

GCSE English Awarding 2012

Supplementary memorandum from Ofqual



Introduction

1. Further to the memorandum submitted on 6 September, this memorandum provides the Select Committee with additional information about our work on GCSE English awarding in advance of the oral hearing on 11 September. The body of the paper sets out our view about what happened with GCSE English awarding this year and summarises the issues arising. Annexed to the paper is a more detailed discussion of some of the issues we have been considering, and how we have been thinking about them.

2. We welcome the opportunity to answer Select Committee's questions tomorrow.

What happened

3. Overall, GCSE English results were down at A*-C by 1.5 percentage points. This was in line with expectations because of changes to the student mix. But there has been an unusual distribution pattern, a greater variation between schools than expected. And for some schools, the results are a far cry from their expectations. We have been looking carefully at what lies behind that.

4. There are three points it is important to make up front.

5. First of all, there has been no political interference.

6. Secondly, awarding and grade boundary setting worked as it should have done for the English suite, as for other GCSEs and A levels. We played our proper role, regulating standards, and using the comparable outcomes approach. Awarding in January and before was generous, but neither the exam boards nor the regulators could have seen it at the time. All the evidence at the time was that awarding decisions were if anything harsh.

7. Thirdly, GCSEs are modular and students' overall GCSE results are influenced by the routes they take through the qualifications – which units they take, and when, and whether they resit.

The impact of choices

8. Route choices are generally made at school or college level, by schools rather than students. Schools make those choices using their best judgement as to how to get the best outcomes – to make sure that the students get the results they deserve. These choices are particularly important, particularly sharp in English. As is normal with modular qualifications, students' results are influenced by the decisions their schools and colleges made about how to approach the qualification as well as by the quality of their work.

9. Some schools chose to put their students through all the modules at the end of the course – to put students through a modular qualification in a linear fashion, with all units sat at the end, in June 2012. The rules allow for this. We can confirm that some of those taking the qualification in this way did worse than they expected. They did worse than those who sat the qualification as it was designed, in a modular fashion, and noticeably worse than the small numbers who finished in January.

10. A small proportion put candidates through the English qualifications early, with final modules taken in January 2012. We are looking into that. Those students did comparatively well, and one of the reasons for that will be the grade boundaries for units sat in January 2012. We now know they were generous, but exam boards could not have known that at the time.

11. Candidates taking qualifications in a modular way will tend to have an advantage – they can see how they are getting along during their period of study, as unit results come in. They are clocking up points. And if they do not do as well as expected in a unit they can resit that unit at a later session. So you can expect them to do better, all other things being equal. That is the case for all GCSEs – indeed all modular qualifications - not just the English suite. And on average they did do better in the English suite, this time, with the effect being exacerbated by the generous grading in January and before.

12. So, some students did worse than expected. Expectations in some schools were raised because they saw the January grade boundaries when they were published by exam boards in March. Schools are told not to rely on those grade boundaries (though there has been some suggestion that some schools may have got different messages from exam boards, and we are exploring that), and they know they can change and do change, but we suspect they were relied on particularly here because the qualifications were new and there was little else to go on. And, because they were generous, expectations were raised. We know of some schools expecting a 15 percentage point increase in achievement at Grade A*-C, for example.

13. From the work we have done so far using the data we have collected from exam boards, the difference in outcomes at Grade C between those who took a linear route and those who took the modular route is not particularly unusual and not out of line with what we would expect from analysis we have done in the past on this issue. We will do more work to test this.

14. We also know – from the information that schools give to exam boards - that schools were predicting that 15% more students would achieve at least a C grade, as compared to last year. Schools generally over-predict. In usual years they predict, on

average a 12% increase at this level – so predictions, and expectations were more optimistic than usual.

15. We are turning now to school variations, and looking at those schools where the results were most out of line with their expectations. We know that the proportion of schools that had very significant negative differences is very small – less than one per cent. We need to know more, school by school, about what lies behind their results. We know that one consideration will be the student mix, and the proportion of students judged to be on the C/D borderline.

Underlying considerations.

16. There are underlying considerations that are relevant, and that might be particularly relevant to some schools.

17. Firstly, controlled assessment. 60% of the English suite assessment is by way of controlled assessment, supervised and marked by teachers in schools. A third of that controlled assessment is assessing what are known as speaking and listening skills. Assessing speaking and listening skills is notoriously difficult, but speaking and listening is in the national curriculum. Most subjects have controlled assessment as part of the overall assessment at GCSE level. For most other subjects where controlled assessment is used, the proportion of assessment that is controlled assessment is 25%. For the English suite it is 60%.

18. Teachers mark controlled assessment. A sample is then moderated by exam boards, and exam boards allow a tolerance of 6%. We think that controlled assessment is problematic, and we are reviewing it in all subjects, but we can see that it is particularly problematic in English, and for speaking and listening.

19. Teachers received the moderated papers back from exam boards with marks confirmed, in the main. Some teachers assumed then that the mark meant the grade – by reference to January grade boundaries. It did not – but again, expectations were raised. We are looking closely at how controlled assessment worked for the English suite.

20. Secondly, there is a need to make sure that standards are maintained in a unitised qualification. 40% of the assessment must take place at the end, and this rule – known as the terminal rule – gives headroom to enable exam boards to get it right. We know that exam boards followed the rules of awarding. The three regulators played their part, querying and challenging provisional outcomes. We challenged two of the four providers – Edexcel and WJEC – to bring their qualifications in within acceptable boundaries. They did so – we did not need to direct them to do so. There is a suspicion that this put undue pressure on the grade boundaries in June. It did not. It is the regulator's job to challenge outcomes, and we did. But there is a recognised tension, for modular specifications, between getting the unit grading right and making sure the qualification level outcomes are right.

21. Thirdly, these were new qualifications. The system is used to changes in qualifications, with GCSEs changing every five years or so. The changes in the English suite were not insignificant, and there were changes to entry patterns as well. Schools made choices, student by student between English, or English Language

and English Literature, and some of the units were what are known as common units – that is, they could be used to contribute, ultimately, to English or to English Language. Schools could delay final decisions. All this made the job of predicting and awarding complex for exam boards.

22. Lastly, of all qualifications and grades, the most weighty pressures come to bear on English Grade C. It has the most significant role, above any other qualifications measure, in the way that schools themselves are assessed.

23. We believe that the combination of pressures has created particular strains and tensions, now evident.

Emerging considerations

24. At this stage some important points emerge.

25. Modularisation has increased the difficulty for exam boards of maintaining comparability across all specifications and pathways through English/English Language GCSE: this is problematic given the particularly high significance placed on Grade C in English (and mathematics) in accountability systems.

26. The ability to defer some of these decisions (ie choice of GCSE (English or English Language) as well as timing of unit entries and resits) until late in the course has seemingly intensified tactical concerns, as schools gather intelligence about the apparent attractiveness of available pathways. This intelligence includes news from other schools about their successes and disappointments in individual units, as well as previous unit grade boundaries published by exam boards, and also affects school expectations of outcomes.

27. The increase in the teacher-marked controlled assessment (previously coursework) component of the qualification from 40% to 60% of the qualification may have contributed to very high school expectations of outcomes on the new English/English Language GCSEs. Many schools are disappointed, not because their results are down, but because they were not up more.

28. There is some confusion in schools between marking and grading, especially in the context of controlled assessment. Some schools appear to have believed that controlled assessments were graded to pre-set pass marks.

29. Accountability pressures encourage schools to focus on tactical decision-making about their choices from among the proliferating routes through GCSE English/English Language as well as on teaching and exam preparation.

30. The possible consequences of the extensive changes to English GCSE do not appear to have been fully considered at the time that the new set of English qualifications was designed.

Ofqual, 10 September 2012

Annex: further discussion

Comparable outcomes

31. In 2010 Ofqual adopted the 'comparable outcomes' approach to securing standards in new GCSEs, an approach we had already started to use for the new A levels. The meaning of 'comparable outcomes' is best understood from the perspective of the person sitting the exam. The basic principle is that, other things being equal, a young person should have the same chance of achieving a given grade, no matter which exam board their school uses and no matter which year they take the exam. The outcomes should be comparable.

32. The exam boards use data on Key Stage 2 attainment to judge the expected attainment of each GCSE cohort, for that majority of GCSE entrants who have a Key Stage 2 result ('matched candidates'). So the national predictions are adjusted if the mix of candidates changes. This year there were several changes in the entry cohort that affected the mix, as we noted in our report.

33. 'Comparable outcomes' are used by exam boards as a check at qualification level. When considering what the grade boundaries should be, examiners see a comparison between the grade outcomes predicted by candidates' Key Stage 2 test results and the proposed actual outcomes given their judgement about where the standard should be set. This testing of examiner judgement against statistical predictions is central to comparable outcomes, and has the potential to give us the best of both worlds. At AQA, for example, about 0.3% more of this year's matched candidates reached Grade C or better in English/English Language combined than the statistics would predict. There are tolerance limits agreed between exam boards and Ofqual which allow for some divergence (typically $\pm 1\%$), reflecting the fact that statistical predictions can never be wholly accurate. Beyond these tolerance limits, the exam boards can justify improvements in grades by reference to genuine educational improvement or deterioration.

34. This approach has been used for some years to improve comparability between exam boards, and also within exam boards to contextualise their judgements.

35. The change that has coincided with the introduction of this qualification is the strengthening of the requirement for exam boards to justify increases beyond tolerance to the regulator. This change was agreed in December 2010, in advance of the first awards of the new GCSEs in most subjects in 2011. This approach has also been applied in A level and AS awarding since 2009 and, in the A level context, has been the subject of a positive independent evaluation by NFER last year. We published details of our approach to try and help schools, colleges and others to understand what we were doing.

36. Questions have been asked about the appropriateness of the application of comparable outcomes in the first awards of a completely new GCSE. This is actually the point at which it is most important to be able to anchor examiner judgement. Their job is to carry forward over time the collective understanding of the level of educational achievement that deserves an A, C or G grade, and to interpret that in the context of changing assessments and other contextual factors. There is always

the risk that performance will dip in the first year of a new qualification, as teachers get used to new curricula and assessment approaches, and comparable outcomes is intended to mitigate that – and then make sure that the standard is carried forward into future years as teachers get used to the new qualifications.

37. One of the biggest challenges in maintaining qualifications standards is to distinguish between improvements in outcomes which are attributed to genuine improvements in attainment – as a result of better teaching and learning – from those which arise from other factors, such as greater teacher familiarity with the assessment or changes to the qualification structure (sometimes called ‘grade inflation’). This is particularly the case when qualifications change. If there were a substantial change in outcomes at national level at the point where a qualification changed, it would be more likely that this could be attributed to the change to the qualification rather than any real changes in attainment – given that schools, teachers and the National Curriculum are all much as they were in preceding years. If the change was a fall in attainment, it may be that a lack of teacher familiarity with the assessment approaches or new subject content could explain the change. If a rise in attainment were reported it could be that the assessment structure benefited students – particularly, as with GCSE English in 2012, when there was a move from a linear to a modular structure, allowing for more retakes. For this reason, we think it is right to have a strong framework to enable challenge to and testing of examiner judgements in the first years of a new qualification to help carry forward standards.

38. Furthermore we know that in any education system, national outcomes change only slowly, even though variations may be quite significant at school level. In the context of English/English Language GCSE, we know that schools, teachers and the National Curriculum are all much as they were in preceding years. If there had been dramatically different outcomes in this year’s GCSEs as compared with last year’s, they would almost certainly have been attributable to the changes in the form of assessment rather than to any underlying improvement in young people’s knowledge of and ability to use English.

39. We believe the comparable outcomes approach was the best way of securing standards year on year in GCSE English in 2012. We know that the exam boards found it unusually difficult to set the standards in the units awarded before June 2012. Given that, it was essential that there was a robust approach to qualification level awarding in the summer, to avoid grade inflation, and comparable outcomes provided a way of doing this. The evidence we have looked at so far suggests that the awarding process in the summer was done appropriately, with the right accountabilities and appropriate engagement with the regulator. This is, though, something we will be looking at further over the coming weeks.

Maintaining standards in modular qualifications

40. The comparable outcomes approach is not specific to modular qualifications, but it helps to manage the specific challenges that arise with grading modular qualifications. These challenges are discussed in some detail in the 2009 paper by Isabel Nisbet, Ofqual’s previous chief executive, which has been highlighted in the media in recent weeks.

41. Qualification awards are made by adding up the awards made at unit level.

There are no awarding decisions at qualification level distinct from the cumulative awarding decisions at unit level. That means it is critical that unit awarding is done consistently and accurately.

42. There are two issues in particular that need to be managed. The first is that, if the full benefits of a modular system are to be realised, then students need to know how they are performing in each assessment as the course progresses, so they know where to focus their efforts. This means that exam boards need to award grades whenever a unit can be taken (up to four times during a GCSE course), even though some of the earlier awards may be made on the basis of only limited information about expected performance, and with only small numbers of entries and a possibly unrepresentative cohort. This means that many of the checks on examiner judgement which would typically be available when awarding grades are not available. When qualifications come to be awarded at the end of the course, if the statistics lead examiners to revisit their judgements about where the standard should be set (as would often happen particularly with a new qualification) there is no opportunity to revisit the standards for units already awarded.

43. The second problem with a modular system is that different choices can be made about the routes through the qualification which can affect student performance. As a result, the students taking each unit can be a mix of school years and the ability profile of the entry can vary from one exam series to the next. We know this makes it more difficult for examiners to judge where the standard should be in terms of the quality of work, particularly in a subject like English where maturity is a factor. For example, students who take the qualification on a modular basis – taking and retaking units – will tend to do better than students who do it on a linear basis, only taking assessments once, at the end of the course. That would be the case even without the generous awarding decisions that were made for some units in January 2012 and before. Given this point, the planned additional resit opportunity in November is appropriate, because it will go some way to levelling the playing field between ‘modular’ and ‘linear’ candidates by giving linear candidates a further assessment opportunity.

44. There is a level of transparency in the GCSE system now which means that schools and colleges are generally in a position to make informed judgements about the most appropriate way of teaching and structuring the qualifications. Information about how awarding is done, the standards that are set, and the assessments that are offered is routinely available to schools, and – because of the pressures of the accountability system – we know that schools focus very heavily on understanding what they need to do to secure the grades their pupils deserve. In general, modular awarding at GCSE has been successful – for most subjects, new GCSEs were awarded in 2011 and this happened successfully; and in 2012, new GCSEs in subjects other than English were awarded without difficulty. But we recognise the significance of the issues with GCSE English this summer: the generosity of the awards in assessments taken before June 2012 has had a serious impact on perceptions of fairness at qualification level. If we were not already planning to remove modular GCSEs after the current school year, we think there would now be a strong case for doing so.

45. Some commentators have suggested that the June awarding must have been

harsh to balance the generosity of the January awarding. Had that been the case, this would have further disadvantaged candidates who had taken the qualification on a linear basis. However, we have no evidence that this happened. If it had done so, we would have expected to see two things:

- Examiner reports at the time of June awarding consistently expressing concerns that the statistics were pointing to boundaries that were higher than examiner judgements would suggest they should be. In fact, as set out in our 31 August report, there is no evidence of that; as would be expected in a normal results season, there were a range of different boundary-setting decisions.
- Data suggesting that like-for-like outcomes were identical to the previous year. In fact, there were small (but within tolerance) increases in like-for-like outcomes in English (though the overall result was down as a result of changes to the cohort).

Numbers affected

46. It is very difficult to quantify the numbers of students affected by the issues with GCSE English awarding this year, because there are many reasons why candidates do not perform as expected, and more generally it is difficult to separate cause and effect.

47. There will also have been some students who got a D grade when they might have been expected to have got a C. As set out above, there are a number of reasons why this might have been the case, and it is impossible to disentangle them and estimate the numbers involved:

- Because they chose not to take assessments as they went along, so they could not resit units where they had done less well than expected.
- Because they took all their assessments in June, and therefore did not benefit from the generous grading in earlier assessments.
- Because schools made over-optimistic predictions about their attainment, perhaps based on incorrect assumptions that the January grade boundaries would be carried forward.

Variations at centre level

48. One of the issues that we will be exploring over the coming weeks is that there was an unusual level of turbulence in the results in GCSE English this year. Although year-on-year the results were very similar – as would be expected given our comparable outcomes approach – this masks significant variability, with some schools and colleges seeing results that were lower than the previous year's, and lower than expected, and others seeing results that were higher. We need to understand this better and consider what it tells us about how the system is used and managed. It may be that some turbulence is inevitable with a new qualification – some schools will be better prepared for a new qualification, or will use approaches which suit some qualification types more than others, or make better choices about

how to manage the assessments (for example, the timing of assessments). So we need to understand whether the levels of turbulence, and the reasons for it, are unusual for GCSE English this year. That will require analysis of results at centre level which we are now embarking on, and we will report further in our final report.