing with those from other years for access to higher education and employment. It gives some students an undeserved advantage if they get better results simply because they were taking an exam that their teachers were used to preparing them for. Students should not be advantaged or disadvantaged just because they were the first to sit a new set of examinations. The only exception to

¹⁰² Cresswell, M. J. (2003) *Heaps, prototypes and ethics: the consequences of using judgements of student performance to set examination standards in a time of change*. University of London Institute of Education

this would be when the new syllabuses were designed to be at a different standard, as they were in the new GCSE Science exam this summer.

There are also good reasons to prioritise comparable outcomes after the first year of a new syllabus. If we accept that performance in the first year's exams will be a little lower than in subsequent years, and then we use the performance in those first year's exams as a benchmark for the subsequent years, we will be locking in a lower standard. Over time, this is likely to result in grade inflation with, each year, gradually more candidates achieving each grade.

This is the approach we have taken to the first awards of new AS levels in 2009, new A levels in 2010, as well as for the subsequent awards of AS and A levels in 2011 and 2012. Exam boards use statistical predictions based on prior attainment at GCSE and they report their outcomes against those predictions. We have agreed tolerances with the exam boards, and if their results are outside tolerance (up or down) then they have to provide evidence to justify those results.

This approach to AS and A levels has produced a stable set of results year on year, as can be seen from Table 12 below.

Table 12: A level outcomes for all exam boards at A*, A and E, 2008–2012

A level	2008	2009	2010	2011	2012
Grade A* (%)	-	-	8.1	8.2	7.9
Grade A (cum. %)	25.9	26.7	27.0	27.0	26.6
Grade E (cum. %)	97.2	97.5	97.6	97.8	98.0

(source: JCQ provisional data for all subjects)

In 2011, we used the same approach for the awards of the first new unitised GCSEs in all subjects except English, mathematics and ICT, using prior attainment at Key Stage 2.

Table 13 below shows the outcomes for all exam boards for the GCSEs that were new in summer 2011¹⁰³.

¹⁰³ All subjects except the English suite, mathematics, the science suite and ICT.

Table 13: Outcomes at A*, A and C for all exam boards in new GCSEs, 2011 and 2012

Grade	2011 (%)	2012 (%)
A*	10.2	9.9
A*/A	27.8	27.7
A*–C	73.5	73.9

Is it the best approach?

Setting standards in GCSE and A level qualifications is a complex process requiring awarders to balance the effects of many different factors that can play out in different ways in different qualifications.

Using comparable outcomes is the best approach we have to maintain standards in new qualifications, to stabilise results, and to end the year-on-year grade inflation that has undermined public confidence in our qualifications system. It also gives us a way of achieving comparable grade standards between the different exam boards. These are the benefits, but the approach also has limitations.

The predictions used to achieve comparable outcomes assume no change in the value-added¹⁰⁴ between the prior attainment measure and the subsequent qualification, in this case between Key Stage 2 and GCSE. Change in the qualifications system (either at Key Stage 2 or at GCSE) could alter the value-added rate.

Predictions are based on the entire cohort and are most reliable when used at a cohort level. But different subsets of that cohort may have different value-added rates. If the cohort mix changes, the predictions may be less reliable.

There are also questions about the suitability of the predictions in practical and performance subjects, such as art and design, music and PE.

A more practical limitation is that Key Stage 2 tests are not taken in Wales or Northern Ireland, so the predictions can only ever be used for candidates in England.

There are also suggestions that this approach does not recognise genuine improvements in national performance in a subject. That is not the case. If exam boards had evidence that the level of performance was at odds with the statistical predictions (either because performance was better or worse than expected) we

¹⁰⁴ 'Value-added' is a term used to describe student performance between different stages of education compared to other students with the same level of attainment at one point. For more detail see www.education.gov.uk/performance/tables/schools_04/sec3b.shtml

would expect them to be able to provide evidence for us to consider. In the time we have been using the comparable outcomes approach, there have been some examples of this – for example in AS level World Development and A level Critical Thinking.

In June 2009, WJEC reported to us that outcomes for AS level World Development (for which WJEC is the only provider) would be out of tolerance at both A grade and E grade. WJEC provided evidence from the awarding committee who had reviewed samples of student work and did not find that the grade boundaries suggested by the statistical evidence reflected the correct standard. We accepted this, and subject outcomes were 2.5 per cent lower at A grade and 3.2 per cent lower at E grade than predictions.

In January 2012, OCR reported that they were concerned that unit outcomes in AS Critical Thinking unit F501 were much higher than predictions, and that may lead to subject outcomes in June 2012 being out of tolerance. OCR and AQA conducted an analysis of the centres taking AS level Critical Thinking in the two years and found a change in the mix of schools offering the qualification in 2012 compared to 2011. This affected all schools taking the qualification, which is only offered by AQA and OCR. We accepted this analysis and revised predictions were calculated, meaning that students received higher grades than the comparable outcomes approach indicated.

On balance, it is our view, endorsed by our Standards Advisory Group¹⁰⁵, that this is currently the best approach we have to maintaining standards in the first years of new qualifications. In March of this year we discussed with the Standards Advisory Group a need to review the approach as we move into a period of greater stability in the short term, but with longer-term changes on the horizon. We have already started to discuss these issues with the exam boards and we will be talking again to our Standards Advisory Group in the coming months.

¹⁰⁵ The Standards Advisory Group is a committee of the Ofqual Board, formed earlier this year to consider and advise Ofqual on qualification and assessment standards issues. The group is made up of independent assessment specialists with a wide range of expertise and members of the Ofqual Board, and is chaired by Amanda Spielman. A full list of members is provided in Appendix 5 to our initial report.

Glossary

Accreditation – A regulatory activity, checking that the syllabus for a qualification submitted by an exam board meets the requirements of the relevant qualification and subject criteria.

Assessment criteria – Set out how well a student has to perform in order to be awarded a particular mark or grade.

Awarding – Where the marks candidates have been given for all units/components of a qualification are converted into overall grades.

Awarding committee – The group of people responsible for setting the grade boundaries for an exam or controlled assessment. The group comprises a chair of examiners, chief examiner and one or more principal examiners. The committee will also review overall grade outcomes for a qualification.

Centre – An organisation (such as a school or college) undertaking the delivery of an assessment to candidates on behalf of an awarding organisation.

Chair of examiners – An individual responsible to the awarding organisation for maintaining standards across different specifications in a subject within a qualification and from year to year.

Chief examiner – An individual responsible to the chair of examiners for ensuring that the exam as a whole, including both internal and external assessment, meets the requirements of the specification and maintains standards from one year to the next. The chief examiner also acts as a principal examiner or moderator for at least one component.

Comparable outcomes – An approach used to maintain standards.

Controlled assessment – Internal assessment carried out by students under controlled conditions in their school or college. This replaced coursework in 2009.

Diploma – A type of qualification first introduced in 2008 and made up of three parts: principal learning, generic learning and additional and specialist learning. Not available after December 2013.

English Baccalaureate – Introduced by the Department for Education as a performance measure in the 2010 performance tables. It recognises where students have secured a C grade, or better, across a core of subjects – English, mathematics, history or geography, the sciences and a language. It is not a qualification in itself.

Exam board – Also referred to as an awarding organisation or an awarding body. An organisation recognised by us for the purpose of awarding regulated qualifications. We generally use the term ‘exam board’ to refer to awarding organisations offering GCSEs and A levels.

Examiners – Individuals with subject expertise who are responsible for marking candidates’ responses. They usually work as part of a team led by senior examiners whose roles include chair of examiners, chief examiner and principal examiner. Many examiners are serving teachers.

Foundation tier – For some subjects, all candidates will sit the same exam, but in other GCSE subjects there are two tiers: ‘higher’ or ‘foundation’, and students are entered for one or the other. Each tier leads to a different range of grades: higher tier exams lead to grades A*–D, foundation tier exams lead to grades C–G.

GCSE (General Certificate of Secondary Education) – Generally taken as a two-year course by students aged 14–16. GCSEs sit in the National Qualifications Framework (NQF) in England, Wales and Northern Ireland. GCSEs are graded A* to G, grades D–G being at level 1 and grades A*–C being at level 2 in the NQF.

Grade boundary – The minimum uniform mark that a candidate needs to score to achieve a particular grade for an exam or controlled assessment

Higher tier – See Foundation tier.

Internal assessment – Internal assessment is work done and assessed internally, by a school or college. Conversely, external assessment is work, usually an exam, assessed by exam boards. Internal standardisation is carried out on internal assessment within the school or college to check the marking of an assessment is consistent across staff.

Legacy GCSEs – GCSEs that are currently being delivered in schools, but which are being replaced by new syllabuses.

Linear GCSE – GCSEs where all assessment is completed at the end of the course.

Mark scheme – A scheme detailing how credit is to be awarded in relation to a particular assessment unit or component; a mark scheme normally characterises acceptable answers to questions or tasks or parts of questions or tasks and identifies the amount of credit each attracts. Also referred to as marking criteria.

Matched candidates – Those candidates that can be referred to when applying a comparable outcomes approach to standards. Matched candidates are those from Year 11 where their Key Stage 2 results are also known.

Moderation – The process through which the marking of assessments by schools or colleges is monitored to make sure it meets required standards and through which adjustments to results are made, where required, to ensure that results are based on the required standard.

Modular GCSE – Modular GCSEs are GCSEs made up of several units that are tested individually throughout a GCSE course. They are also known as unitised GCSEs. Breaking the course into modules means that candidates are able to sit assessments throughout the course. However, candidates must be entered for at least 40 per cent of the assessment for a GCSE subject in their final sitting.

Principal examiner – An individual responsible for the setting of the question paper or task and the standardising of its marking.

Programme of study – Sets out the prescribed syllabus or specification that students must be taught in that subject.

Qualifications and Credit Framework – A credit transfer system for qualifications in England, Northern Ireland and Wales. Every unit and qualification in the framework has a credit value, as well as clear details about difficulty (from Entry level up to Level 8) and content.

Qualification criteria – Set out the basis for developing qualifications at a certain level or type. The criteria provide a framework for an exam board to follow when creating a qualification including titling, subject matter, structure and assessment.

Raw marks – What candidates actually scored for a unit/component. The raw marks are converted into uniform marks for use in awarding.

Responsible Officer – The person in each awarding body who is ultimately responsible for the standards of all exams offered by that awarding body, as required by our General Conditions of Recognition.

Subject criteria – Set out the knowledge, understanding, skills and assessment objectives for a particular subject. The criteria provide a framework for an exam board to create a qualification syllabus.

Syllabus – A syllabus belongs to a specific exam board, and describes the content of a qualification and how it will be assessed. Also known as a specification.

Terminal rule – A rule imposed to make sure that a certain percentage of any GCSE has to be sat or submitted at the end of the course of study.

Tolerance limits – Part of the comparable outcomes approach to standard setting. Tolerance limits are set around predicted qualification outcomes to take into account the number of entries and other statistical factors for particular subjects. The

tolerance limits act as triggers for exam boards to report differences from expectations.

Uniform marks – Uniform marks are given using the Uniform Mark Scale (UMS). The UMS balances out differences between exams/controlled assessments which are used for different sittings of the same unit.

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2012

© Crown copyright 2012

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: psi@nationalarchives.gsi.gov.uk

This publication is also available on our website at www.ofqual.gov.uk

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346