

Inter-board comparability of grade standards in GCSEs, AS and A levels 2017



December 2017

Ofqual/17/6330

Contents

Introduction	3
What do we mean by comparability?	3
The principle of comparable outcomes.....	4
Setting grade boundaries	5
Awarding committees.....	5
Predictions.....	5
Research evidence to support the use of predictions.....	6
What exam boards do when predictions are less reliable	6
Matched entry.....	6
Reporting tolerances.....	8
Ofqual's role.....	9
Notes.....	10
Data	10
GCSE English/English language and mathematics	10
GCSE science suite	10
A*	10

Introduction

We monitor GCSE, AS and A level awards each summer to make sure, so far as possible, that there is a level playing field for students. One of the ways that we do this is to make sure that the grade standards are comparable, so that it is not easier or more difficult to get a particular grade in a subject with one exam board than with another.

We monitor the outcomes of several hundred GCSE, AS and A level awards during July and August. Typically, a small number are outside the reporting tolerances, which means exam boards must provide additional information to us to support their decisions. In some cases, we accept the additional evidence provided; in other cases, we do not. Occasionally, we can require a board to change a grade boundary, in order to bring it into line with the grade standards of other boards in that subject. Each year we publish more detail on the subjects that were out of tolerance, including the reasons provided by the exam board¹.

Across all awards in summer 2017, we concluded that the exam boards had maintained appropriate standards and that grade standards within each subject were aligned.

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 predictions², 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. Each year, we report on our monitoring of standard setting in GCSEs, AS and A levels³. 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, AS and A levels this is generally achieved

¹See for example <https://www.gov.uk/government/publications/monitoring-gcse-as-and-a-level-exams-summer-2017>

² <https://ofqual.blog.gov.uk/2017/04/21/prediction-matrices-explained/>

³ <https://www.gov.uk/government/publications/ofqual-2017-summer-exam-series-report>

through the rules we put in place about the subject content⁴ 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, AS 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 close⁵ to their predictions, we judge that their grade standards are aligned, and therefore it is not easier or more 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 topic⁶.

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 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 years will perform less well than their predecessors⁷; 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

⁴ For example: <https://www.gov.uk/government/collections/gcse-9-to-1-requirements-and-guidance>

⁵ In general, 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. However, there may be instances where we judge that an exam board is out of line with others, even though its award is within tolerance.

⁶ <https://www.gov.uk/government/news/can-different-gcse-and-a-level-subjects-be-compared-accurately>

⁷ <https://www.gov.uk/government/publications/investigation-into-the-sawtooth-effect-in-gcse-as-and-a-levels>

prioritise what Cresswell (2003)⁸ referred to as ‘comparable performance’ – 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 at the time agreed that exam boards should prioritise comparable outcomes, to avoid disadvantaging the students who were the first to sit these new qualifications.

This is the approach we took for new AS in 2016 and for new GCSEs, AS and A levels in 2017. This approach was also used for the last round of new A levels in 2010 and new GCSEs in 2011 and 2012.

Newton (2011)⁹ 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 proportion 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 subject¹⁰ to recommend minimum marks at key grade boundaries¹¹ for each unit of the qualification. The committees include senior examiners/question writers as well as exam board staff.

Predictions

Exam boards have always used statistics to guide their decisions about where to set grade boundaries. These have become more sophisticated over time, taking account of the prior attainment of the cohort, and the regulators require exam boards to use these statistics in a consistent way. Predictions are based on the relationship between prior attainment and national results in a reference year. Exam boards use

⁸ 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

⁹ Newton, P. *A level pass rates and the enduring myth of norm referencing*. In *Research Matters* (October 2011), Cambridge Assessment

¹⁰ 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.

¹¹ A*, A, C and F or 9, 7, 4 and 1 for GCSE, A and E for AS and A*, A and E for A level

prior attainment at Key Stage 2¹² when predicting GCSE outcomes, and prior attainment at GCSE when predicting AS and A level outcomes.

Predictions provide a common basis for all exam boards to use 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 of that board's students.

For example, the predictions used for GCSE mathematics in summer 2017 were based on the relationship between GCSE results achieved by year 11 students in summer 2016 and key stage 2 results for those students in 2011 (that is, when they were in year 6).

Research evidence to support the use of predictions

We have previously commissioned research to evaluate the effectiveness of the prediction methodology used for GCSEs¹³ and A levels¹⁴. The research supported 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 small changes suggested 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 large numbers of students. With smaller numbers of students, they are less reliable. We and the exam boards take that into account. Where we use reporting tolerances, they 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

¹² 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')

¹³ 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

¹⁴ 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-gcse>

attainment measure and who are in the target age group for the qualifications (16 for GCSE, 17 for AS and 18 for A level¹⁵).

For GCSE it includes only those year 11 students who can be matched to their key stage 2 prior attainment¹⁶. For A level it includes only those year 13 students who can be matched to their GCSE prior attainment, and for AS it includes those year 12 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 key stage 2 or GCSEs or who are in a different year group (students in year 10 taking GCSEs early, or adult learners, for example).

Exam boards 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. It shows the cumulative percentage of students on each mark that might have been considered as the grade C/4 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.

¹⁵ Age at the end of the relevant academic year.

¹⁶ 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 %
65.2%	36	60.3	57.3
	35	64.6	59.9
	34 C/4	64.9	60.4
	33 D/3	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 provide additional evidence to explain those results.

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

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 small¹⁷, no tolerance is applied to A*.

¹⁷ 100 or fewer matched students at cumulative grade A.

Setting standards in new qualifications

In the first awards of reformed GCSEs, AS and A levels this summer, we said that exam boards would rely more heavily on statistics to set standards, so that students were not disadvantaged by being the first to sit these new qualifications. As a result, we did not set reporting tolerances for the first awards of the new qualifications, or the second awards of those AS qualifications that were first awarded in 2016¹⁸. We asked exam boards to set grade boundaries so that they met predictions as closely as possible.

Ofqual's role

Ofqual collects data from the exam boards in England, Wales and Northern Ireland on behalf of Qualifications Wales, the regulator in Wales, and CCEA, the 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 that is published on our website each year¹⁹.

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 summer²⁰. If necessary, we can direct an exam board to set different grade boundaries to achieve comparability over time and between boards, although to date we have not needed to do this.

18

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/377771/2014-09-12-board-paper-for-new-gcses-in.pdf

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474099/08_15_-_Awarding_Reformed_AS_and_A_Levels.pdf

¹⁹ The summer 2017 data exchange procedure is available at:

<https://www.gov.uk/government/publications/data-exchange-procedures-for-a-level-gcse-level-1-and-2-certificates>

²⁰ Summer 2017 out of tolerance information is included in:

<https://www.gov.uk/government/publications/summer-2016-gcse-as-and-a-level-a-summary-of-our-monitoring>

Notes

Data

We have published separate spreadsheets to show the comparisons between predictions and each exam board's outcomes. We have only included subjects in the spreadsheets that 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 complete. They are not required to provide us with updated outcomes for the matched students, unless there is a significant change.

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 outcomes at all grade boundaries are within tolerance of predictions. Exam boards have to consider the comparability of standards across the different titles, and therefore it is 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 points²¹. 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.

²¹ Where there are at least 500 matched students and where the cumulative number of matched students at grade A is more than 100. For unreformed qualifications only.

The diagram below explains the headings used in the spreadsheets. There are separate tables for AS, 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.

These columns identify the exam board, subject 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: +/- 1, 2 or 3 percentage points (and no tolerance for 500 students or fewer). Different boards will often have different tolerances in a subject, because of their different entry sizes.

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

Exam board	Subject	Specification number	Total entry	Matched entry	Tolerance at A* grade	Grade A* Actual cum. %	Difference from prediction	Tolerance at grade A	Grade A Actual cum. %	Difference from prediction	Out of tolerance Y/N
AQA	Art and Design (all)	2204D, 2203C, 2205E, 2202B, 2201A, 2206F	20,265	15,381	+/-2%	12.4	0.7	+/-1%	28.1	1.2	Yes
CCEA	Art and Design (all)	A3512	733	672	+/-2%	12.9	-2.2	+/-3%	36.3	2.6	Yes
OCR	Art and Design (all)	(H560- H566)	6,112	5,111	+/-2%	12.2	-0.5	+/-1%	29.4	0.5	No
Pearson	Art and Design (all)	9AD01toPY01	11,785	9,912	+/-2%	13.7	0.9	+/-1%	27.9	-0.9	No
WJEC	Art and Design (all)	300101, 301101, 302101, 303101, 304101, 305101	3,589	3,119	+/-2%	10.4	-0.6	+/-1%	25.8	0.2	No

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

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 AQA award is out of tolerance at grade A. The tolerance is +/-1% and the award is 1.2 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.



© Crown copyright 2017

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit <http://nationalarchives.gov.uk/doc/open-government-licence/version/3> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: publications@ofqual.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

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
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346