



Llywodraeth Cymru
Welsh Government

GCSE English Language 2012
An investigation into the outcomes
for candidates in Wales

Qualifications and Learning Division
10 September 2012

GCSE English Language 2012

An investigation into the outcomes for candidates in Wales by the Qualifications and Learning Division of the Welsh Government

Introduction

1. On 23 August 2012, following the publication of provisional GCSE results, the Minister for Education and Skills, Leighton Andrews, asked officials to conduct an investigation into the performance in GCSE English Language of candidates in Wales. The investigation was commissioned in the light of concerns about:
 - i) the significant fall in outcomes at grades A* to C for candidates in Wales, compared with the outcomes in 2011;
 - ii) significant variations in performance between schools and colleges in Wales, with a few centres experiencing significant improvements and many experiencing significant falls in outcomes when compared with 2011; and
 - iii) significant changes to grade boundaries for some units between January and June, by some awarding organisations.
2. This report has been prepared by the Qualifications and Learning Division of the Department for Education and Skills, which, on behalf of the Welsh Ministers, undertakes the functions relating to qualifications in Wales as set out in Section 30 of the Education Act 1997.
3. In conducting this investigation, officials have drawn upon:
 - extensive data provided by awarding organisations
 - data provided by other Divisions within the Welsh Government
 - discussions with awarding organisations
 - discussions and records of meetings with the regulators of qualifications regulators in England (Ofqual) and Northern Ireland (CCEA)
 - the report of the regulators of qualifications in England (Ofqual): GCSE Awards 2012 – a Regulatory Report (published on 31 August)
 - discussions with and evidence submitted by schools and colleges in Wales.

Summary of Findings

4. Having considered the available evidence, officials have reached the conclusion that, in relation to the concerns listed in paragraph 1 above:
 - i) The methodology for determining the grade boundaries for candidates in Wales did not deliver comparable outcomes for the 2012 cohort when compared with the 2011 cohort and that some candidates, particularly those close to the C/D boundary, are likely

to have been disadvantaged when compared with their peers in 2011. Recommended action to reduce this disadvantage is set out in paragraph 81 below.

- ii) The range of variability in centres' year-on-year outcomes at grades A* to C in 2012 was greater than in 2011. It is also clearly the case that the lower overall outcomes than expected will have contributed to this issue. However, it is also likely to have some connection with the very high proportion of controlled assessment in the new English Language specifications and possibly to different approaches to controlled assessment across centres. This particular issue will need further investigation over the coming weeks and action to carry out this work is recommended in paragraph 82 below.
- iii) Grade boundaries for some GCSE English Language units taken by candidates in Wales in June 2012 were *more different* to those taken by candidates in January 2012 than has been the case in previous years. This was most significant for some of the 5% of candidates in Wales who took their awards with AQA. AQA candidates who submitted controlled assessment tasks in June may have received a lower unit grade outcome than if they had submitted the same task in January and, in addition, AQA's Foundation Tier examination paper required ten more marks for a C grade in June than in January. WJEC did not offer controlled assessment units in January so that particular aspect of inconsistency was not an issue for their candidates – who represent 95% of candidates in Wales. There were, however, some slightly larger differences than usual between the marks required for a C grade in some of WJEC's examination papers in June and in January. These issues are explored further in paragraphs 53 to 56 below.

5. Other significant issues raised in this report include:

- i) the extent to which the awarding of qualifications in Wales is currently bound to, and determined by, policies and methodologies designed to be fit for purpose in England – and the challenges this presents to a devolved education system;
- ii) the fitness for purpose of the current suite of GCSE English Language qualifications; and
- iii) the need to give further consideration, in the development of future qualifications policies for Wales, to the relative merits and methods of norm-referencing style approaches to grading compared with more criteria-referenced approaches to grading.

GCSE English Language – overall outcomes in Wales

- 6. In Wales, 35331 candidates were entered for GCSE English Language in June 2012. Of these, 95% were entered with WJEC

and 5% were entered with AQA. No candidates in Wales were entered for GCSE English Language with Edexcel/Pearson, OCR or CCEA, the other awarding organisations which are able to offer this qualification to centres in Wales. In investigating the issues relating to outcomes, officials have therefore focused on data provided by WJEC and, to a slightly lesser extent, AQA.

7. GCSE English Language is a high entry, high stakes qualification. The vast majority of 16 year olds in Wales are entered for this qualification and a pass at Grade C and above is set down by most providers as a pre-requisite for progression onto many post-16 qualifications (such as A levels), apprenticeships and other employment opportunities. It is also an entry requirement for many degree and professional programmes. Furthermore, outcomes for GCSE English Language at grade C and above contribute to performance measures and as an indicator for banding for schools in Wales. It is essential, therefore, to learners and to providers that the outcomes for GCSE English Language are accurate, appropriate and comparable over time.
8. The percentage of candidates in Wales achieving GCSE English Language at grade C and above in 2012 was 57.4%. This was down 3.9 percentage points from 61.3% in 2011 and, when seen in the light of outcomes over a five year period, represents a significant and unexpected fall (see Figure 1 below).

Figure 1: Overall outcomes at Grade C and above for GCSE English Language¹ in Wales 2008 – 2012 : Source JCQ

Examination Series	% of candidates in Wales with Grades C to A*
June 2008	62.4%
June 2009	60.4%
June 2010	61.6%
June 2011	61.3%
June 2012	57.4%

9. It is relevant, also, to consider whether there have been any significant changes in the cohort in Wales between 2011 and 2012. We therefore asked WJEC, who had 95% of the candidature in Wales, to provide us with information on the numbers of entries and the grade distributions for learners who were aged 16 on 31st August in 2011 and 2012. This data (see Figure 2 below) showed us that there were very similar numbers of 16 year olds entered for

¹ Prior to the 2012 awards, the qualification which was replaced by GCSE English Language was called GCSE English. From the 2012 awards onwards, GCSE English is a different qualification and co-exists in England with GCSE English Language. This is explained later in the report.

GCSE English Language in both years, indicating a stable cohort at this age. It also showed that 16 year old candidates had a similar, though slightly greater, fall in outcomes at grade C and above when compared to the all-age Wales cohort, with 16-year olds seeing a fall of 4.1 percentage points between 2011 and 2012.

Figure 2: GCSE English Language candidates in Wales aged 16 as at 31st August in each relevant year. Source: WJEC

Year	Entry – 16 yr olds	% at Grade C and above
2011	30830	63.1
2012	30933	59

10. We also looked at whether there had been a shift in the type of centre which candidates for GCSE English Language in Wales attended. Looking again at data provided by WJEC, we found that there had been a slight shift in the proportion of candidates from each type of centre. This showed that, in Wales, for WJEC's specification, there had been a small reduction in the percentage of entries from maintained schools (down from 92.4% to 91.9%), small increases in the proportion of entries from FE and sixth form colleges (up from 4.2% to 4.8%) and from independent schools (up from 2.1% to 2.5%).² There does not appear to be any indication from this data that the balance of centre types should have any significant direct consequence upon outcomes at C and above.

Introduction of new GCSE English Language qualifications

11. Across Wales, England and Northern Ireland a new suite of GCSEs in subjects relating to English were introduced for teaching from September 2010. These replaced the former GCSEs in "English" and "English Literature". While direct replacements for GCSE English and GCSE English Literature were developed, it was also proposed that a combined GCSE, simply titled GCSE English and combining elements of English Language and English Literature, should be offered to enable candidates who had less of an interest and or aptitude in English Literature to meet the requirements of their national curriculum programme of study while only completing one GCSE in the subject area instead of two.
12. In Wales, this proposal was not accepted as it was judged that GCSE English alone would not enable learners to cover the full national curriculum programme of study for English Language or for English Literature in Wales. GCSE English is, therefore, not approved for use in schools in Wales and learners in maintained

² These sets of figures do not add up to 100% as a small proportion of entries are categorised as 'other'.

schools in Wales are required to cover the fuller programme of study in English Language that GCSE English Language promotes. The programme of study also requires that learners in Wales should study elements of literature and the majority of learners in Wales, therefore, also take GCSE English Literature.

13. In England the proposal to offer a reduced, combined assessment model was accepted and learners in England are able to study either for GCSE English (alone) or GCSE English Language and GCSE English Literature. It is likely that this policy difference in England has had some impact on the determination of outcomes for GCSE English Language in Wales as suggested in paragraphs 34 to 36 below.
14. The new GCSE English Language specifications which were introduced for teaching from September 2010, and which were awarded for the first time this summer, had two particularly significant features which differentiated them from the previous GCSE English specifications. These were unitisation and controlled assessment.

Unitised specifications

15. With unitised specifications, candidates are able to enter for units prior to the end of the course and are awarded a grade for those units which can subsequently contribute to the overall qualification grade. This enables candidates to 'bank' units prior to the end of the course so that their GCSE outcome need not be entirely dependent upon assessment at the end of two years of study. If candidates are unhappy with their unit grade, they are permitted to retake any unit, once only, to try to improve their grade. In addition, candidates must always take a minimum of 40% of the qualification at the end of their course.
16. In England, the Westminster Government and subsequently Ofqual, announced in 2011 that candidates in England would no longer be able to take unitised GCSEs. Therefore, for courses beginning in September 2012 onwards, all units must be submitted and awarded at the end of the course and no retaking of units will be permitted. The Welsh Government has not implemented this policy and the issue is being considered by the Review of Qualifications 14-19 in Wales. In the meantime, until at least September 2014, centres in Wales may choose to enter learners for unitised specifications. For the qualifications awarded this summer, candidates in both Wales and England were able to take some of their units early and to retake one or more of these units (once only) to try to improve their grade.
17. Data received from WJEC about their candidates shows that candidates who took one or more units before the June 2012

examination series (ie in a unitised way) were substantially more likely to have gained an overall A* to C than those candidates who took all of their units for the first time in June (ie in a linear way).

Figure 3: Impact of taking one or more units before June 2012 on final outcomes for candidates in Wales for WJEC's GCSE English Language.
Source: WJEC

	Candidates in Wales who took one or more units before June 2012.	Candidates in Wales who took all of their units in June 2012
% achieving grade C or above	64%	52.8%
% achieving grade A or above	13.4%	11.1%

18. The same data indicates that there are significant differences in the proportion of candidates taking units early in Wales and in England. Well over two thirds of WJEC's candidates for GCSE English or English Language in England (69.2%) took one or more units before June 2012 whereas well under half (42%) of candidates in Wales took one or more units early.

Figure 4: Percentage of WJEC GCSE English/English Language candidates in Wales and England who took one or more units before June 2012.
Source: WJEC

	GCSE English/English Language candidates in England who have taken one or more units early, expressed as a percentage of all candidates in England	GCSE English Language candidates in Wales who have taken one or more units early expressed as a percentage of all candidates in Wales.
Foundation tier	25.3%	17.4%
Higher tier	35.4%	21.1%
Mixed tier ³	8.5%	3.5%

19. It therefore appears likely that some centres in Wales, due to not having entered as many candidates for units earlier than June as did centres in England, have unknowingly rendered their candidates more likely to receive lower grades than those who took one or more units early. There are likely to be two main reasons for these differences in attainment levels between the two entry patterns. The first is that the January 2012 grade boundaries were set lower

³ Mixed tier candidates will have had grades taken into account for one Foundation tier unit and one Higher tier unit. This might, for example, have occurred where candidates took a unit early and achieved a grade C at Foundation tier, then retook the unit at Higher tier in the hope of gaining a B or above but instead gained a grade D: in this case the C from the Foundation tier paper would count in their final grade.

than the June 2012 boundaries (as discussed later in this report); the second is that unitised specifications delivered the benefits for which they were initially designed – by enabling candidates to be assessed in stages rather than leaving all the assessment to the end of the course. In either case, this significant difference between centres' approaches to entry patterns in Wales and England may be a significant factor in the failure to maintain comparable outcomes in Wales using the methodology for reviewing potential GCSE outcomes against cohort predictions based on Key Stage 2 outcomes in England. WJEC were required to follow this methodology due to the majority of their candidates for GCSE English and English Language being candidates in England. The methodology is discussed in paragraphs 24 to 33 below.

Controlled Assessment

20. Controlled Assessment was introduced into revised GCSEs which have been taught from September 2009 for non-core subjects; from September 2010 for English Language, English Literature, Welsh Language, Welsh Literature and ICT; and from September 2011 for the Sciences. Controlled assessments replaced coursework and aimed to encourage a more integrated approach to teaching, learning and assessment, and to enable teachers to confirm that students had carried out the work involved.
21. In 2011, in the light of significant concerns emanating from the teaching profession, the regulators in England, Wales and Northern Ireland commissioned an investigation into the introduction of the first round of controlled assessment in those GCSEs that had been introduced for teaching from September 2009. It was clear from this report ⁴that there was considerable variety between centres in the effectiveness with which controlled assessment had been introduced. While many teachers could see some of the benefits of controlled assessment in reducing some of the security concerns relating to coursework, many others raised concerns about, for example, the limitations of guidance from awarding organisations, potential inconsistencies across centres in the application of the control of tasks, and the impact on teaching and learning time.
22. While the moderation by awarding organisations of internally assessed controlled assessments is designed to ensure that candidates' work is fairly and consistently assessed and marked, it is difficult to be confident that the introduction of this new form of assessment has been implemented equally effectively across all centres. For example, candidates are able to take a limited amount of notes into their assessment and there is the potential for a different amount and quality of support to be provided to learners in

4

<http://wales.gov.uk/topics/educationandskills/qualificationsinwales/qualificationregulation/regpublication/reports/assessment/?lang=en>

different centres in the preparation of these notes. Indeed, the advice of one of the independent scrutineers, with whom officials contracted to conduct the initial part of our scrutiny review of WJEC English Language⁵, while commending the awarding organisation's attempts to manage the challenges of moderating the assessment of candidates' work, expressed in clear terms the risks presented by the inclusion of controlled assessment in this qualification – particularly when it accounts for 60% of the final mark.

“[An] area of concern is the actual mechanics of Controlled Assessment. Not only is this an organisational problem for many schools but there is clearly no way that the awarding organisation can police whether Controlled Assessment is being carried out according to the regulations”

Source: Independent subject specialist reporting to the Welsh Government

Related concerns were raised by officers and moderators at AQA as referred to in paragraphs 67-68 below.

23. Given the potential for variation in the quality of preparedness of candidates across centres, the introduction of controlled assessment contributing 60% of the marks of such a high entry, high stakes qualification as GCSE English Language, where consistency and comparability across Wales is of crucial importance, was perhaps questionable. The Review of Qualifications (14-19) in Wales is giving consideration to the future of controlled assessment in qualifications, especially GCSEs and to the future assessment of literacy and numeracy.

Maintaining Standards when Qualifications Change

24. One of the central factors in considering the reason for the fall in outcomes for GCSE English Language in Wales is the introduction in recent years of a methodology that is designed to ensure that 'standards' in qualifications are maintained from one year to the next and across awarding organisations. This particular programme of work was established in the autumn of 2008 at the time that schools and colleges commenced teaching the current suite of A level qualifications. There had been significant changes to the structure of these A levels which had moved, for the most part, from six modules to four. Regulators and awarding organisations were mindful of the fact that when the *previous* generation of A levels had first been introduced, there had been significant increases in outcomes in the first year of awarding and that this had generated much concern both in the media and in the education sector.
25. At the same time, it was noted that candidates and teachers undertaking a new specification are faced with a different order of

⁵ See paragraphs 72 to 74 below for an explanation of the scrutiny programme.

challenge from those taking a well-established specification with the benefit of past papers, examiners' reports and teacher experience – and that this should not have a negative impact on the chances of learners in this first cohort achieving the grades they deserve.

26. For A levels, the methodology used to determine whether a set of outcomes is in line with the outcomes of the previous cohort, is based on the relationship between the previous cohort's GCSE results and their A level outcomes. This relationship is tabulated to generate the most common outcome at A level for any given 'score' in GCSEs. This relationship is then applied to the actual GCSE 'scores' of those candidates entered for the current year's A levels for whom GCSE results are known, in order to generate a predicted percentage of these 'matched' candidates at each grade for each specific A level. It is important to note that the methodology is applied at a cohort level, rather than at the level of the individual learner. This methodology is complicated – but it has been demonstrated to have value by the findings of an independent research project commissioned by the regulators in England, Wales and Northern Ireland.⁶
27. Although initiated by the regulator in England (Ofqual), Welsh Government officials were engaged with, and contributed to the refinement of, this methodology for determining expected outcomes for A levels. In general, and with some reservations, this methodology is applicable and relevant for A levels across England and Wales, while GCSE outcomes remain reliable and consistent across the two nations.
28. Whereas Wales and England have a common qualification (GCSE) which is generally taken by learners prior to A levels, there is no such common qualification taken by learners in both nations prior to GCSE. In England, learners who are currently aged 16 took externally set and marked tests at the end of Key Stage 2, the data from which is accessible to awarding organisations. These tests were not taken by learners in Wales, who, instead, at the end of Foundation Phase, Key Stages 2 and 3, were assessed by statutory teacher assessments which were locally moderated and reported to the Welsh Government. The awarding organisations do not have access to this data.
29. With the introduction of significantly changed GCSEs, and in the light of the relative success of the predictor methodology in determining and considering comparable outcomes for A levels, the qualifications regulators were keen to identify whether any similar methodology could be employed to review GCSE outcomes in order

⁶ The NFER report is available at <http://wales.gov.uk/topics/educationandskills/qualificationsinwales/qualificationregulation/regpublication/reports/gcses/?lang=en>

to maintain outcomes for these qualifications. When one of the awarding organisations suggested in 2009 that Key Stage 2 outcomes might be used to predict GCSE outcomes there was considerable cynicism. Welsh Government officials were clear, from the outset, that such a methodology could not, and should not, be used in relation to candidates in Wales. Following some initial modelling and extensive discussions, Ofqual and the awarding organisations in England agreed that the Key Stage 2 model was the best methodology available for indicating any given cohort's expected performance at GCSE. Welsh Government officials continued to reject the use of this methodology for awarding decisions where substantial numbers of candidates from Welsh centres were involved.

30. As a consequence of these discussions, it was agreed that, for the awarding of the June 2011 examination series, which saw the first awards of the new GCSEs in non-core subjects, the awarding organisations based in England would use Key Stage 2 data in determining expected outcomes while WJEC (who had the majority of candidates in Wales) and CCEA (who had the majority of candidates in Northern Ireland) would not use Key Stage 2 data but would seek to maintain comparable outcomes using, in the main, a methodology which looked at predicted outcomes for 'common centres' – a methodology that draws on data from centres with entries in the same subject (with any awarding organisation) in the current and previous years. The common centres methodology is based on the assumptions that over time GCSE outcomes are stable and that, while there may be changes at centre level, these are evened out for the whole cohort.
31. In the awarding of these new non-core GCSEs in 2011, no significant issues arose in the outcomes for Wales. The 'common centre' predictors were of some use in maintaining comparable outcomes, but in some cases they produced unconvincing results and it was agreed between the regulators in Wales and England, that a 'common sense' approach to considering both the common centre predictions and the year-on-year overall outcomes should be adopted.
32. Early in 2012, Ofqual officials alerted Welsh Government officials that they were minded, in the light of WJEC having substantial numbers of entries from candidates in England for some of their core subject GCSEs, to require WJEC to employ the use of Key Stage 2 predictors to determine the expected outcomes for WJEC's specifications. Welsh Government officials expressed serious concerns that it was not appropriate for results for Welsh candidates to be determined on the basis of prior achievement by candidates in England. Nevertheless, when it became clear that a significant majority of WJEC's candidature for GCSE English Language was from England, and that this represented the second

largest awarding organisation entry for candidates from England, it was hard to justify continued resistance to the use of the methodology. Simultaneously, WJEC also suggested that they would be prepared, with some reservations, to use the Key Stage 2 methodology for the awarding of GCSE English and English Language. It was therefore agreed that, for these subjects only, WJEC would report their projected outcomes against the Key Stage 2-based predictions.

33. It should be stressed that at no point in these early discussions was it apparent that this would have any significant impact on the outcomes for candidates in Wales.

WJEC GCSE English Language Provisional Results, Summer 2012

34. Towards the end of July 2012, when marks and awards were being prepared by awarding organisations, it became clear that the split of the candidature in England between GCSE English and GCSE English Language appeared to be causing difficulties with the predictor model. As many centres in England appeared to have entered their lower ability candidates for the new GCSE English specification, the entry profile of candidates for GCSE English Language was significantly different in England to that for its direct predecessor, the former GCSE English. As a consequence, there was a need to adjust the predictions for these two qualifications to take account of the different candidate profiles in England.
35. A 'fix' was proposed and accepted by awarding organisations which was based on the assumption that, in the main, the 'typical' candidates for GCSE English would have been those who previously would have only taken GCSE English Language and not English Literature, whereas the 'typical' candidate for GCSE English Language, in England, would have previously taken both. Two sets of prediction matrices were accordingly compiled and awarding organisations were instructed to achieve outcomes that were within 3% of the predictions for each of the separate predictions and within 1% of the overall predictions for the combined cohort for both GCSE English and English Language.
36. Given that all candidates in Wales previously took GCSE English and this year took GCSE English Language (and given that the number of 16 year olds taking the qualification remained stable in Wales) it is clear that the position in Wales is much more stable and therefore comparable, year on year, than the position in England where the cohort has been split between GCSE English and GCSE English Language. It is unfortunate, therefore, that sufficient consideration was not given, in the determination of WJEC's outcomes, to maintaining comparable outcomes for the Wales element of WJEC's cohort.

37. The first indication that awarding processes adopted could have an impact on candidate outcomes for Wales began to emerge late on Friday 27th July following the meeting of the awarding committee at WJEC. Officials in DfES were alerted that results in GCSE English Language for Wales may be below those in 2011 and on Monday 30th July, the Chief Executive Officer of WJEC gave further notice to the Welsh Government that the outcomes for Wales were likely to be significantly down from the previous year.
38. In the following two weeks, officials from the Welsh Government engaged with WJEC and with Ofqual to attempt to reach an agreement on the final agreed outcomes for GCSE English Language. While taking account of the guidelines for using the predictors, WJEC had been unable to confirm outcomes which fell within tolerance of the predictors that had been based on Key Stage 2 performance in England. Although the outcomes from the awarding meeting already represented a significant fall in outcomes for the Wales cohort (a likely overall fall of 3.4 percentage points), Ofqual officials were not prepared to accept these outcomes as they regarded these as too 'generous' in comparison with the Key Stage 2 predictions.
39. Following extensive correspondence and discussions, at a time when there was considerable pressure to agree final outcomes in order for candidates to receive their results on time, one of three options provided by WJEC of outcomes which came closer to the Key Stage 2 predictions for England was accepted by Ofqual. Regulatory officials in Wales continued to express strong reservations about the methodology, but at this late stage, recognising the need to reach agreement, as this option produced the least damaging impact on outcomes for Wales, Welsh Government officials reluctantly agreed to accept this amendment to the outcomes which resulted in an overall 3.9 percentage point fall in outcomes for Wales. We address this issue further in paragraphs 79 and 80.
40. Following the announcement by the Minister for Education and Skills that there would be an investigation into the outcomes for GCSE English Language in Wales, officials have given further detailed consideration to the data relating to WJEC's potential and actual outcomes and to a range of supporting evidence.

Further Consideration of WJEC's outcomes

41. WJEC's original proposed outcomes were determined in the meeting of examiners that was held on 27th July. This meeting was observed by another independent subject specialist, contracted by the Welsh Government as part of its scrutiny programme. The observer reported that "*appropriate statistical evidence was combined with examiners' judgements*" in determining the award

and that *“the entire Awarding process was underpinned by the expectation that work considered at a certain boundary would on balance demonstrate the knowledge, understanding and skills set out in the grade descriptions.”* We deduce from these, and from other comments in our observer’s report, that the awarding meeting was fairly and appropriately conducted.

42. As stated in paragraph 38 above, the outcomes recommended by WJEC’s awarding committee resulted in outcomes which would have led to a 3.4 percentage point fall in outcomes for Wales at grade C and above (compared with the final 3.9 percentage point fall that was implemented at the request of Ofqual). However, WJEC has provided us with data that shows the significant effect that would be brought about by adjusting downward, by one mark, the grade boundary recommended for *one* of the controlled assessment units. Such a slight amendment would result in an increase of 2.4 percentage points in the GCSE English Language outcomes for Wales at grade C and above. The overall outcomes at grades C to A* would then be increased to 59.8% which is only 1.5 percentage points below the 2011 outcomes for Wales (compared with 57.4% with the currently published provisional outcomes). This outcome would be closer to the current year-on-year fall in overall outcomes in England, though it would, in itself, if adopted for candidates from England as well as from Wales, reduce further the fall in England.
43. Welsh Government officials recommended this option for adjustment to Ofqual officials on Thursday 30th August but Ofqual rejected the recommendation on the grounds that it would place WJEC’s outcomes for England too far out of tolerance of the Key Stage 2 predictions.
44. While respecting the view of the regulator in England (Ofqual) that the Key Stage 2 predictor model may be useful in comparing the outcomes of candidates in England, Welsh Government officials have concerns about the level of confidence that can be placed in the degree of precision that this model can provide in determining awarding outcomes. Furthermore, officials have very serious concerns, as outlined above) about using the model where a cohort has significant entries from Wales and especially where those entries comprise a high proportion of the Wales national cohort (in this case 95%).
45. While externally assessed Key Stage 2 data is not available for candidates in Wales, the Welsh Government does hold teacher assessment results for English for the 2012 cohort and its predecessors at the former Key Stage 1, at Key Stage 2 and at Key Stage 3. This data (Figure 5 below) shows that, when compared with its predecessor cohort, the 2012 GCSE cohort was performing slightly below the 2011 GCSE cohort at Key Stage 1, on a par with

the 2011 GCSE cohort at Key Stage 2, and slightly above the 2011 cohort at Key Stage 3. This indicates that, had a similar prediction model been devised for Wales candidates to that used in England, they may have been expected to achieve a similar outcome for GCSE English as their immediate predecessors. Indeed, if a prediction model had been based on the Key Stage 3 outcomes, the GCSE English Language outcomes in 2012 would have been expected to surpass the outcomes in 2011. We are not suggesting that these data should have been used as an alternative methodology, however we believe that the data is relevant in cross-checking the reliability of the methodology used for candidates in England.

Figure 5: Percentage of the Wales cohort achieving the expected level of attainment in teacher assessments at Key Stages 1 to 3 (and in GCSE at Key Stage 4) over three years. Source: Welsh Government

Subject	Year GCSE taken	KS1	KS2	KS3	KS4⁷
English	2012 cohort	82.3	78.6	72.5	57.4
	2011 cohort	83.0	78.6	70.6	61.3
	2010 cohort	83.4	79.3	69.5	61.6

46. In addition to these indicators of prior attainment, we asked WJEC to calculate the predicted outcomes for the Wales cohort only, based on the “common centre” benchmark model that is used for the remainder of WJEC’s GCSE specifications (as referred to in paragraph 30 above). This model (Figure 6 below) includes the vast majority of WJEC’s candidates in Wales and illustrates a significant variation between the “common centre” prediction for candidates in Wales and their actual outcomes. If this predictive model had been used for Wales candidates alone, it would have resulted in a significant increase in outcomes at Grade C and above for candidates in Wales.

⁷ percentage of candidates achieving A*-C in GCSE English (2010 and 2011) and GCSE English language (2012)

Figure 6: Comparison of predicted outcomes based on ‘common centre’ methodology with actual outcomes for WJEC’s GCSE English Language candidates in Wales in 2012. Source: WJEC

GCSE English Language:

Wales Common Centre predictions compared with outcomes in 2012

	Cumulative Prediction	Cumulative Results	Difference
A*	3.6	3.0	-0.6
A	14.3	12.8	-1.6
B	35.0	31.1	-3.9
C	64.6	59.7	-4.9
D	84.8	82.6	-2.1
E	94.9	93.5	-1.4
F	98.7	98.3	-0.5
G	99.9	99.7	-0.1
No. of Cands	30,247	30,247	

47. As has been identified in paragraphs 15 to 19 above, the introduction of opportunities to take units early may have significantly increased candidates’ likelihood of achieving a grade C or above. Candidates in England who have taken one or more units early have achieved over 10.3 percentage points higher outcomes than candidates who took all units at the end of the course. Candidates in England have taken up the opportunity to take units early significantly more frequently than candidates in Wales. Given that there were no opportunities to take units early in the previous specifications for GCSE English Language, it would appear that over two thirds of the candidature in England may have increased their potential outcomes at grade C and above by some ten percentage points due to the ability to take units early when compared with the candidates in 2011 (and 2010 which is the comparator year for the England predictor model) who were not able to take any units early. Whether it was due entirely to different grade boundaries, or also to the benefits of staggered assessment, it seems likely that the significant difference between entry patterns in 2010 and 2012 could have had an impact on the reliability of the Key Stage 2 predictor matrix, for candidates in England as a whole. In attempting to maintain a stable relationship between Key Stage 2 outcomes and GCSE outcomes, the methodology attempts to smooth out any other factors which may change this relationship. Where these factors have impacted differently upon candidates in England and in Wales, then the ‘smoothing’ carries increased risks of distortion.

48. This issue highlights a fundamental tension between the aims of maintaining comparable outcomes while maintaining comparable 'standards'. If the ability to take units early is helpful to learners in promoting improvements in their performance, resulting in improved chances of success, consideration needs to be given as to whether these improved chances should be reflected in improved national outcomes – or whether similar outcomes to previous cohorts without the ability to take units early should be the desired aim. In potentially suppressing the improvements in the outcomes for the English cohort brought about by the ability to take units early, through the KS2 predictor model, there is a strong chance that the Wales cohort has been relatively disadvantaged. This tension will need further consideration in the coming months and in the development of any future policies on the grading of new qualifications.
49. In trying to juggle to fit the outcomes for candidates in England to a prediction model based on a different entry profile and different entry patterns, it seems possible that a serious distortion to the outcomes of the candidates in Wales has been caused. Certainly the provisional outcomes for Wales, published on 23 August, contradict one of the early principles that was stated by Ofqual in the minutes of one of the first Maintaining Standards meetings which was that:
- “on a national level, overall there is no reason to believe that outcomes in terms of grade distribution in the first year should be very different from those before the changes”* Source: minutes of Maintaining Standards meeting published by Ofqual in October 2008.
50. A consideration of the technical and qualitative data available strongly suggests that, in the context of maintaining 'comparable outcomes', the 2012 GCSE English Language outcomes for candidates in Wales, at grades C and above, would have been expected to be at least as high as, or even higher than, the equivalent outcomes in 2011. In summary:
- a. the cohort in Wales in 2012 is very similar in size and centre type to the cohort in 2011;
 - b. the 2012 cohort in Wales performed as well as the 2011 cohort in Key Stage 2 English teacher assessments and 1.9 percentage points better than the 2011 cohort in Key Stage 3 English teacher assessments;
 - c. the “common centres” predictor model, which is the preferred predictor model for Wales, indicated an expected outcome for candidates which was 4.9 percentage points higher than the actual published outcomes;
 - d. the effect and likelihood of candidates in England taking units early does not appear to be taken into account in the England predictor model using Key Stage 2 outcomes in England; and

- e. very minor amendments to the grade boundaries determined at awarding – a process which, in itself, was influenced by the requirements to keep within or very close to predictions - would result in significantly improved outcomes in Wales which would be much closer in line with the implications of this evidence.
51. In reviewing the outcomes for candidates in Wales, it is the considered view of regulatory officials that the 3.9 percentage point fall in outcomes for GCSE English Language at grades C and above in 2012, when compared with 2011, as announced in the provisional results for GCSE English Language, is unjustifiable and almost certainly unfair to candidates.
52. Given that 95% of the candidates for GCSE English Language in Wales entered with WJEC, it is recommended that the Minister should require WJEC to re-award this qualification, adjusting grade boundaries to bring the outcomes for candidates in Wales as close as possible to the outcomes for GCSE English of the candidates in Wales in 2011 (see paragraph 81 below).

Variability of centre outcomes across Wales

53. In addition to the concerns outlined above about the overall outcomes for candidates in Wales, concerns have been raised about the wide variability of outcomes on a centre by centre basis, when compared with previous years. WJEC and AQA have provided us with data that shows that, in 2012, a total of 85 centres in Wales had centre outcomes at grade C and above for GCSE English Language which differed by more than 15 percentage points from their outcomes for GCSE English in 2011.
54. For 21 of these centres, their outcomes at grade C and above had improved by 15 percentage points or more, but 64 centres had seen a fall in outcomes of 15 percentage points or more. We asked WJEC to provide comparative data from 2011 and this did confirm that more centres in 2012 had variations of >15 percentage points than in 2011 (Figure 7 below). The increase in variability appears, from this data, to have been due to worsening outcomes for more centres, but it is interesting to note that despite the significant fall in national outcomes at grades C and above, some centres still managed to achieve significant improvements.

Figure 7: Number of WJEC centres with >15 percentage points change in outcomes at grade C and above for GCSE English Language and direction of change in 2011 and 2012. Source: WJEC

Year	No. of centres with >15 percentage points change	No. of these with improved outcomes	No. of these with lower outcomes
2011	51	21	30
2012	84	20	64

55. As has been explained in paragraphs 20 to 23 above, the introduction of controlled assessment into GCSE English Language, and the difficulties inherent in ensuring that centres conduct these assessments consistently, are unlikely to reduce variability between centres unless efforts are made to improve both the quality and consistency of assessment and the preparedness of candidates.
56. In the context of its ongoing work to improve literacy and numeracy outcomes in Wales and in response to both the fall in outcomes in 2012 at Key Stage 4, the Schools Standards Unit in the Department for Education and Skills has already planned a programme of data analysis and visits to schools to review the features of more successful and of less successful schools and to investigate, in particular, the delivery of controlled assessment. WJEC has also offered to conduct some further analysis of its data and of outcomes and it is recommended that the Schools Standards Unit liaises with WJEC in this area of work.

Changes to unit grade boundaries between January and June

57. The aspect of the awarding of GCSE English Language (and of GCSE English) which has raised most concern across the UK, was the fact that grade boundaries for some units, with some awarding organisations, changed between January and June 2012. As has been explained above candidates in Wales were only entered with two awarding organisations: WJEC (who had 95% of the Wales candidates) and AQA (who had 5% of the Wales candidates). While some other awarding organisations made changes to grade boundaries, these are not relevant to Wales and have therefore not been reviewed as part of this investigation.
58. With traditionally examined units, it is normal for the marks required to reach any grade boundary to vary between examination series. Despite the best efforts of examiners, exam papers can turn out to prove relatively easier or harder for candidates from one series to the next. Typically, and depending on a number of technical factors, including the overall number of marks available, the marks required for a unit to be awarded a C can vary by up to 3 marks in either direction.

59. On the other hand, marks for coursework (the predecessor for controlled assessment) have, historically, generally remained the same from one series to the next, as learners are completing similar tasks. Controlled assessments are different in some respects to coursework, but in January 2012 and June 2012, for AQA, candidates completed the same tasks. WJEC did not award any controlled assessment grades in January 2012.
60. In reviewing the C/D boundary marks for WJEC's examined papers for GCSE English Language, we can see that, between January and June, the boundary marks changed by 4 marks for two papers which is a slightly larger change than might be accepted without question.
61. We asked WJEC to provide an explanation for each of these four mark changes. For one WJEC Foundation tier paper, the boundary mark fell by four marks from the January paper, meaning that candidates needed to achieve fewer marks in June to gain a grade C than they needed to gain a grade C in January. WJEC explained to us that examiners felt, in January, that some of the questions had proved to be insufficiently demanding, especially the first question on the paper. Therefore, candidates needed to gain more marks than expected in order to demonstrate the requirements of a grade C. In June 2012, the examination paper met expectations more closely, and in fact the mark required to achieve a grade C in June 2012 was the same as had been required in June 2011 (when the paper was taken 'early' by some year 10 candidates). We believe this to be a satisfactory explanation.
62. WJEC's other 4 mark change in a C/D boundary mark between January and June was for one of the two Higher tier papers (Higher tier papers are taken by candidates wishing to access grades B and above). In this case, candidates in June 2012 had to achieve 4 more marks to achieve a unit outcome of grade C than candidates in January had needed to achieve. The June 2012 mark was, in this case, also 2 marks higher than the June 2011 mark. WJEC has explained that examiners in January 2012 had found that they needed to make allowances for one question that candidates had found more challenging than expected – and that this brought the grade boundary down, in January, to a point which would normally have been unacceptable. The issue for this unit appears to be more that the standard achieved by candidates in January was lower than might be expected, rather than that the candidates in June were disadvantaged. However, if a regrading of the award is required to take place, it may be appropriate for WJEC to give further consideration as to whether they are fully satisfied that the C/D boundary was set appropriately for this unit.
63. Turning to the grade boundary changes implemented by AQA, it can be seen in Figure 8 below that AQA implemented four changes

to the C/D boundary marks for GCSE English Language between January and June.

Figure 8: AQA Grade boundary marks in June and January 2012. Source: AQA

June 2012					
English Language (4705)					
	A	C	F		Max
ENG1F		53	19		80
ENG1H	55	44			80
ENG02	38	28	13		45
ENL03	64	46	18		60
January 2012					
English Language (4705)					
	A	C	D	F	Max
ENG1F		43		13	80
ENG1H	54	41	33		80
ENG02	37	25		11	45
ENL03	64	43		16	80

64. AQA's paper ENG1F is the foundation tier externally assessed examination paper, "Understanding and producing non-fiction texts", and the boundary mark set for grade C in June was 10 marks higher than the grade boundary set in January. Evidence provided to Welsh Government officials and reported by the regulator in England (Ofqual) demonstrates that AQA places a strong emphasis on statistical indicators in determining grade boundaries and that, in both January and June, examiners based their judgements around statistically recommended boundaries while giving due cognisance to the quality of candidates' work.
65. The Chair of examiner's report for ENG1F in June 2012 states that *"the Tier F agreed 'C' boundary, initially 52, was moved up to 53 on revisiting the tick charts in the light of further statistical information. Although this mark was significantly higher than the mark for the reference year it was felt that this higher mark was a truer reflection of the quality of candidates' work in relation to the 'C' grade criteria"* and the principal examiner's report for this unit claims that *"the overall demands of the paper were very similar to previous series"*. The implication of these comments, and the conclusion of Ofqual as the regulator in England, is that the boundary mark in January 2012 was set too low and that the boundary mark in June 2012 was the necessary boundary mark in order to meet the predicted outcomes (using the Key Stage 2 prediction methodology for candidates in England). Of the 1084 candidates from Wales who sat this paper in 2012, 488 took the January paper and the remainder took the June paper.
66. AQA's paper ENG1H in June 2012 required 3 more marks to achieve the C grade boundary than did the equivalent paper in

January 2012. This difference falls within the category of a 'normal' change in boundary marks between one series and the next.

67. Units ENG02 and ENL03 comprise the two controlled assessment units for AQA's GCSE English Language specification. For both these units, the boundary marks were 3 marks higher than they were in January 2012, even though candidates had completed the same task. According to Ofqual, AQA officers noted concerns, in January, that marks appeared to be high and that there was some evidence of over-marking by teachers. However, the grade boundaries that had been used previously in June 2011 were maintained in January 2012 as the January entry represented only a small subset of the wider cohort and it was not clear at that point in time as to whether higher grade boundaries would have a detrimental effect on AQA's ability to meet the predicted outcomes for the whole qualification in June 2012.
68. For ENG02, which is AQA's controlled assessment unit for Writing, there are indications of concern about the difficulties of arriving at an appropriate grade boundary in June 2012. Prior to the awarding meeting the principal moderator for this unit suggested that marks from the previous series should "*probably*" be carried forward and described this as the "*ideal scenario*". The principal moderator went on to comment that "*I have no overall sense of how vast numbers of teachers have been rewarding their students. Anecdotally I fear the raw statistical award would need to be high, but I have been working very hard with the AQA research team....to prepare for this award and its aftermath. I suspect there will need to be a judicious decision made on a host of factors, which will move the award upwards, but maybe not as high as it 'should' be.*" It is clear that the changes to the boundary mark, which rose by three marks between January and June 2012, were made in the interest of meeting the statistical predictions for candidates in England. This appears to have created a distinct inequality between candidates who received a unit award in January and those who were awarded for this unit in June and it is relevant to note that **none** of the candidates from Wales who took GCSE English Language with AQA were entered for unit ENG02 in January and all were therefore graded against the higher grade boundaries.
69. ENL03 is the Speaking and Listening controlled assessment unit and again here the mark required at the C/D boundary in June 2012 was three marks higher than in January 2012. The principal moderator's initial report for this unit in June 2012 simply recommended that the mark from January 2012 be carried forward. However, a subsequent report by the same principal moderator explicitly welcomed the fact that moderators are able to amend the grade boundaries for controlled assessment. The principal moderator goes on to comment that "most of the work seen at the

borderlines in controlled assessment by the Principal Moderator has been over-rewarded". There are other suggestions throughout the documentation received from AQA that there are significant reliability (and presumably variability issues) with teachers' marking of controlled assessment. Nevertheless, it remains the fact that for this unit too, there is an inequality in the awarding between candidates who took this controlled assessment unit in January and those who took it in June. Of the 1890 candidates from Wales who took this unit in 2012, only 200 received their unit awards in January, with the majority being awarded against the higher grade boundaries in June.

70. In conclusion, with regard to the determination of grades by AQA between January and June, it is clear that AQA set their grade boundaries in line with the requirements set down by the regulators that Key Stage 2 predictions for England should be used to inform a statistically based decision.
71. While the arguments against the use of the Key Stage 2 predictor methodology for candidates in Wales, as rehearsed earlier in this report, could be deemed to apply equally to those candidates whose centres chose to enter them for AQA qualifications, the fall in outcomes for AQA's 16 year old candidates from Wales (from 59.7% to 57.8% is less than the equivalent fall in outcomes for WJEC's candidates from Wales. In addition, candidates in Wales form a tiny minority of AQA's overall candidature (less than 1%). On balance it is not possible to determine definitively, from the evidence available, whether outcomes for AQA's Welsh candidates for June 2012 are any less reliable than the equivalent outcomes for June 2011. Should further evidence be received, or further action subsequently taken, in respect of AQA grading elsewhere, we would be prepared to look at this issue again.

Scrutiny of WJEC's GCSE English Language specification

72. Each year, as part of its ongoing programme of monitoring, the Qualifications and Learning Division of the Welsh Government conduct a number of 'scrutinies' of WJEC's general qualifications. A scrutiny is an in depth review of the quality of identified qualifications in terms of the design, delivery and standards of those qualifications and in terms of the awarding organisation's compliance with the Code of Practice for General Qualifications. In 2012 the nominated specifications for scrutiny are GCSE English Language, GCSE Mathematics, GCSE History and GCSE Science.
73. For each of these scrutinies the Welsh Government contracts with a team of subject specialists, who, under the leadership of an experienced Welsh Government official, conduct a series of activities designed to elicit qualitative and quantitative evidence on the performance of each qualification. These activities include:

observations of various meetings of WJEC’s examiners; a review of the question papers, tasks and supporting documentation; a review of candidate scripts; comparison with the question papers, tasks and candidates’ work from other awarding organisations and meetings with staff from centres who have delivered the qualification.

74. The scrutiny of GCSE English Language is currently underway and will not be complete until the end of this year. However, in addition to the comments cited earlier in this report in relation to the awarding meeting, the scrutineers have provided this investigation with some preliminary findings which indicate that:

- *“In both Foundation and Higher Tier examinations and also in the internally-assessed components, the level of demand was appropriately challenging for the full range of the candidature and offered scope for students to show what they knew, understood and could do.”*
- *“The levels of demand are appropriately differentiated.”*

The next stage in this scrutiny will be the review of candidates’ work at key grade boundaries. This will be followed by a number of meetings with selected centres across Wales. Any concerns raised about the standard of candidates’ work, or about the design or delivery of the qualification will be identified in a report to WJEC. Regulatory officials will, where necessary, agree the nature and timescale of any appropriate action to be taken, with WJEC and will monitor compliance against a formal action plan.

Matters arising from Ofqual’s report on GCSE English Awards 2012

75. Welsh Government officials are grateful to officials at Ofqual for the opportunity to observe their meetings with awarding organisations in the week leading up to the publication of Ofqual’s report on GCSE English on 31st August 2012 – and for the willingness expressed by Ofqual to share the data which they received from awarding organisations as well as some of their initial findings in the preparation of their report.

76. Ofqual’s report makes clear that it is a report about and for England. Similarly, it should be clearly understood that this is a report about, and for, Wales.

77. We fully respect the integrity of the recommendations made by Ofqual in the light of its own initial findings for England from the evidence it has considered. However, our findings for Wales differ in part to Ofqual’s findings for England. In particular, we cannot agree that, in relation to candidates in Wales that:

“the standard set for these English [including English Language] GCSEs is comparable with the standard set in previous years”.

In addition, while we share Ofqual’s view that *“each exam board set standards for the qualifications in the manner expected”* we have less confidence than Ofqual that, in relation to candidates for GCSE English Language in Wales, *“candidates’ work has been properly graded”*.

78. Ofqual express a clear view that the decision to move away from a modular system for GCSEs in England was the right decision for England. The findings of this investigation do not point to a need to move away from a modular system in Wales. Rather they reinforce, in our view, the need to take into account a full, appropriate and relevant range of both qualitative and quantitative data, for the particular cohort in question, when determining grade boundaries. In particular, our findings suggest that, in determining outcomes for candidates in Wales the most relevant data is that which relates to candidates in Wales rather than that which relates to candidates in England.

Reflection on the challenges presented by qualifications that are delivered by Wales and England

79. One of the findings of this investigation is that a key difficulty in maintaining comparable outcomes for candidates in Wales is the need to reach a compromise where the same qualification is offered in both England and Wales. If we were designing a methodology to ensure comparable outcomes over time for a qualification taken by the vast majority of learners in Wales, we would not choose to do so by using a methodology which only made reference to the predicted and actual outcomes of candidates in England. Ofqual has regulatory jurisdiction over the majority of candidates for this qualification and it was judged important to reach a compromise on determining candidates’ outcomes, in the days immediately before the publication of the provisional GCSE results. In hindsight, and in the light of the evidence which has subsequently become available, this compromise was, in our judgement, an inappropriate solution to accept on behalf of candidates in Wales.
80. In order to ensure that awarding decisions for candidates in Wales are made entirely on the basis of data relevant to Wales, it would be necessary to ensure that candidates in Wales followed different qualification specifications to candidates in England. To a certain extent this will be happening for many GCSE courses which start this September as candidates in Wales are able to continue to enter candidates for modular (or ‘unitised’) GCSEs which candidates in England will not be able to take. Although the linear and unitised qualifications are very similar, they exist and have been separately accredited: linear specifications by the regulators in Wales, England

and Northern Ireland, and unitised specifications only by the regulators in Wales and Northern Ireland. It is not inconceivable that any future divergence of policy on qualifications could lead to entirely separate qualifications at this level in Wales, England and Northern Ireland – and indeed this is one of the issues which is currently being debated by the Review of Qualifications 14-19 in Wales.

Recommendations of this investigation

81. Having considered all the available evidence, it is the conclusion of this investigation that the published provisional outcomes for candidates in Wales for GCSE English Language at Grade C and above, which show a fall of 3.9 percentage points from the equivalent outcomes in 2011, are not secure or supported by any reasonable justification. It is therefore recommended that the Minister for Education and Skills, representing the Welsh Ministers in fulfilling their responsibilities in relation to relevant qualifications as set out in Section 30 of the Education Act 1997, should:
- a. **request that WJEC undertake a regrading of GCSE English Language in order to achieve outcomes that are as similar as possible to the outcomes achieved by candidates in 2011, on the basis that there is no reason to believe that the 2012 Wales cohort was significantly different to the 2011 Wales cohort. It would be strongly preferable for this regrading to be applied to all of WJEC's candidates in both Wales and in England but, in the event of the regulator in England (Ofqual) refusing to endorse this regrading, it should be applied only to candidates in Wales;**
 - b. **take no action in respect of the published provisional outcomes for AQA on the basis that, on balance, there is no clear evidence that candidates in Wales who were entered for GCSE English Language with AQA received inappropriate grades, though this could be reviewed in the light of any further action taken by the regulators in England or Northern Ireland.**
82. With regard to the variability in outcomes between centres, it is recommended that the School Standards Unit should:
- c. **in consultation with WJEC and with schools across Wales, continue to undertake the programme of activities that it has already commenced in order to learn the lessons of successful centres in Wales and improve the potential outcomes for less successful centres, with particular regard to the implementation of controlled assessment.**

83. With regard to the difficulty of maintaining comparable outcomes for Wales when qualifications are subject to regulation by a regulator in England which is tasked with implementing policies and methodologies that are designed for England, it is recommended that the Review of Qualifications 14-19 in Wales should:
- d. take due cognisance of the issues raised in this report in forming its views on the extent of future divergence from qualifications in England.**
84. With regard to the difficulties presented by the large proportion of controlled assessment in GCSE English Language, it is recommended that the Review of Qualifications 14-19 in Wales should:
- e. give consideration as to whether controlled assessment is an appropriate assessment methodology for GCSE English Language.**
85. With regard to the difficulties of achieving a good balance between approaches to grading which are akin to norm-referencing and criteria based approaches grading, it is recommended that the Review of Qualifications 14-19 in Wales should:
- f. give consideration as to the most appropriate method of grading the next generation of general qualifications in Wales.**

**Qualifications and Learning Division
Department for Education and Skills
Welsh Government
September 2012**