



REPORT INTO THE AWARD OF GCSE ENGLISH LANGUAGE SUMMER 2018

September 2018

CONTENTS

- 1. Foreword**
- 2. Executive summary**
- 3. Background**
- 4. Concerns raised with us**
- 5. Our approach to the review**
- 6. Setting standards in the new GCSEs**
- 7. The awarding process**
- 8. Summer 2018 GCSE English Language award**
- 9. Analysis of results**
- 10. Conclusions**
- 11. Looking forward**

Appendix 1: Overview of the new GCSE English Language qualification

Appendix 2: Comparable outcomes

Appendix 3: Grade boundaries

Appendix 4: The sawtooth effect

Appendix 5: Statistical prediction methodology

Appendix 6: Glossary

1. Foreword

Qualifications Wales was established as the independent regulator for qualifications other than degrees in September 2015. We are responsible for maintaining standards over time in qualifications such as GCSEs, and consider this to be a core aspect of our work. Our principal aims are to ensure that qualifications, and the qualifications system, meet the reasonable needs of students and promote public confidence.

There are many interpretations of the word 'standard'. Standards are founded upon the principle of fairness. Ultimately, maintaining standards ensures that the value assigned to the grade awarded to a student remains constant year-on-year. For example, a grade C result in any given subject should be as similar as possible in its reflection of student attainment from one year to the next.

It is crucial to maintain a single standard at national level - this ensures fairness for students past, present and future. At a national level the awarding process can take account of factors such as age distribution and is designed to ensure stability. Variations at this level can be accommodated through the careful balance of statistical evidence and senior examiners' judgement that is used when establishing grade boundaries.

Changes in results at a local level can be influenced by a range of factors. These include: the overall ability of the year group, exam entry decisions, familiarity with the qualification, how a course is taught. All these factors, and others besides, can lead to changes in a school's results from one year to the next. However, changes at a local level usually cancel each other out when the results are aggregated together. All things being equal, these will give stable outcomes at a national level.

The awarding process, through which grade boundaries are established and standards maintained, is complex. It is a specialist area that is not always easy to understand or communicate, involving a range of statistical and judgemental evidence.

Because it is complex, there are many misunderstandings about how it works, for example that there are fixed quotas of specific grades available in each exam series. This is not the case. There are no limits to the number or proportion of particular grades awarded in any given year – the comparable outcomes approach to awarding sets expectations, not limits.

This year concerns have been raised with us about where the grade C boundary for GCSE English Language has been set. In summer 2018, the grade C boundary was set at 55% of the total marks available for the qualification compared to 50% in summer 2017 and 51% in November 2017.

As the regulator, we monitored the WJEC meeting where the grade boundaries were set for this year's award. We found that the award was compliant with our requirements and the outcomes were in line with our expectations. However, in response to the concerns raised about grade boundaries and the potential impact of changes to entry patterns, we decided to look again at this year's award and outcomes. In doing so, we looked at additional information not available at the time of the award and kept an open mind as to what this evidence might tell us.

This report is intended as a comprehensive response to those who have raised their concerns with us. It includes details of the review, including the evidence that we considered and information about how grade boundaries are determined. We are publishing it to provide full transparency in our work and to report to those who may have unvoiced concerns.

Philip Blaker
Chief Executive
September 2018

2. Executive summary

In Wales this summer, 271,761 full-course GCSEs were awarded, including GCSE English Language. This was one of 21 reformed GCSEs awarded. Results were announced on 23 August 2018 and overall they remained broadly stable.

After results were published, some schools, local authorities and a regional education consortium in north Wales contacted us to raise concerns about grade boundaries for GCSE English Language. As the regulator, we took swift action to investigate.

The concerns focused on the impact on students of setting the grade C boundary in summer 2018 at a higher mark than in previous awards. Those who contacted us questioned whether the summer 2018 grade boundary was justified and whether it had been skewed by the effect of the very high levels of early entry in 2017.

With the principle of ensuring fairness to students past, present and future firmly in mind, we reviewed this summer's GCSE English Language award to investigate:

- **whether standards for the qualification had been maintained in line with previous exam series.**
- **whether students who sat the qualification for the first time this summer had been disadvantaged compared to those who sat the qualification last year.**

Having reviewed this summer's award, our conclusions are that:

- **the national standard for GCSE English Language was maintained in the summer 2018 award.**
- **there is no evidence that the GCSE English Language award in summer 2018 unfairly disadvantaged students sitting the qualification for the first time compared to those who sat the qualification in 2017.**
- **we are confident that students' achievements have been fairly recognised. Users can rely on the fact that the grades awarded in summer 2018 are equivalent to those awarded in 2017.**

On the basis of these conclusions, there are no grounds for reconsidering the GCSE English Language award for summer 2018.

Our analysis of results by entry practice at a centre level found that there is evidence of a statistical relationship between centre entry practice and results. However, the relationship differs across grades. Even for centres that used similar entry strategies, the relationship with results is not consistent. Other factors, in addition to entry practice, also influence outcomes at the centre level.

There is no evidence that this relationship is the result of inconsistent standards being applied across different exam series. However, it does show that the significant differences and changes in centre entry strategies over recent years do not promote a level playing field for students.

The action taken by the Welsh Government to discourage widespread early entry has helped to substantially reduce the levels of early entry that we saw last summer. We welcome these developments as they are likely to lead to more stable results in future years that will help to promote fairness for all students and users of qualifications.

3. Background

There is a new GCSE English Language qualification designed specifically for Wales. It was introduced for first teaching in 2015 alongside GCSEs in English Literature, Welsh Language, Welsh Literature, Mathematics and Mathematics-Numeracy¹. WJEC is the only awarding body that offers these reformed GCSEs in Wales.

The new GCSE English Language qualification is very different to its predecessor, both in structure and in content. It is an untiered qualification, which means that all students sit the same exam paper, regardless of the grade they are aiming for. It is also a linear qualification, which means that students must sit their exams all together, at the end of the course.

There are three units in the qualification. Unit 1 assesses speaking and listening skills (referred to as oracy) and contributes 20% towards the final grade. Units 2 and 3 both assess reading and writing skills (in different ways) and each contribute 40% towards the final grade. Further details about these units are included in Appendix 1, together with a link to the qualification specification.

GCSE English Language was awarded for the first time in summer 2017. In November 2017 there was an opportunity for students to sit the qualification again, provided they had already sat the qualification in summer 2017.

With multiple exam series across the two years that students study GCSE English Language, schools can enter students for more than one exam sitting (as with other linear qualifications). This system has lent itself to an approach whereby students can be entered for examinations before the end of Year 11 (early entry), and sometimes on multiple occasions (multiple entry).

In recent years, we have seen record levels of early and multiple entry in Wales, with schools registering significant numbers of students to sit linear GCSE exams before the end of Year 11.

In January 2017, we conducted research² into this growing practice and recommended changes to the Welsh Government to ensure that the interests of students are put first.

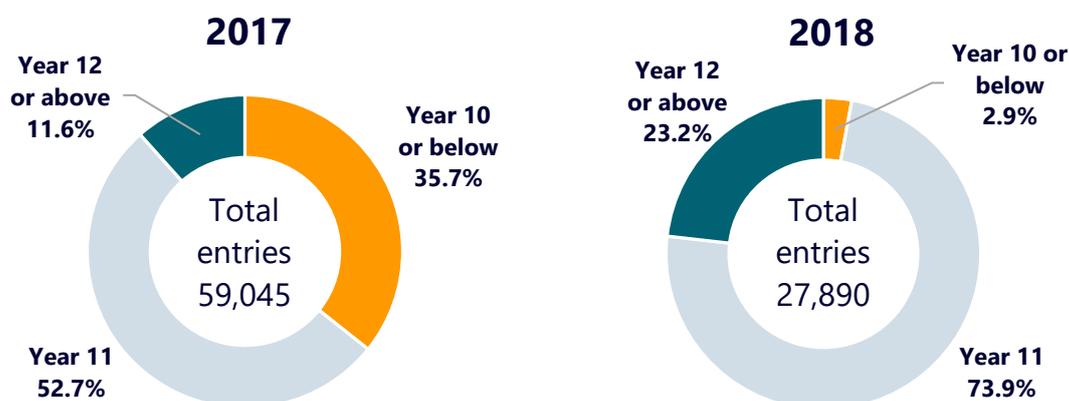
¹ These were the first GCSE subjects to be reformed.

² <https://www.qualificationswales.org/media/2825/approaches-to-early-and-multiple-entry-2017-e.pdf>

Responding to our findings, Welsh Government announced that only the first grade awarded to a student counts towards a school's performance measures (the previous policy allowed schools to count the best grade achieved by a pupil from multiple sittings). This new policy took effect from summer 2018 (for summer 2019) and led to a marked change in the number of student entries for GCSE English Language in summer 2018 compared to summer 2017.

There was a significant decrease (96%) in year 10 entries as well as a decrease (34%) in year 11 entries (many of whom had sat the exam in June 2017 and November 2017). These decreases in the Year 10 and Year 11 entries meant that a higher proportion of the results in GCSE English Language this summer were for Year 12 and above students who were resitting the qualification.

Figure 3.1 GCSE English Language entries by age



4. Concerns raised with us

The 2018 GCSE results for Wales and the rest of the UK were published by exam boards on 23 August 2018. In Wales, there had been significant changes to the entry patterns for this summer's exams and several new GCSEs were awarded for the first time.

On 28 August, we were contacted by GwE³ (the school effectiveness and improvement service for north Wales) alerting us to concerns raised by schools in its region about the award of the GCSE English Language qualification. In the following weeks we also received communication directly from some local authorities and several secondary school headteachers in north Wales expressing similar concerns.

The concerns focused on the impact that setting the grade C boundary in summer 2018 at a higher mark than in previous series had on student results. Specifically, the concerns were that:

- the total number of marks required to achieve a grade C in summer 2018, compared to summer 2017, was too high. In summer 2018 the grade C boundary was set at 220 marks (out of a total of 400 marks), in summer 2017 the grade C boundary was set at 200 marks and in November 2017, the grade C boundary was set at 204 marks.
- this increase in the grade C boundary mark may have unfairly disadvantaged 16-year-old students who sat the qualification for the first time in summer 2018, compared to those students who had been entered early for the qualification in 2017, when the grade C boundary mark was lower.
- results for schools who had not entered candidates early in 2017 may have been negatively affected in comparison to schools who had used early entry.

We had closely monitored the award and the outcomes for GCSE English Language this summer and were not aware of any specific cause for concern⁴. However, in response to the concerns raised by GwE, we decided to take a second look at this summer's award. We ensured that we considered all subsequent communication we received from schools and local authorities as part of this work.

³ One of the four regional education consortia in Wales.

⁴ We outline later in this report our approach to monitoring GCSE awards.

5. Our approach to the review

Our overall aims in conducting this review were to:

- understand the nature and the basis of the concerns raised by schools and others.
- explain how GCSE English Language was awarded this summer.
- identify any issues and take swift action to address them.
- identify whether there are any broader questions or areas of further work for us and others.

We sought to investigate:

- whether standards for the qualification have been maintained in line with previous exam series, specifically:
 - that the award was conducted appropriately and that grade boundaries were set in accordance with our regulatory requirements.
 - that this year's grade boundaries reflected a comparable standard to previous awards.
- whether students who sat the qualification for the first time in 2018 were unfairly disadvantaged compared to those who sat the qualification in 2017, specifically:
 - whether centre-level results were significantly different in summer 2018 compared to previous years.
 - whether any differences in centres results appear to be related to whether or not centres practised early entry.

In response to the concerns raised, we have reviewed the evidence relating to this summer's award of GCSE English Language in detail, including:

- correspondence from GwE including some limited data about GCSE English Language results of schools in its area.
- letters from schools and Local Authorities raising concerns and questions about this summer's GCSE English Language award.
- our observer's report on the awarding meeting.

- our Data Exchange Procedures for summer 2018.
- WJEC's Chair of Examiners' report.
- WJEC's Statistical Officer's report.
- WJEC's Principal Examiners' report.
- the report on the summer 2018 GCSE English Language award that we required WJEC to produce for this review.
- data received from WJEC including:
 - predictions and outcomes provided during awarding process.
 - centre variation data.
 - comparison of candidate outcomes between series.
 - best grade by route analysis.
- national outcomes published by the Joint Council for Qualifications.

Our review and this report are focused only on GCSE English Language. The correspondence we received from centres focused on GCSE English Language. A few schools also asked about the award of the GCSE Mathematics and GCSE Mathematics-Numeracy qualifications. Those concerns, however, were not specific and did not lead us to expand the scope of our review to look again at how those qualifications were awarded.

6. Setting standards in the new GCSEs

In this section we explain the method for maintaining standards that we require WJEC to use when awarding reformed GCSEs to ensure that outcomes remain comparable year on year.

Our approach

Six new GCSE qualifications, designed specifically for students in Wales, were awarded for the first time in summer 2017 including GCSE English Language. Our aim in overseeing each series of exams and awards is to make sure that standards are consistent over time. In other words, to ensure as far as possible that a student of a given ability is equally likely to achieve a certain grade, no matter which exam series they sit. We need to consider fairness to students past, present and future. We must also consider the needs of users who rely on GCSE results and expect grades to reflect consistent standards of attainment.

During this period of reform, our priority has been to transfer the 'standard' from the legacy qualifications to the new qualification and then to maintain it over time. We required WJEC to use the established 'comparable outcomes' approach to ensure students were neither advantaged nor disadvantaged simply by being among the first cohorts to sit the new qualification⁵. This approach can compensate in situations where a small drop in assessment performance may occur when new qualifications are introduced. By assessment performance we mean the marks gained by a student in their assessments, not the overall grade that those marks equate to.

Comparable outcomes can also be used to predict the expected outcomes at a national level. These predictions help to safeguard against increases in the proportion of students achieving higher grades without evidence of real improvements in attainment (grade inflation).

In setting standards for the first award of the new GCSEs and A levels we, along with the regulators in England (Ofqual) and Northern Ireland (CCEA), have carried forward the standards from the old qualifications. In subsequent years, each regulator independently requires awarding bodies to apply methods for maintaining grade standards over time. These are based on similar and well-established principles. The specific detail of the approaches differs slightly from one country to another

⁵ We describe in more detail the comparable outcomes approach in Appendix 2, including how it is used to maintain standards as new qualifications are introduced and become established.

reflecting the cohorts taking the qualifications, the data available in each country and different emphasis on expert judgement by senior examiners. We, along with Ofqual and CCEA are working together to keep our respective approaches to maintaining standards under review and to ensure GCSE qualifications are valued equally and assessed fairly, wherever they are taken.

Where the cohort of students taking the qualification is similar to previous years, we expect results to be similar. However, there are no quotas or allocations to be filled and we do expect to see some variation. If results vary more than we might normally expect, we require evidence to support a genuine change in student attainment. Where exam boards provide us with evidence to support an improvement or indeed a decline in student attainment, we consider whether the evidence justifies the proposed change to outcomes.

The comparable outcomes approach will produce similar results year-on-year if the cohort for the subject remains similar in terms of its overall ability. However, when a qualification changes, there can be more year-on-year variability than usual in the results for individual schools and colleges, as they may have taken different approaches to delivering the new qualification. This variability can be due to a variety of factors, including a change in entry strategy or a change in approach to delivering a qualification. For new qualifications, familiarity with the assessment arrangements can also play a part.

Maintaining standards over time

The process for maintaining GCSE standards over time takes place at the point at which a qualification is awarded and is closely monitored by us. Each year, statistical analysis of previous cohorts is used to predict this year's outcomes. When generating GCSE predictions, we require WJEC to use the 'common centres' method, based only on results for 16-year-old candidates in Wales⁶.

As the regulator, we review the outcomes for each qualification. Where an award leads to outcomes that are significantly different to predictions, we review the evidence submitted by the exam board to decide whether the award is justified.

In reviewing the evidence, we consider:

- maintenance of standards.
- fairness for students.
- public confidence.

⁶ For a more detailed explanation of the 'common centres' method, see Appendix 5.

7. The awarding process

In this section, we explain how awarding works in practice, including the key roles and responsibilities of those involved.

Preparing for the examination series

In the months leading up to each exam series, we take a close interest in WJEC's preparations for awarding each qualification. We check to make sure that we are content with the approach it proposes to take to ensure that standards are maintained. For each exam series, we produce a regulatory document which outlines the procedures by which WJEC is required to award the qualifications and the data it must provide to us during the examination series.

Exam boards aim to develop question papers with the same level of difficulty year-on-year. However, it is very hard (almost impossible) to produce two exam papers that are exactly as difficult as each other. Clearly it would be unfair for students to get a lower grade just because they sat a more difficult paper. So, in each exam series, the exam board sets grade boundaries for each unit and for the overall qualification. This is only one of the reasons why grade boundaries often vary from one exam series to the next. Another reason grade boundaries can change is to compensate for where students' assessment performance in a newly reformed qualification typically falls in the first year, before gradually improving over time.

The awarding committee

WJEC convenes an awarding committee for each qualification. The awarding committee is made up of senior examiners who have responsibility for writing the question papers and overseeing the marking of those papers. The committee is responsible for recommending appropriate grade boundaries to maintain standards year-on-year. The committee is provided with statistical evidence and support from the WJEC research and statistical team. The committee needs to balance a range of evidence when deciding where to recommend the grade boundaries. It must consider the statistical evidence, including predictions, alongside its review of student work.

Setting grade boundaries

Grade boundaries are the minimum mark required for a particular grade. For example, a C boundary of 250 means that you need at least 250 marks to get grade

C. In GCSEs, the awarding committee sets the judgemental grade boundaries (A, C and F) and the other grade boundaries are then calculated. We describe this process in more detail in Appendix 3.

Once all the marking has been completed, the awarding committee meets to recommend where to set the grade boundaries for that examination series. The committee is aiming to find the minimum mark for this year's work that best reflects the standard at which last year's grade boundary was set. To do this it looks at student work from this year that is close to the grade boundary proposed from the statistical evidence and compares it with student work at that grade from last year. The awarding committee is also presented with statistical information about the exams themselves. This includes an analysis of marks students achieved on each individual question and an indication of where grade boundaries could be set to achieve similar overall outcomes to last year's exams.

When a new qualification is introduced, students' exam performance in the first few years is likely to dip compared to students taking the previous qualification who were more familiar with the exams. This is followed by improved performance over time as familiarity with the new qualification increases. This is known as the sawtooth effect which we describe in more detail in Appendix 4. The awarding committee uses all the evidence available to it to identify what adjustment is needed to allow for this effect.

The awarding committee needs to balance both the statistical evidence and the standard of student work that it has reviewed. As the starting point, it uses the statistical analysis of how this year's assessments have functioned including the predicted outcomes to identify the range of student work to be looked at. It considers candidate work from the current exam series and compares it to candidate work that was at the grade boundary in the last exam series. The deciding factor is the view of the awarding committee about which mark best represents a comparable standard of work to last year.

After carefully considering all the evidence, the awarding committee recommends the minimum mark needed for each grade to ensure the standards required for this year are the same as for previous years. The exam board then makes quality checks to ensure that the committee has considered all the available evidence and that the grade boundary decisions are justified. The final grade boundaries are reviewed and signed-off by the exam board's responsible officer.

Role of the regulator

We scrutinise WJEC's process for setting appropriate grade boundaries throughout the awarding period. Before the award of a qualification is confirmed, WJEC must report their proposed outcomes to us at a Maintenance of Standards meeting. Following this meeting, provided that all outcomes have been agreed, WJEC then processes these grade boundaries and confirms the grade each student has achieved.

8. Summer 2018 GCSE English Language award

In this section we give a detailed account of how GCSE English Language was awarded in summer 2018. We explain the evidence considered and decisions taken by the awarding committee when setting the grade boundaries.

Entries

The number of students who sat GCSE English Language fell this summer. This was due to the drop in Year 10 students sitting exams early and a drop in Year 11 students sitting exams this summer because they had sat the qualification in an earlier series and not returned to resit. This decrease in the cohort size meant that a higher proportion of the results in GCSE English Language this summer were for Year 12 and above students, most of whom were resitting the qualification.

Table 8.1 shows the number of students sitting GCSE English Language in June 2017 and June 2018 by age.

Table 8.1 Certifications by age (Wales).

	Number of students				Percentage of total		
	Year 10 and under	Year 11	Year 12 and above	Total	Year 10 and under	Year 11	Year 12 and above
June 2017	20,979	31,064	6,814	58,857	35.6%	52.8%	11.6%
June 2018	709	20,575	6,571	27,855	2.5%	73.9%	23.6%

Source: Joint Council for Qualifications.

Given this significant change to the nature of the cohort sitting the GCSE English Language in summer 2018 compared to summer 2017 we did not expect results at a national level to be similar. After the exams had been sat, we published an article highlighting the changes to the cohort sitting GCSE English Language and what that may mean for the summer results⁷.

Pre-award analysis

For all reformed GCSEs in English, Welsh and Mathematics subjects, we required WJEC to generate predictions based on expected best-grade outcomes for

⁷ www.qualificationswales.org/media/3516/spotlight-on-english.pdf

16-year-olds as set out in the data exchange procedures (*Wales Summer 2018 Data Exchange Procedures*⁸). This approach was consistent with the prediction methodology used in June 2017 to manage the transition of the standard from the legacy qualifications. We required WJEC to use the June 2017 prediction approach to ensure that the national standard was carried forward in the context of early and repeat entry in the legacy qualifications.

Given the significant early and repeat entry undertaken by the 2018 16-year-old cohort, we judged that the same prediction approach would be the most appropriate in order to secure a comparable best-grade outcome for 16-year-olds in summer 2018 compared to 16-year-olds in summer 2017. In effect, by using this prediction method, we expected the national best-grade outcomes for 16-year-olds in summer 2018 to be broadly similar to the national best-grade outcomes for 16-year-olds in summer 2017. We explain the best-grade prediction method in detail in Appendix 5. The awarding committee used the prediction as part of the statistical evidence to be considered alongside its review of student work.

The award

The summer 2018 awarding committee for GCSE English Language comprised the same members who awarded both the summer 2017 qualification and the November 2017 qualification. The WJEC statistical officer presented the statistical evidence including the initial recommended boundaries (IRBs) to the awarding committee. These were used to identify the range of marks from which to draw samples of student work for the committee to look at. The committee also looked at work at the grade boundaries from previous exam series.

The Principal Examiner for each unit gave an overview of how the question paper functioned. For the oracy unit (Unit 1), the Chair's report stated that the Principal Moderator highlighted *'the reading of scripts still being in evidence'* and *'that some centres are not teaching candidates the necessary skills nor giving them opportunity to practise these skills, before the actual assessments take place.'* In Units 2 and 3, the Chair's report noted *'written accuracy proved a significant area of concern'*. *'Candidates whose work is characterised by numerous errors, a struggle to control tense and agreement, or an inability to punctuate with control and coherence did struggle.... This is likely to be the most significant factor affecting candidates who do not achieve their desired mark.'*

⁸ www.qualificationswales.org/english/publications/data-exchange-procedures-wales-summer-2018/

Furthermore, the Principal Examiner explained to the committee that there was some evidence that students were using strategies to pick up marks as some exam papers *'included a tick list of different types of punctuation... which (candidates) then tried to shoehorn into their writing.'* This information formed the basis of the Examiner report that was published on results day⁹.

Table 8.2 gives the ranges of marks considered for each key grade boundary and the initial recommended boundaries (IRB), each unit is marked out of 80.

Table 8.2 Script review ranges and IRBs.

	Unit 1 (NEA)			Unit 2			Unit 3		
	A	C	F	A	C	F	A	C	F
Range	62-66	46-50	22-26	54-58	42-46	14-18	50-54	38-42	10-14
IRB	64	48	24	56	44	16	52	40	12

Each committee member independently scrutinised a range of student work and recorded whether or not they considered it worthy of the grade. These individual decisions were collected and summarised. The committee then determined collectively the range of marks in which the final grade boundary could be set. The Chair then selected the recommended boundary marks, taking the committee's views and the statistical evidence into account.

Table 8.3 shows the boundaries recommended by the committee for each unit at each key grade. At C, the awarding committee recommended that the grade boundaries for two of the units should be set one mark higher than the initial recommended boundary (IRB). The committee considered this was necessary to ensure that the award reflected the standard of attainment expected for a grade C. The Chair's report notes that the marks set for the C/D boundary on each unit represented the minimum standard expected of a C grade and was supported by all members of the committee.

Table 8.3 IRBs and Recommended Boundaries.

	Unit 1 (NEA)			Unit 2			Unit 3		
	A	C	F	A	C	F	A	C	F
IRB	64	48	24	56	44	16	52	40	12
Recommended boundary	63	48	24	56	45	17	53	41	13

⁹ www.wjec.co.uk/examiner-reports/2018/gcse/wjec-gcse-english-language-new-report-summer-2018-e.pdf?language_id=1

WJEC used the grade boundaries recommended by the committee to generate provisional outcomes which were compared to the predicted outcomes. Table 8.4 below shows the predictions and provisional outcomes reviewed by the committee.

Table 8.4 Predictions and Outcomes.

Series		Total Wales [†]	A*	A	B	C	D	E	F	G	U
June 2018	2018 Outcomes	28,080	3.1	13.2	34.5	64.2	82.4	92.0	97.2	99.3	100.0
June 2018	2018 Predictions	28,080	3.0	13.9	37.8	64.9	82.6	92.4	97.6	99.6	100.0
	Outcomes - Predictions		0.1	-0.7	-3.3	-0.7	-0.2	-0.4	-0.4	-0.3	0.0

[†]excludes students who did not sit all of the units for the qualification
Counts have been rounded to the nearest five.

Table 8.5 shows the final unit grade boundaries endorsed by the awarding committee for the three exam series.

Table 8.5 Unit Judgemental Grade Boundaries.

	Unit 1 (NEA)			Unit 2			Unit 3		
	A	C	F	A	C	F	A	C	F
June 2017	64	46	24	56	41	16	52	36	12
Nov 2017	64	46	24	57	42	17	53	37	13
June 2018	63	48	24	56	45	17	53	41	13

Each unit is marked out of 80 marks. Table 8.6 shows the grade boundaries as a percentage of the maximum mark.

Table 8.6 Unit Judgemental Grade Boundaries as Percentages.

	Unit 1			Unit 2			Unit 3		
	A	C	F	A	C	F	A	C	F
June 2017	80.0%	57.5%	30.0%	70.0%	51.3%	20.0%	65.0%	45.0%	15.0%
Nov 2017	80.0%	57.5%	30.0%	71.3%	52.5%	21.3%	66.3%	46.3%	16.3%
June 2018	78.8%	60.0%	30.0%	70.0%	56.3%	21.3%	66.3%	51.3%	16.3%

The qualification is not equally weighted across the three units. Units 2 and 3 are worth 40% of the qualification, Unit 1 is worth 20% of the qualification. So before the unit marks can be aggregated, a scaling factor of 2 needs to be applied to Units 2 and 3 to ensure the overall qualification weightings are met. Table 8.7 shows the grade boundaries for the qualification after marks have been aggregated.

Table 8.7 Qualification Level Boundaries.

	Qualification grade boundaries			Subject grade boundaries as percentages		
	A	C	F	A	C	F
June 2017	280	200	80	70.0%	50.0%	20.0%
Nov 2017	284	204	84	71.0%	51.0%	21.0%
June 2018	281	220	84	70.3%	55.0%	21.0%

Table 8.8 shows how the grade boundaries are calculated from the raw marks.

Table 8.8 Grade C boundary.

Unit	June 2017 Aggregating			June 2018 Aggregating		
	Raw	Factor	Aggregated	Raw	Factor	Aggregated
1	46	1	46	48	1	48
2	41	2	82	45	2	90
3	36	2	72	41	2	82
Total	123	N/A	200	134	N/A	220

The increase in the grade C boundary of 20 marks at qualification level equates to students having to achieve 11 more raw marks to gain a grade C in summer 2018 when compared to summer 2017.

Regulatory oversight

As part of our monitoring programme, a subject expert observed and reported on the two-day GCSE English Language award. We were satisfied that the award was conducted in line with our requirements and expectations. Our observer's report provided evidence that grade boundaries were set in accordance with our rules, having considered both statistical evidence and expert judgement.

Having reviewed a range of student work, the committee concluded that the initially proposed grade boundaries for grade C needed to be increased by one mark to ensure that the award reflected the standard of attainment expected for that grade. The committee were particularly concerned about the quality of writing that they saw when they reviewed the student work. Our observer noted that the awarding committee commented '*on the relatively poor quality of writing across the grade boundaries*' leading to '*concern about an obvious imbalance between skills in reading and writing.*' This imbalance of skills meant that '*candidates at the key boundaries could pick up enough marks on the lower-tariff reading questions to compensate for*

deficiencies in the writing question(s) and achieve a total mark which gets them over the threshold.'

Given that this was the third exam series for the award of this qualification, we would expect there to be upward pressure on grade boundaries in order to adjust for student and teacher familiarity with the assessments. The grade boundaries that were set reflected a comparable standard to previous awards.

On 3 August 2018, WJEC presented the GCSE English Language award along with the other GCSE awards that they had made for the summer 2018 examination series to us at the Maintenance of Standards meeting. We were content that the GCSE English Language award had been conducted appropriately and that the standard had been maintained from the previous series. Following this meeting, WJEC finalised the award.

9. Analysis of results

In this section we present the findings from our analysis of the summer 2018 GCSE English Language outcomes at both a national and centre level.

We analysed a range of evidence to investigate the GCSE English Language results, including:

- entries¹⁰ and outcomes in current and previous exam series.
- best-grade outcomes.
- best-grade by entry route.
- the relationship of outcomes to predictions.
- the relationship between resitting and outcomes.
- changes in centre-level results ('centre variation', both schools and colleges).

The key questions for the statistical analysis were whether there was any evidence that:

- the standard from summer 2017 had not been maintained in the summer 2018 award.
- early and repeat entry impacts on centre level results.

In summary, we found:

- there was evidence to indicate that the national standard has been maintained.
 - The outcomes against prediction from the award provide evidence that the national standard has been maintained.
 - The variation in centre level results in English Language is similar to the Welsh and English GCSEs, and to that observed in summer 2017.
 - There are fewer centres with large variations this year, compared to last year and fewer seeing large reductions in outcomes.
- there was evidence to indicate a relationship between early and multiple entry practice and results. The relationship varies across the grades and there is a mixed picture at the centre level.
 - Centres that entered early saw, on average, an improvement in year-on-year best grade outcomes at A*-C and A*-F, while there was a

¹⁰ Entries here refers to the number of candidates achieving a grade, including a grade U.

reduction in centres that did not enter early. However, there was also evidence that centres that entered early saw poorer year-on-year outcomes at A* and A*-A, while centres that did not enter early saw an improvement at A* and a smaller reduction at A*-A.

- There is strong statistical evidence of a relationship between early and multiple entry practice and better average year-on-year outcomes at A*-C. However, this average picture masks variation across centres that adopted similar entry strategies. This suggests that other factors, in addition to or instead of entry practice, were associated with whether outcomes increased or decreased at a centre level.

Cohort definitions and entry opportunities

The approach to maintaining standards for GCSE qualifications awarded each summer focuses on candidates that are 16-years-old by 31 August. In other words, the focus is on candidates who are sitting their exams at the end of Year 11. In this analysis the '2017 Cohort' refers to candidates who were in Year 11 in June 2017, and the '2018 Cohort' refers to candidates who were in Year 11 in June 2018.

The new GCSE English Language was introduced for first teaching in September 2015. The first assessment opportunity for the 2017 Cohort was in June 2017. Because the 2017 Cohort was not able to enter early for GCSE English Language, the whole cohort was assessed in the same exam series, in June 2017.

By contrast, candidates in the 2018 Cohort could enter early for GCSE English Language in June 2017 (when they were in Year 10). Those candidates who did so could sit their exams up to three times by the end of Year 11 - in June 2017, as resitters in the November 2017 series, and again in June 2018.

National Results Analysis

The Joint Council for Qualifications (JCQ) data for the June 2017 and June 2018 series in Table 9.1 shows how the assessment opportunities at cohort level were associated with very different patterns of entry and results (both overall and by age) in these series. The age composition of those taking the exams at different points in time has to be taken into account to make valid comparisons.

Cumulative outcomes at grades A*-A and A*-G were stable over time. Outcomes for all candidates at grade A*-C in June 2018 included relatively poor outcomes for candidates that were 17 or older. This group included resitting candidates from the

2017 Cohort and formed a greater proportion of total certifications in June 2018. There was also a reduction in entries from 16-year-olds in the 2018 Cohort, compared to the 2017 Cohort in the June 2017 exam series, which included practically all 16-year-olds. This is because a substantial proportion of the 2018 Cohort did not return in the June 2018 series, having previously achieved grades in the June 2017 or November 2017 series.

Table 9.1 National Cumulative Results by Age (Percentages).

Age		Number			
		Sat	A*-A	A*-C	A*-G
15 or younger	June 2017	20,979	7.8	54.0	98.3
	June 2018	709	18.9	57.8	96.2
	<i>Change</i>	<i>-20,270</i>	<i>11.1</i>	<i>3.8</i>	<i>-2.1</i>
16	June 2017	31,064	14.3	64.8	98.6
	June 2018	20,575	11.8	45.5	98.2
	<i>Change</i>	<i>-10,489</i>	<i>-2.5</i>	<i>-19.3</i>	<i>-0.4</i>
17 or older	June 2017	6,814	2.1	36.6	97.2
	June 2018	6,571	2.4	27.0	98.4
	<i>Change</i>	<i>-243</i>	<i>0.3</i>	<i>-9.6</i>	<i>1.2</i>
All candidates	June 2017	58,857	10.6	57.7	98.3
	June 2018	27,855	9.8	41.4	98.2
	<i>Change</i>	<i>-31,002</i>	<i>-0.8</i>	<i>-16.3</i>	<i>-0.1</i>

Source: Joint Council for Qualifications

Table 9.2 shows the number of unique candidates taking different entry routes for GCSE English Language in the 2018 Cohort. There were 30,010 candidates who entered the qualification at least once. Of these, 30.1% sat in June 2018 only, while 20.9% were entered in all three series. Just over half of the candidates – 51% – sat the qualification more than once in the 12-month period between June 2017 and June 2018.

We also considered an analysis of best grade outcomes by the entry routes. However, we did not consider this a valid comparison because decisions on whether candidates resit are influenced by the results they achieved. This creates a 'selection bias' in relation to the candidates remaining in each entry route by June 2018, which makes comparison of the best grade results achieved via differing entry routes potentially misleading.

Table 9.2 Number of unique candidates in the 2018 Cohort, by entry route.

Entry Route	N	Percentage of total
June 2017 only	5,675	18.9
June & Nov 2017 only	4,125	13.7
June 2018 only	9,035	30.1
June 2017 & June 2018 only	4,885	16.3
All three	6,280	20.9
Other ¹¹	10	0.0
Total	30,010	100.0

Counts have been rounded to the nearest five.

To make valid comparisons on attainment at the cohort level, we need to look at the best grade achieved by each candidate. Best grades represent the highest grade achieved by a student across all their attempts. Table 9.3 includes information on the best-grades achieved by the 2017 Cohort and the 2018 Cohort in GCSE English Language.

These results were broadly stable. In terms of the judgementally set grade boundaries (A, C and F), there was a small decrease in the proportions of candidates achieving a grade A*-A or A*-C, and a small increase at A*- F. There was a larger decrease at grade B and above, but this grade boundary is arithmetically set once the judgmental grade boundaries have been agreed.

Table 9.3 Best grade cumulative outcomes for 16-year-old candidates.

Grade	2017 Cohort	2018 Cohort	Change
A*	3.2	3.2	0
A*-A	14.1	13.3	-0.8
A*-B	37.9	34.3	-3.6
A*-C	64.4	63.3	-1.1
A*-D	81.5	81.4	-0.1
A*-E	90.5	90.8	0.3
A*-F	95.9	96.1	0.2
A*-G	98.6	98.7	0.1
A*-U	100	100	0
N	30,790	30,015	-780

Counts have been rounded to the nearest five.

¹¹ The 'other' category represents a small number of candidates that resat the qualification and achieved a grade, having entered for a previous series but not present for the assessments.

Table 9.4 shows that actual results compared to the predicted results for those candidates in the 2018 Cohort that were used to create the prediction (referred to as matched candidates). Predictions are a quantitative way of helping to ensure that the standard of the qualification is carried forward each year. The predictions show what outcomes would be expected if the current cohort had taken the exam in the reference year. This prediction was on a best grade basis to reflect the fact that the age composition of the entry into each exam series has changed substantially. The prediction functions as a guide to the results that would be expected rather than determining, for example, the number of candidates who can achieve each grade.

Given that the prediction is based on a large sample of candidates representing most of the national cohort, we would expect it to function as accurate evidence to guide the awarding process. Our rules state that we would expect the actual results for matched candidates to be within plus or minus one per cent of the predicted results at grades A*-A and A*-C. Table 9.4 shows that this was achieved. This is evidence that the standard implied by the best grades achieved by the 2017 Cohort was maintained at the summer 2018 award for the 2018 Cohort¹².

Table 9.4 Cumulative outcomes against predictions.

Series	Wales only grade outcomes	Total Wales cash-in	Total Wales									
			A*	A*-A	A*-B	A*-C	A*-D	A*-E	A*-F	A*-G	A*-U	
June 2018	2018 actual results for matched candidates	28,080	3.1	13.2	34.5	64.2	82.4	92.0	97.2	99.3	100.0	
June 2018	2018 predicted results for matched candidates	28,080	3.0	13.9	37.8	64.9	82.6	92.4	97.6	99.6	100.0	
	Difference		0.1	-0.7	-3.3	-0.7	-0.2	-0.4	-0.4	-0.3	0.0	

Counts have been rounded to the nearest five.

Centre level results analysis (Centre Variation)

We expect results at a centre level to be variable compared to the previous year, even when results at a national level are stable. Results in some centres will improve and some will decrease. We would expect results to vary more (on average) when centres are small (due to low sample sizes) and where entry patterns change considerably relative to the previous year (due to likely changes in the ability in the subject among those being entered by such centres). There are also other factors

¹² The achieved best grade results for the whole cohort in Table 9.3 are slightly different than the actual results for matched candidates in Table 9.4, because the former also includes results from the remainder of the cohort. These candidates had slightly lower levels of attainment than the matched candidates.

that influence how results at a centre level change, including year-on-year changes in the ability of candidates and changes to delivery within a centre.

The changes in the entry by age into individual exam series, as well as the approach to maintaining comparable best grade outcomes in English Language, mean that it makes sense to consider how centre level results have changed on a best grade basis, for the 2018 Cohort compared to the 2017 Cohort.

Table 9.5 summarises how centre level results changed across centres for the 2018 Cohort compared to the 2017 Cohort. On average, centres saw a reduction in best-grade outcomes of 1.21 percentage points at grade A*-C. There were smaller average reductions at grades A*-A and A*-F. This is consistent with the data in Table 9.3. The standard deviation figures shown in Table 9.6 are a measure of the degree of variation in how centre level results have changed across all centres.

Table 9.5 GCSE English Language centre variation (2018 cohort compared to 2017 cohort, best grade).

	A	C	F
N	260	260	260
Mean	-0.94	-1.21	-1.06
Standard Deviation (SD)	6.24	14.56	11.76

Counts have been rounded to the nearest five.

Table 9.6 puts this decrease for English Language in the context of other reformed Welsh and English GCSEs that were awarded for the first time in 2017. For the 2017 Cohort, the comparison is made on the percentage change in best grade cumulative results in centres at grades A, C and F compared to the cohort that entered the legacy English Language GCSE in Summer 2016 (this comparison is labelled '2017 Cohort').

The table shows that the variation in changes to centre results (as measured by the standard deviation) at grades A*-C for GCSE English Language in the 2018 Cohort was similar to that seen for the 2017 Cohort, and similar to the variation seen in other Welsh and English GCSEs in the 2018 Cohort.

Table 9.6 Average centre variation by subject and series.

Subject	Mean change at C		Standard Deviation	
	2018 Cohort	2017 Cohort	2018 Cohort	2017 Cohort
English Language	-1.2	-2.0	14.6	15.9
English Literature	1.8	-0.7	15.6	14.1
Welsh	0.3	2.6	13.4	18.5
Welsh Literature	0.4	0.6	13.0	11.7

Table 9.7 shows the number of centres where there was large variation at grade A*-C. Large variation is here taken to be more than +/- 15.0 percentage points i.e. more than the width of one standard deviation. This shows that last year 60 centres saw a large change in results with 35 having large reductions in outcomes and 25 having large increases in outcomes. This year saw fewer large changes (35 compared to 60) and fewer centres have large reductions (20 compared to 35).

Table 9.7 Centre variation greater than +/-15 pp at grade C.

	N	> 15 %	Increased	Decreased
2017	255	60	25	35
2018	255	35	15	20

Counts have been rounded to the nearest five.

The next analysis explores the centre variation data in relation to entry practice in centres, comparing centres that entered candidates early (here defined as at least 10 candidates from the 2018 Cohort entered before the June 2018 exam series) against those that did not enter candidates early (defined as fewer than 10 candidates from the 2018 Cohort entering before June 2018¹³).

Figures 9.1 to 9.4 show the distribution of centre variation on a best grade basis for the 2018 Cohort compared to the 2017 Cohort at grades A*-A and A*-C, by the entry practice of the centre. The key point here is that the distributions for the different entry practices at these grades look similar. The distributions for grades A*-C show some shift between entry practices. However, closer inspection also shows that the spread of variation is quite wide within both entry practices, and the distributions overlap. In other words, both types of entry practice sometimes lead to poorer outcomes and sometimes to better outcomes at each grade.

¹³ This definition was chosen as early entry practices at centre level tended to involve substantially more than 10 candidates. This approach also provides a larger sample size in the 'not entered early' category for analysis purposes.

Figure 9.1 Centre variation, 2018 Cohort compared to 2017 Cohort for grades A*-C, by entry practice.

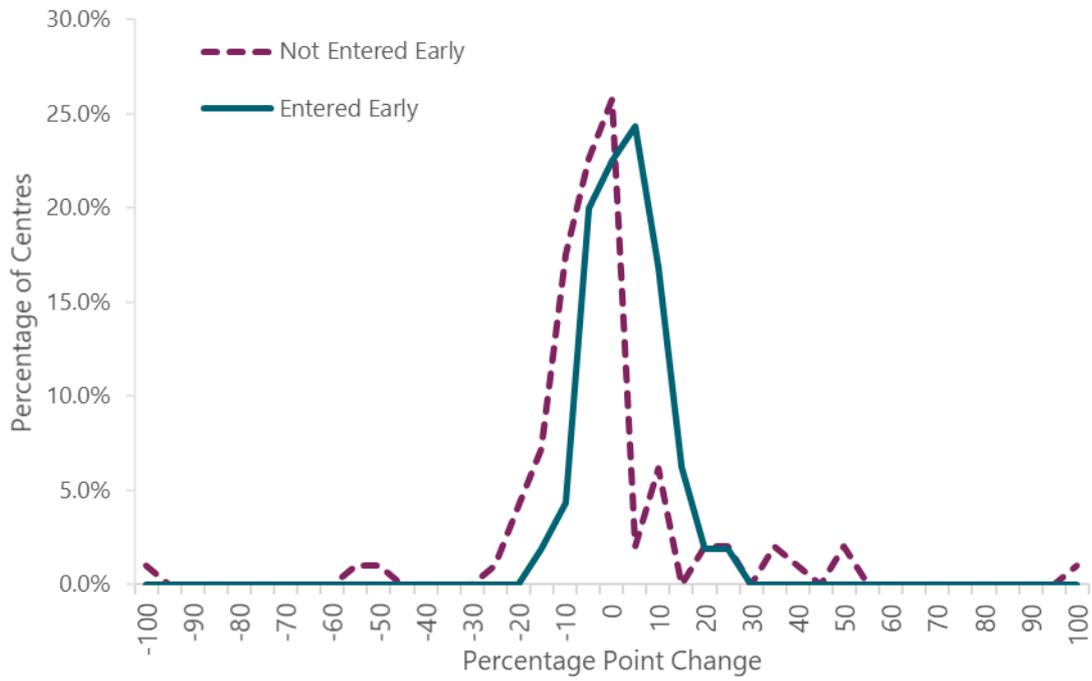


Figure 9.2 Boxplot of centre variation at grades A*-C by entry practice.

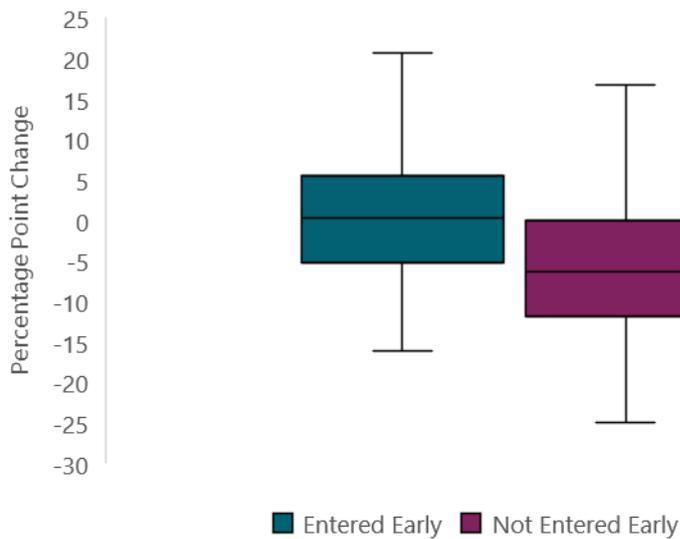


Figure 9.3 Centre variation, 2018 Cohort compared to 2017 Cohort for grades A*-A, by entry practice.

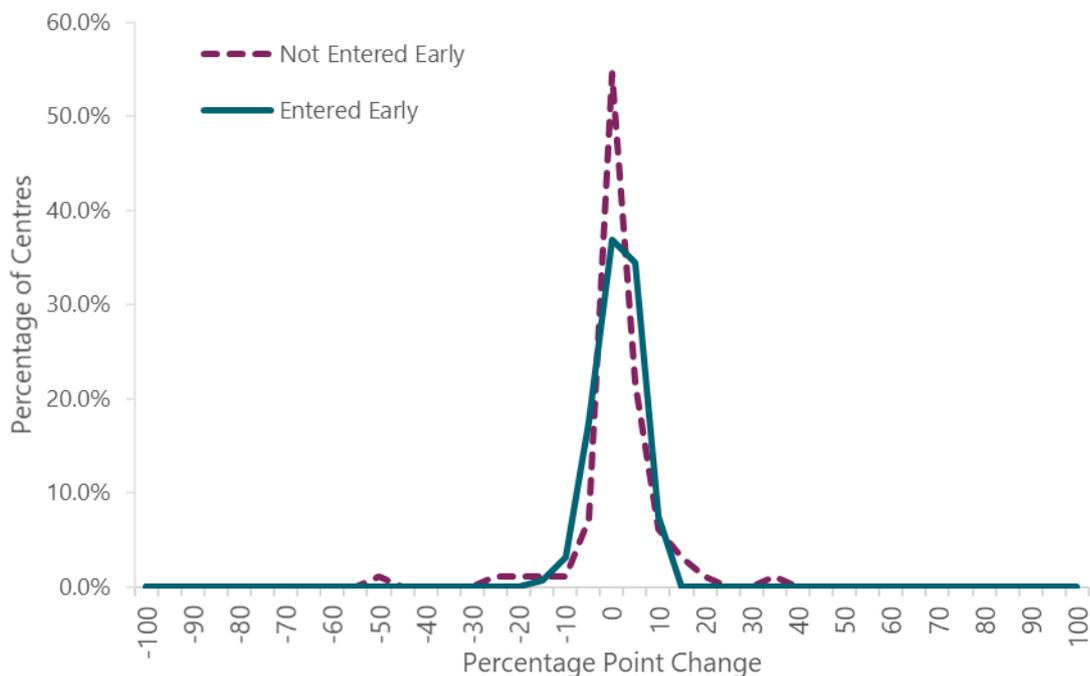


Figure 9.4 Boxplot of centre variation at grade A*-A by entry practice.

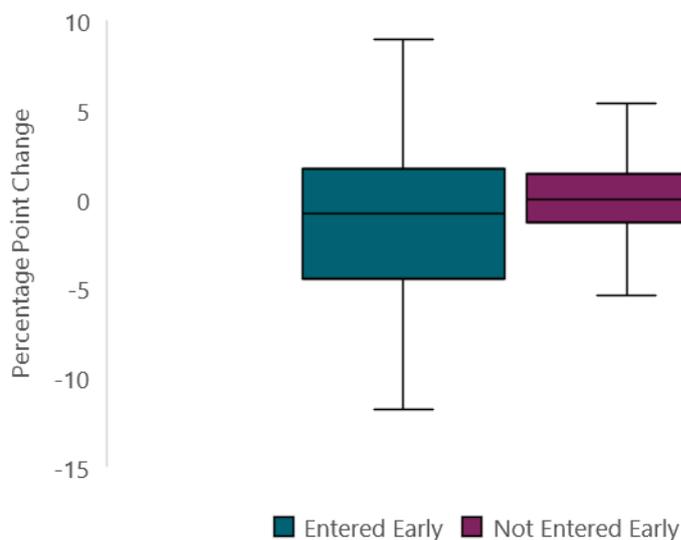


Table 9.8 provides summary statistics describing these distributions (rows 'N' to 'Kurtosis') as well as the results (in terms of p-values) of statistical tests.

The summary statistics in table 9.8 show that centres that did not practise early entry saw, on average, an improvement in their A* results of 1.1 percentage points compared to a decrease of 0.3 percentage points for centres that did enter candidates early. At A*-C, however, this pattern is reversed, with centres that didn't

enter candidates early, on average, seeing a decrease of 4.0 percentage points compared to an increase of 0.5 percentage points for centres that did use early entry. These figures suggest a hypothesis that there are different relationships between centre early and repeat entry practices and overall best-grade outcomes across the grades. More specifically, that early and repeat entry is associated with poorer outcomes at A*-A but better outcomes at A*-C and A*-F.

However, there are large variations around these average changes (as measured by the standard deviation and interquartile range). So, to assess the strength of the evidence of a relationship between entry practice and changes in centre level results, we ran statistical significance tests¹⁴. These tests help to establish whether the differences could have arisen by chance.

The results of these tests show that there is statistical evidence confirming the patterns observed in the mean average changes. The statistical evidence is, however, weaker at grades A*-A than elsewhere, which might imply more inconsistent relationships between entry practices and results at grades A*-A. While this analysis does provide evidence that entry practice has different impacts at different points in the grade distribution, it does not show statistical evidence of candidates being consistently advantaged or disadvantaged by entry practice across the grade distribution.

¹⁴ To determine the most valid approach to testing, the centre variation distributions were tested to see how 'normal' they were. The results of the 'Shapiro-Wilk' test, as well as the skewness and kurtosis statistics, provide strong evidence that the centre variation data is not normally distributed. Given this, a non-parametric test was performed (Mann-Whitney).

Table 9.8 Summary statistics and statistical tests by grade and entry.

	A*			A*-A			A*-C			A*-F		
	Early	Not Early	Total									
N	160	100	260	160	100	260	160	100	260	160	100	260
Lower Quartile	-1.8	0.0	-1.1	-4.5	-1.3	-3.7	-5.3	-11.9	-7.4	-1.1	-4.8	-1.9
Median	-0.1	0.0	0.0	-0.8	0.0	0.0	0.3	-6.4	-1.4	0.1	-0.2	0.0
Mean	-0.3	1.1	0.2	-1.3	-0.4	-0.9	0.5	-4.0	-1.2	0.4	-3.5	-1.1
Upper Quartile	0.7	1.1	0.8	1.7	1.4	1.6	5.4	0.0	4.3	1.6	0.8	1.4
SD	2.3	5.1	3.7	4.7	8.3	6.3	7.5	21.5	14.6	3.6	18.5	11.8
Skewness	0.3	3.7	4.2	-0.3	-1.8	-1.5	0.4	0.5	0.2	2.4	-1.4	-2.5
Kurtosis	1.4	19.1	32.1	-0.1	14.9	17.7	0.0	8.7	17.8	15.1	7.5	22.6
Shapiro-Wilk	<0.01	<0.01	<0.01	0.51	<0.01	<0.01	0.1	<0.01	<0.01	<0.01	<0.01	<0.01
Mann-Whitney			<0.01			0.05			<0.01			<0.01

Counts have been rounded to the nearest five.

Given concerns were expressed at a regional level, Table 9.9 shows the same centre variation data with a breakdown by the four regional consortia areas in Wales. The relationship of entry practice to outcomes varies across the regions and at different grades. These inconsistent relationships between entry practice and changes in centre level results suggest that other factors, in addition to entry practice, also influence outcomes at the centre level. However, it should be noted that the variation across centres, as measured by the standard deviation, is large in some cases. This is due to small sample sizes in some of the categories, as well as variation in results across centres.

Table 9.9 Centre variation in best grade results for the Summer 2018 Cohort, by entry practice within consortia area.

Consortium	Entry	N	A*		A*-A		A*-C		A*-F	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	Early	50	0.4	2.4	0.0	3.9	3.7	7.9	1.5	4.6
	Not Early	20	0.7	4.0	0.4	5.3	-2.3	11.2	0.1	20.6
	All	65	0.5	2.9	0.1	4.4	2.0	9.3	1.1	11.7
2	Early	55	-1.0	2.2	-2.0	5.1	0.9	6.6	0.3	1.8
	Not Early	25	3.2	7.3	0.9	9.1	-0.3	25.7	-8.8	24.4
	All	80	0.3	4.8	-1.1	6.7	0.5	15.0	-2.4	14.0
3	Early	25	-0.1	2.4	-2.1	4.5	-3.1	5.6	-0.7	2.7
	Not Early	40	0.0	1.8	-0.2	6.0	-6.2	25.8	0.1	12.0
	All	65	0.0	2.1	-1.0	5.5	-4.9	20.1	-0.2	9.4
4	Early	30	-0.6	1.7	-1.5	4.4	-2.2	7.2	-0.1	4.1
	Not Early	10	-0.2	3.3	-0.5	4.6	-5.2	7.2	-8.2	14.9
	All	40	-0.5	2.3	-1.2	4.4	-3.0	7.3	-2.4	9.4
All	Early	160	-0.3	0.0	-1.3	4.6	0.5	7.4	0.4	3.6
	Not Early	95	1.1	0.0	-0.3	8.2	-4.0	21.4	-3.5	18.4
	All	255	0.2	3.7	-0.9	6.2	-1.2	14.6	-1.1	11.8

Counts have been rounded to the nearest five.

We also conducted an analysis similar to Table 9.9 broken down by local authority. The analysis focused on comparing the average centre variation at grades A*-C in centres who entered early with centres who did not enter early. We have not presented that analysis in full in this report because small sample sizes in some categories raised concerns about the reliability of the statistics and may have allowed individual centres to be identified.

10. Conclusions

Our primary reasons for initiating this review were to investigate:

- whether standards for the qualification have been maintained in line with previous exam series.
- whether students who sat the qualification for the first time in 2018 were unfairly disadvantaged compared to those who sat the qualification in 2017.

Having reviewed this summer's award, we conclude that:

- **the national standard for GCSE English Language was maintained in the summer 2018 award.**
 - The grade boundaries were set by senior examiners based on a careful review of student work.
 - The results for 16-year-olds were in line with the outcomes predicted through statistical analysis.
 - At centre level, it is normal for results to vary from one year to the next. The extent to which centre level results varied in 2018 was similar to 2017.
- **there is no evidence that the GCSE English Language award in summer 2018 unfairly disadvantaged students sitting the qualification for the first time compared to those who sat the qualification in 2017.**
- **we are confident that students' achievements have been fairly recognised. Users can rely on the fact that the grades awarded in summer 2018 are equivalent to those awarded in 2017.**

There are no grounds for reconsidering the GCSE English Language award for summer 2018.

Many of the concerns raised about this summer's GCSE English Language award were based on a view that centres' decisions about when to enter students for exams had impacted on their results. Specifically, there were concerns that results for centres who did not enter students until summer 2018 were lower than expected and lower than those of centres who entered students early and multiple times.

To investigate these concerns, we analysed results by entry practice at a centre level.

We found that there is evidence of a statistical relationship between centre entry practice and results. The relationship is complex and inconsistent:

- **The relationship differs across grades.** On average, centres that enter candidates early and/or repeatedly have slightly lower outcomes at A* and A, and slightly higher outcomes at grade C.
- **For centres that used similar entry strategies, the relationship with results is not consistent.** Some centres who used early entry and multiple entry had higher outcomes while others had lower outcomes. The same was true for centres who entered their candidates at the end of Year 11.
- These inconsistent relationships between entry practice and changes in centre level results suggest that **other factors, in addition to entry practice, also influence outcomes at the centre level.**

Our research into early and multiple entry may offer some insight into statistical relationships between entry practices and average changes in centre level results at grades A* and C. We know entry strategies are often used by centres to maximise attainment at grade C. We also know that some students who achieve a grade C or above in an early attempt do not return to try and improve on their grade, or if they do, are less motivated to do so.

However, there is no evidence to suggest that this relationship is the result of inconsistent standards being applied across different exam series. What this finding does suggest, is that the significant differences and changes in centre entry strategies over recent years do not promote a level playing field for students.

11. Looking forward

The way in which GCSEs are awarded is complicated and difficult to explain in simple terms.

Maintaining standards in GCSE English Language this year has been challenging, not least because of the complexities that early and multiple exam entry bring to the picture. As we reported last year, we believe that the widespread use of early and multiple entry at GCSE poses risks to students and to the system. In this review, we have seen how it creates an uneven playing field for students and undermines confidence in the qualifications system.

The action taken by Welsh Government to discourage widespread early entry has helped to substantially reduce the levels of early entry that we saw this summer. We welcome these developments as they are likely to lead to more stable results in future years that will help to promote fairness for all students and users of qualifications.

Appendix 1: Overview of the new GCSE English Language qualification

Unit 1 assesses speaking and listening skills (referred to as oracy). This unit is assessed through two practical tasks that are taken by students during the course and marked by teachers. Each task is marked out of 40. Teachers' marks are standardised across teachers within the school and externally moderated by WJEC.

One task requires each student to give a researched presentation, the other task is to take part in a structured group discussion with peers. Tasks must be recorded electronically to allow for subsequent standardisation and monitoring to take place. Units 2 and 3 both assess reading and writing skills. Each unit focuses on different kinds of writing. Unit 2 focuses on description, narration and exposition, while Unit 3 focuses on argumentation, persuasion and instructional writing. Unit 2 and Unit 3 are both assessed by a two-hour exam, each marked out of a total of 80 marks.

Each exam is split into two parts, each marked out of 40. The first part of each exam assesses candidates' reading skills and the second part assesses their writing skills. For each of the writing tasks, half of the marks are awarded for the way candidates communicate and organise their thoughts, and the other half are awarded for how accurately they write, including their use of grammar, punctuation and spelling. Unit 2 also includes a proofreading task and requires candidates to proofread and correct a brief text. This task contributes 2.5% towards the final grade.

The qualification specification can be found on the [WJEC website](#).

Appendix 2: Comparable outcomes

The comparable outcomes approach has been used by UK qualifications regulators for several years as the main mechanism to prevent grade inflation over time.

However, the principal aim of the approach is to protect students so that they are not unfairly disadvantaged from being the first to sit new qualifications. This means that if the national cohort, that is all students sitting a qualification, is similar to last year, then we expect the results at a national level to be similar this year. This comparable outcomes approach is tried and tested, and has previously been used successfully by exam boards to maintain standards at a time of qualification change.

If a cohort has not changed much, then we would not expect the results to change much either. It is well established practice when new qualifications are introduced to give priority to comparable outcomes over comparable performance. If the awards were based on judgements of performance, students could be disadvantaged as they are likely to get lower marks due to a lack of familiarity with the new exams. The aim of the comparable outcomes approach is that the first students to take these new exams should be awarded grades that they would have received had they taken the old exams.

In essence, the comparable outcomes approach compensates for a likely small drop in performance when new qualifications are introduced. The approach also helps compensate for the relative advantage that later cohorts have over the first students to take a qualification.

Appendix 3: Grade boundaries

The grade boundary is the minimum mark required to achieve a particular grade. For most GCSEs, the judgemental grade boundaries are A, C and F. These are set by the awarding committee using both the quantitative and qualitative evidence available to them. The other grade boundaries are calculated based on the judgemental grade boundaries that have already been established. Further details can be found in our regulatory document *Requirements for setting specified levels of attainment for GCE and GCSE qualifications*¹⁵.

Example:

If the A boundary is set at 70, C at 54 and F at 33, then the B boundary is halfway between A and C, in this case at 62. The D and E boundaries are placed equally between D and F, in this example at 47 and 40 respectively. The A* boundary¹⁶ is the same distance above A as B is below A, in this example at 78. The G boundary is the same distance below F as E is above F, in this example at 26.

Grade	A*	A	B	C	D	E	F	G
Mark	78	70	62	54	47	40	33	26

¹⁵ www.qualificationswales.org/english/publications/requirements-for-setting-specified-levels-of-attainment-for-gce-and-gcse-qualificationspdf/

¹⁶ There are further rules around the A*/A boundary set out in our regulatory document referenced above.

Appendix 4: The sawtooth effect

When new qualifications are introduced, students in the first few years are likely to attain lower marks in the exams than students taking well-established qualifications in previous years, who were more familiar with the exams. This is followed by improved performance over time as familiarity with the new qualification increases. This pattern of change caused by assessment reform is known as the 'sawtooth effect'. Specifically, performance on high-stakes assessments is often adversely affected when that assessment undergoes reform, followed by improving performance over time as students and teachers gain familiarity with the new test.

In practical terms, we require exam boards to make an adjustment to grade boundaries to compensate for the sawtooth effect. When grade boundaries are set, evidence will be considered to see if an adjustment is needed to allow for this effect. Where there is evidence to support the presence of the sawtooth effect, grade boundaries are likely to be lowered, to compensate for the dip in performance of those students sitting new qualifications. As qualifications become more established, exam performance tends to improve. This does not necessarily happen to the same extent across all grades, it is more likely to happen because candidates get better at gaining marks on more straightforward and accessible questions.

This effect is understandable. With each exam series, more past exam papers and mark schemes become available, together with exemplar materials based on genuine student work, and examiner reports explaining how students performed on each question and topic area. In other words, teachers become more familiar with how the qualification is assessed, and students have more opportunities to practise on past papers. To make sure that students in later cohorts are not given an unfair advantage compared to students sitting their exams when a qualification is first introduced, grade boundaries are likely to be increased, to compensate for the improvement in performance due to familiarity rather than evidence of real improvements in attainment.

Ofqual has carried out some research into the sawtooth effect in GCSE and AS and A level assessments which can be found at:

<https://www.gov.uk/government/publications/investigation-into-the-sawtooth-effect-in-gcses-as-and-a-levels>

Appendix 5: Statistical prediction methodology

Common Centres Predictions

Statistical predictions based on prior attainment are used for GCSE and A/AS level awarding across England, Wales and Northern Ireland.

In Wales, 'Common Centres (CCs)' predictions are typically used for GCSEs. CCs predict the results of current candidates in a subject using the results achieved by previous candidates in the same subject in a reference year, using centres that entered candidates at both points in time. It is assumed that any factors leading to a more positive or negative set of results within individual centres (compared to the reference year) will average out to produce an accurate national level prediction.

The results for matched candidates, who are the entries from the common centres in the current series, are then compared to the predicted results. Regulatory rules state that we require more justification for actual results that are outside of the tolerances set around the predicted results for these candidates.

Factors such as the first award of a new specification and large changes in entry behaviour necessitate consideration of the most appropriate prediction approach for the purpose of guiding awarding and maintaining standards over time.

Predictions for reformed GCSEs in English, Welsh and Mathematics subjects in Wales

The prediction approach for reformed GCSEs in English, Welsh and Mathematics subjects in Wales in summer 2018 was consistent with the approach used in summer 2017 to manage the transition of the standard from the legacy qualifications. The summer 2017 approach used best-grade common centre predictions, which had been chosen to ensure that a national standard was carried forward in the context of early and repeat entry in the legacy qualifications.

During preparations for the summer 2018 series, we considered whether the best grade prediction approach was still appropriate given the extent of early and repeat entry in the 2018 Cohort.

In the case of reformed English Language GCSE, the two cohorts of 16-year-old students completing Year 11 by June 2018 had experienced very different patterns of entry.

The 2017 Cohort: 16-year-olds during the June 2017 series. These candidates could only enter the reformed GCSE once (in the June 2017 series) by the end of Year 11, as this was the first time the qualification was awarded. The best grades achieved by this cohort in reformed English Language were therefore also the first grades they had achieved.

The 2018 Cohort: 16-year-olds during the June 2018 series. These candidates could enter up to three times by the end of Year 11: In the June 2017 series (when they were in year 10), as resitters in the November 2017 series (if they had sat in June 2017), and in the June 2018 series.

We chose to maintain the consistent best grade common centres prediction approach across the reformed GCSEs in English, Welsh and Mathematics subjects. This was written into the Data Exchange document for the summer 2018 series as:

WJEC must create for each specification predictions for the Wales cohort of 16-year-old matched cash-in students. The predictions must be based on the data from common centre analysis of the best grade achieved by candidates from November 2015 to summer 2017 (as applicable) and applied to the combined outcomes for the November 2016 to summer 2018 cohorts (as applicable).

We judged that a consistent best grade prediction approach would be the most appropriate way of ensuring that a comparable national standard would be maintained. This decision was informed by the varying early and repeat entry practice across and within centres, as well as the small proportions of 16-year-olds in the 2018 Cohort sitting for the first time in summer 2018 (only 30.1% in English Language GCSE).

The steps in the Common Centres method applied to reformed GCSEs in English, Welsh and Mathematics subjects are:

- calculate the best grade achieved by all candidates who are 16-years-old in June 2017, excluding partial absences¹⁷
- calculate the cumulative grade distribution for each common centre (centres that have entered candidates in both series)

¹⁷ Candidates that received a grade but did not sit all of the units for the qualification. Predictions are based only on candidates that sat all of the assessments in the reference year.

- for each common centre, use the cumulative grade distribution and the number of candidates who are 16-years-old in June 2018 in that centre to calculate the expected numbers of candidates achieving each grade.
- sum the expected numbers of candidates achieving each grade across all common centres to obtain the expected number of candidates achieving each grade at the national level.
- use the total entry of candidates who are aged 16 in June 2018 to calculate the cumulative grade distribution on a percentage basis. For the purpose of determining whether the award was out of tolerance, these predicted results were compared to the best grade results achieved by the June 2018 Cohort across all their previous entry attempts.

Appendix 6: Glossary

Assessment – The evaluation, test, measurement and/or documentation of knowledge, skills or understanding.

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

Awarding body – Also referred to as an exam board. An organisation recognised by us for the purpose of awarding regulated qualifications, such as WJEC.

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.

Best grade – the highest grade achieved by a student across all their attempts.

Centre – An organisation (such as a school, college or work-based learning provider) that enters candidates for qualifications and/or delivers assessments on behalf of an awarding body.

Chair of Examiners – An individual responsible to the awarding body for maintaining standards in a qualification 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 year to year. The chief examiner also acts as a Principal Examiner or Moderator for at least one component.

Cohort – A group of candidates that shares a meaningful characteristic, such as age, for administrative and/or statistical purposes.

Common centres – A centre that has entered students for a subject in two different points in time.

Common centres approach – Using statistical comparisons to predict likely examination results year on year, based on the assumption that - all else being equal - a centre's results are unlikely to be very different across two years. This is explained in more detail in Appendix 5.

Comparable outcomes – The main mechanism to prevent an unfair variation in awarding standards at the national level. It uses scrutiny of student work alongside statistical evidence across a cohort to calculate any necessary adjustments to grade boundaries. This is explained in more detail in Appendix 2.

Data Exchange - A regulatory document outlining the procedures which exam boards use to award the qualifications and what data that they provide to us during the examination series.

Early/multiple entry – The practice of entering candidates early or multiple times to achieve a grade. The definition of 'early' is relative to the design of the qualification.

The analysis presented in this report defined an early entry centre as one which entered at least 10 candidates who turned 16 years old in the 2017/18 academic year before June 2018. This definition was chosen as early entry practices at centre level tended to involve substantially more than 10 candidates.

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, such as WJEC.

Examiners – Individuals with subject expertise who are responsible for marking candidates' responses.

GCSE (General Certificate of Secondary Education) – Designed to be 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.

Grade boundary – the minimum mark that a candidate needs to score to be awarded a particular grade for a unit or qualification.

Linear qualification – Where all examination components are taken in one examination series, usually at the end of the course.

Mark scheme – A scheme detailing how marks are 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.

Monitoring – Activities that Qualifications Wales carries out to review the quality of an awarding body's systems, processes and/or qualifications.

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

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

Reformed GCSEs – GCSEs in a number of subjects have been redesigned and introduced in three waves, for first teaching in 2015, 2016 and 2017.

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.

Series – The period of time when a set of exams is taken by the same cohort.

Sawtooth effect – The change of pattern caused by assessment reform, where students sitting the first few years of a reformed qualification are likely to get lower marks than students in previous years, who were more familiar with exams. Subsequent years are likely to see a rise in performance. Grade boundaries are adjusted to minimise this risk. This is explained in more detail in Appendix 4.

Scaled marks – The conversion of raw marks into a scale to accurately weight the contribution of an assessment unit or component towards the overall grade of a

qualification. This allows for exam papers or assessment tasks to be designed to be marked out of an optimum range of marks.

Tolerance limits – These are set around predicted qualification outcomes taking into account the number of entries. They are a trigger for exam boards to report additional evidence to the regulator, to justify the award in cases where the actual results differ from predicted results by more than the applicable tolerance.