Inter-board comparability of grade standards in GCSEs and A levels

Summer 2015

Contents

Introduction

Our evaluation of grading standards in summer 2015 shows that there is no discernible advantage or disadvantage in entering students for different exam boards in a subject. We explain the evidence and our rationale below.

One of our aims in monitoring the GCSE, AS and A level awards each summer is to make sure so far as possible that there is a level playing field for the students. One of the ways that we do this is to make sure that the grade standards are comparable, so that it is no easier or more difficult to get a particular grade in a subject with one exam board than with another.

During July and August we monitored the outcomes of 743 GCSE, AS and A level awards. Of these, 51 were outside the reporting tolerances, which meant that the exam boards provided additional information to us to support their decisions. In all cases we accepted the additional evidence provided. We have already published more detail on the subjects where one or more exam boards were out of tolerance.[1]

Across all awards we concluded that the exam boards had maintained appropriate standards and that grade standards within a subject were in line across exam boards.

This report explains more about how we judge comparability within a subject. The accompanying spreadsheets provide a comparison of each board's provisional outcomes compared with the predictions, for all subjects offered by more than one exam board.

What do we mean by comparability?

There are many different ways of considering the comparability of qualifications. In this paper we report on the comparability of grade standards, within individual subjects, in GCSEs, AS and A levels awarded in summer 2015. We monitor this closely and our aim is that, all things being equal, a student should receive the same grade in a subject, for example GCSE geography, regardless of which exam board they entered with.

Our monitoring of summer awarding does not consider the comparability of the content that must be studied. In GCSEs and A levels this is generally achieved

---

[1] https://www.gov.uk/government/publications/out-of-tolerance-report-gcse-as-and-a-level-summer-2015

through the rules we put in place about the subject content[2] that must be included in all qualifications with the same title.

Our monitoring also does not consider how demanding the assessments are. One of the reasons that we accredit GCSEs and A levels is so that we can judge whether they are sufficiently demanding before allowing the qualification to be offered. We also monitor the demand of qualifications in other ways once they are operational.

There are different ways to measure comparability of grade standards. We use statistical predictions to judge the comparability of grade standards across all exam boards in a subject. Where all boards' results are reasonably close[3] to their predictions, we judge that their grade standards are aligned, and that it is no more easy or difficult to get a particular grade with one board than with another.

This report does not discuss comparability between different subjects. We have already published separate information on this topic.[4]

The principle of comparable outcomes

Exam boards have always used statistics, alongside senior examiner judgement, to guide their decisions about grade standards. These decisions are based on consideration of different sources of evidence – including current and past student work, data about the prior attainment of this year's students compared with previous years, and recommendations from the senior examiners. In recent years, Ofqual and the regulators in Northern Ireland and Wales have required exam boards to use these statistics in a consistent way, an approach which has become known as 'comparable outcomes'. But the principle of comparable outcomes pre-dates Ofqual's existence.

When qualifications change, it is likely that students in the first year will perform less well than their predecessors: their teachers will be less familiar with the new qualifications and there will be fewer past papers and other support materials available. In 2001 and 2002 discussions between exam boards and regulators about new 'Curriculum 2000' A levels were focused on how best to avoid disadvantaging the first cohort to sit these new A levels. A decision could have been made to prioritise what Cresswell (2003)[5] referred to as 'comparable performance' –

---

[2] For example: https://www.gov.uk/government/collections/gcses-9-to-1-requirements-and-guidance
[3] We define this as being within the published tolerance of plus or minus 1, 2 or 3 percentage points (depending on entry size) of the prediction.
[4] https://www.gov.uk/government/news/can-different-gcse-and-a-level-subjects-be-compared-accurately
[5] Cresswell, M.J. (2003) *Heaps, prototypes and ethics: the consequences of using judgements of student performance to set examination standards in a time of change*. University of London Institute of Education

expecting students to demonstrate the same level of knowledge, skills and understanding as in previous years in the previous qualifications. Or we could have prioritised 'comparable outcomes' – roughly the same proportions of students achieving each grade as in previous years, providing the cohort of students is similar to previous years.

The regulators agreed that exam boards should prioritise comparable outcomes, to avoid disadvantaging the students who were the first to sit these new qualifications. This approach has also been used for new A levels since 2010 and for new GCSEs since 2011 and 2012.

Newton (2011)[6] argues that exam boards have always based their awarding decisions on the idea that if the cohort of students taking an exam hasn't changed much, then we wouldn't expect the proportions of students achieving each grade to change much either. He refers to this as the 'similar cohort adage'.

Setting grade boundaries

Awarding committees

Grade boundaries are set once the exam scripts have been marked. Exam boards convene awarding committees for each subject[7] to recommend minimum marks at key grade boundaries[8] for each unit of the qualification. The committees also include the chief examiner, principal examiners and principal moderators, and may also include exam board technical experts.

Predictions

Exam boards have always used statistics to guide their decisions about where to set grade boundaries. In recent years these have become more sophisticated, taking account of the prior attainment of the cohort, and the regulators require exam boards to use those statistics in a consistent way. Predictions are based on the relationship between prior attainment and national results in a reference year. Exam boards use prior attainment at Key Stage 2[9] when predicting GCSE outcomes, and prior attainment at GCSE when predicting AS and A level outcomes.

---

[6] Newton, P. *A level pass rates and the enduring myth of norm referencing.* In *Research Matters* (October 2011), Cambridge Assessment
[7] In some cases and where different qualifications share the same assessments (for example the GCSE science suite) an awarding committee might cover a number of related qualifications.
[8] A*, A, C and F for GCSE, A and E for AS and A*, A and E for A level
[9] Key Stage 2 tests are only taken in England. Where the entry for a qualification is predominately from Northern Ireland or Wales, CCEA and WJEC use a prediction based on the overall performance of schools that have taken the qualification in previous years (referred to as 'common centres')

Predictions provide a common basis for all exam boards to us and so give us a way to compare grade standards across boards. Each board's prediction is based on the same national results but reflects the prior attainment profile that board's entry.

For example, the predictions used for GCSE English in summer 2015 were based on the relationship between GCSE results achieved by year 11 students in the academic year 2013/14 and KS2 results for those students in 2009 (that is, when they were in year 6).

Research evidence to support the use of predictions

We have commissioned research to evaluate the effectiveness of the prediction methodology used for GCSEs[10] and A levels[11]. The research supports the approach taken. Small changes were suggested to improve the effectiveness of the predictions but these have to be balanced against the disadvantages of additional complexity. After considering the risks with exam boards, we decided not to make the recommended changes as the likely benefits were outweighed by the potential risks of implementing a new approach.

What exam boards do when predictions are less reliable

Predictions are most reliable when they are based on larger numbers of students. With smaller numbers of students, they are less reliable. We and the exam boards take that into account. Our reporting tolerances are wider for qualifications with relatively small entries, and we do not set a reporting tolerance for qualifications with an entry of 500 or less. Where the entry numbers are relatively small, exam boards will balance the use of statistics with the judgements of their senior examiners.

Matched entry

The data in the accompanying tables includes details of the total entry for each specification, as well as details of the matched entry. The matched entry is a subset of the total entry, comprising students who can be matched to the relevant prior attainment measure and who are in the target age group for the qualifications (16 for GCSE, 17 for AS and 18 for A level[12]).

---

[10] Benton, T., and Sutch, T. (2014) Analysis of use of Key Stage 2 data in GCSE predictions. Coventry: Ofqual. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/429074/2014-06-16-analysis-of-use-of-key-stage-2-data-in-gcse-predictions.pdf

[11] Benton, T., and Lin, Y. (2011) Investigating the relationship between A level results and prior attainment at GCSE. Coventry: Ofqual. Available at: https://www.gov.uk/government/publications/investigating-the-relationship-between-a-level-results-and-gcses

[12] Age at the end of the relevant academic year

For GCSE it includes only those year 11 students who can be matched to their KS2 prior attainment[13]. For A level it includes only those year 13 students who can be matched to their GCSE prior attainment.

The predictions used by exam boards to guide their decisions include only the matched students. The other students may be those who did not sit KS2 or GCSEs or who are in a different year group (students in year 10 taking GCSEs early, or adult learners, for example).

Exam boards will make decisions about grade boundaries and outcomes for the cohort on the basis of the matched students. Figure 1 below shows how this might work. Figure 1 shows the cumulative percentage of students on each mark that might have been considered as the grade C boundary. The cumulative percentage of **matched** students at a mark of 34 (that is, the proportion of students scoring 34 or more) is the mark that would most closely meet the prediction for the matched entry.

Grade boundary decisions based on matched students apply to all students. In the example below, the exam board set the grade boundary at 34 on the basis of the matched students, but that grade boundary also applies to the unmatched students. The same standard is therefore applied to all students, but the cumulative percentage will be different when the grade boundary decision is applied to all students.

---

[13] As previously mentioned, for Northern Ireland and Wales the prediction would be based on previous performance of common centres if the majority of the entry is from Northern Ireland or Wales

Figure 1 Example of matched entries

prediction	Mark		Matched students cumulative %	All students cumulative %
	36		60.3	57.3
	35		64.6	59.9
65.2%	34	C	64.9	60.4
	33	D	65.9	61.2
	32		67.2	65.3

Reporting tolerances

Exam boards send us the results for each individual specification as it is awarded. The accompanying data tables include some of the data they send us.

Reporting tolerances are set according to the size of the matched entry. Where the entry is 500 or less, no tolerance is applied.

If the results for the matched entry are within the reporting tolerances set out below, exam boards are not required to explain those results.

Table 1 Reporting tolerances (used at grade A for AS and A level, and grades A and C at GCSE)

Matched entry	Reporting tolerance
500 or less	no tolerance applied
501 – 1000	+/-3 percentage points
1001 – 3000	+/-2 percentage points
3001 +	+/- 1 percentage point

There is also a reporting tolerance at A* in A level and GCSE. The reporting tolerance for A* is +/-2 percentage points. Where the entry is small[14], no tolerance is applied to A*.

---

[14] 100 or fewer matched students at cumulative grade A

Ofqual's role

During the summer Ofqual collects data from the exam boards in England, Wales and Northern Ireland on behalf of Qualifications Wales, the regulator in Wales, and CCEA regulator in Northern Ireland.

Each year the regulators and exam boards agree the bases of the predictions to be used to guide awarding decisions. These are described in the data exchange procedure, which is a regulatory document and published on our website each year.[15]

The data exchange procedure sets out the reporting tolerances which the exam boards must apply when sending data to us. If the results differ from the prediction by more than the tolerance, then they must provide additional evidence to support their decisions.

We review the data from every GCSE, AS and A level award, and we pay particular attention to awards where one or more grade boundaries are out of tolerance. We can, and do, challenge exam boards where we do not believe the supporting evidence is sufficiently strong, and we publish a summary of the out of tolerance awards at the end of the summer.[16] If necessary, we can direct an exam board to set different grade boundaries, to achieve comparability over time and between board, although to date we have not needed to do this.

Notes

Data

We have published separate spreadsheets to show the comparisons between predictions and each exam board's outcomes. We have only included in the spreadsheets those subjects which are offered by more than one exam board.

Note that the data we review is not complete, as exam boards are still processing some marks at the time of awarding. The outcomes in the spreadsheets are only for the matched students, who are a subset of the overall entry. Therefore, the outcomes reported in the attached spreadsheets will be different from any final published outcomes for the specifications. In some cases, exam boards will carry out a check on the provisional outcomes, once they are closer to the data being

---

[15] The summer 2015 data exchange procedure is available at
http://webarchive.nationalarchives.gov.uk/20150710183905/https://www.gov.uk/government/publications/data-exchange-procedures-for-a-level-gcse-level-1-and-2-certificates
[16] Summer 2015 out of tolerance summary is available at
https://www.gov.uk/government/publications/out-of-tolerance-report-gcse-as-and-a-level-summer-2015

complete. They are not required to provide us with updated outcomes for the matched students, unless there is a significant change.

GCSE English/English language and mathematics

These qualifications have seen considerable changes in entries in recent years, as a result of changes to the qualifications themselves, and in particular the move to linear from 2014 onwards in England.

In the English qualifications, we have seen schools moving from English to English language and vice versa, fewer early entries and a decline in entries as some school move to international GCSEs. There are also different versions of the qualifications for Northern Ireland and Wales.

In mathematics, the shift has largely been from the modular specifications to the linear versions, given that the assessments in the modular specifications must now all be taken at the end of the course. We have also seen a reduction in early entry and in students being entered for more than one GCSE in mathematics.

These shifts in entries can make it difficult for exam boards to meet predictions at the level of individual syllabuses. The regulators have agreed, therefore, that exam boards should focus on the overall prediction for combined English/English language and combined linear/modular mathematics, in relation to combined predictions. As a result, outcomes for individual specifications in some cases are outside tolerance, but the overall combined outcomes were within tolerance of the combined prediction.

GCSE science suite

Many of the exam boards offer a suite of qualifications in which the assessments are shared between different titles. For example, a biology unit might be part of a GCSE in science, and part of a GCSE in biology. Where the assessments are the same, the grade boundaries for any one exam series must also be the same, in order to be fair to students.

This complexity can make it more challenging to set grade boundaries in the suite of science subjects so that all boundaries ware within tolerance of predictions. Exam boards have to consider the comparability of standards across the different titles, and it is therefore sometimes the case that one or more boundaries are outside tolerance. We have accepted a number of these, where the exam board provides assurance that the standards across the suite of qualifications is appropriate.

A*

The reporting tolerance at A* is fixed at +/-2 percentage points.[17] Because the A* grade is at the very top end, there are often very few students on the marks near the proposed grade boundaries. Small changes to grade boundaries can therefore have larger changes on the percentage of students achieving A*. As a result, the A* boundary can be more likely to be out of tolerance.

---

[17] where there are at least 500 matched students and where the cumulative number of matched students at grade A is more than 100

The diagram below explains the headings used in the spreadsheets. There are separate tables for A level and GCSE. The A level data includes grades A* and A, and the GCSE data includes grades A*, A and C. The data on the proportion of students achieving each grade (outcomes) are cumulative, so the outcomes for grade A should be read as A and above.

Table 2 Guide to accompanying spreadsheets

These columns identify the exam board and syllabus/specification code

Total entry is all students taking the qualification. Matched entry is only those students with matched prior attainment data (previous GCSE or KS2 results). Subsequent columns focus on matched entry only.

The reporting tolerance varies according to the size of the matched entry – plus or minus 1, 2 or 3 percentage points (and no tolerance for 500 students or fewer). Different boards will often have different tolerances, because of their different entry sizes.

This columns shows whether one or more grade boundaries was out of tolerance.

Exam board	Specification code	Total entry	Matched entry	Outcome at award (%)	tolerance	Difference from prediction	Outcome at award (%)	tolerance	Difference from prediction	Out of tolerance?
A level Biology				Grade A*			Grade A (cumulative)			
AQA	2411	24,311	17,577	10.2	+/- 2%	0.9	30.2	+/-1%	-0.2	No
CCEA	A1012	2,696	2,351	10.8	+/- 2%	1.0	35.7	+/-2%	3.7	Yes
OCR	H421	21,154	13,633	10.1	+/- 2%	1.0	29.1	+/-1%	-0.9	No
Pearson	9BI01	6,578	4,925	8.8	+/- 2%	1.2	25.9	+/-1%	-0.2	No
WJEC	3071/01	5,161	4,069	11.8	+/- 2%	1.9	32.4	+/-1%	0.5	No

For each of the key grades, the outcome at award is the percentage of the matched entry achieving that grade

This shows how far away each board was from its own prediction, based on its matched entry. If the difference in this column is greater than the tolerance, the board will have reported an out of tolerance award and submitted further evidence to explain the decisions.

The CCEA award is out of tolerance at grade A. The tolerance is +/-2% and the award is 3.7 percentage points above prediction.

We wish to make our publications widely accessible. Please contact us at publications@ofqual.gov.uk if you have any specific accessibility requirements.

OGL

This publication is available at www.gov.uk/ofqual.

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

Office of Qualifications and Examinations Regulation

Spring Place   2nd Floor
Coventry Business Park   Glendinning House
Herald Avenue   6 Murray Street
Coventry CV5 6UB   Belfast BT1 6DN

Telephone   0300 303 3344
Textphone   0300 303 3345
Helpline   0300 303 3346