

RESEARCH AND ANALYSIS

Inter-board comparability of grade standards in GCSEs, AS and A levels 2018

ofqual

Contents

Introduction	3
<i>Background</i>	<i>3</i>
How grade boundaries are set.....	3
The use of statistics in awarding	3
What we mean by comparability.....	4
<i>Interpreting the data tables</i>	<i>5</i>
Matched entry.....	5
Reporting tolerances.....	7
Setting standards in the first awards of new qualifications	8
Ofqual's monitoring role	8
Data	8
Grade A*	9

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 our 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. 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 2018, 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.

Background

How grade boundaries are set

Grade boundaries are set once the exam scripts have been marked. Exam boards convene awarding committees for each subject (or related suite of subjects). These committees include senior examiners/question writers as well as exam board staff. The committee will recommend the minimum mark at each key grade boundary³. In linear qualifications, grade boundaries are set at the overall subject level, although exam boards also publish notional component grade boundaries.

The use of statistics in awarding

Exam boards have always used statistics, alongside senior examiner judgement, to guide their decisions about where to set grade boundaries. These decisions are based on considering 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.

¹See for example, [Monitoring GCSE, AS and A level exams](#)

²[Prediction matrices explained](#)

³A*, 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

Statistics take the form of predictions for the proportion of students likely to achieve a certain grade, for example grade 4 and above. Predictions are based on the relationship between prior attainment and national results in a reference year. Exam boards use prior attainment at key stage 2⁴ to predict GCSE outcomes, and prior attainment at GCSE to predict AS and A level outcomes. For example, the predictions used for GCSE maths in summer 2018 were based on the relationship between GCSE results achieved by year 11 students in summer 2017 and key stage 2 results for those students in 2012 (that is, when they were in year 6).

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

There are several advantages in using predictions to guide awarding decisions.

- Using statistical predictions in a consistent way gives us the best tool to make sure that grade standards between different exam boards are aligned.
- When qualifications change, it is more difficult for senior examiners to judge the quality of student work. In general, students in the first cohorts are likely to perform less well than their predecessors⁵, because teachers are less familiar with the new qualifications and there are fewer past papers and other support materials available.

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 a matched 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.

We have commissioned research to evaluate the effectiveness of the predictions used for GCSEs⁶, AS and A levels⁷. The research supports the approach taken.

What we mean by comparability

There are many different ways of considering the comparability of qualifications. Our focus in this report is **comparability between exam boards in a subject**, for

⁴ Key Stage 2 tests are only taken in England. Where the entry for a qualification is predominantly 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')

⁵ [Investigation into the sawtooth effect in GCSEs, AS and A levels](#)

⁶ Benton, T., and Sutch, T. (2014) [Analysis of use of Key Stage 2 data in GCSE predictions](#). Coventry: Ofqual. Available at:

⁷ Benton, T., and Lin, Y. (2011) [Investigating the relationship between A level results and prior attainment at GCSE](#). Coventry: Ofqual.

example GCSE geography. Our aim is that, all things being equal, a student should receive the same grade in a subject, regardless of which exam board they entered with. Each year, we report on our monitoring of standard setting in GCSEs, AS and A levels⁸.

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 through the rules we put in place which set out the subject content⁹ 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 close¹⁰ 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 published separate information on this topic¹¹.

Interpreting the data tables

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 only those 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 level¹²).

For GCSE it includes only those year 11 students who can be matched to their key stage 2 prior attainment¹³. For A level it includes only those year 13 students who

⁸ See <https://www.gov.uk/government/publications/ofqual-2018-summer-exam-series-report>

⁹ For example: [GCSEs \(9 to 1\): requirements and guidance](#)

¹⁰ 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.

¹¹ [Reviewing Inter-Subject Comparability](#), November 2018

¹² Age at the end of the relevant academic year.

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.

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. They do this because the matched students generally make up the largest part of the cohort and they tend to be most stable.

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 of students achieving a particular grade will be different when the grade boundary decision is applied to all students.

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 4/C	64.9	60.4
	33 3/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 provide additional evidence to explain those results.

Table 1 Reporting tolerances (used at grade A for AS and A level, and grades 7/A and 4/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 legacy (unreformed) A level and GCSE. The reporting tolerance for A* is +/-2 percentage points. Where the entry is small¹⁴, no tolerance is applied to A*.

¹⁴ 100 or fewer matched students at cumulative grade A.

Setting standards in the first awards of 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 GCSE, AS and A level qualifications that were first awarded in 2017¹⁵. We asked exam boards to set grade boundaries so that they met predictions as closely as possible. In the third year onwards of the new qualifications (ie for AS qualifications first awarded in 2016), we expect senior examiner judgements about the quality of work to play a greater part in the decisions, and so the reporting tolerances were applied.

Ofqual's monitoring role

Each year, we set rules about the bases of predictions and how they are to be used in awarding. These are described in the data exchange procedure, which is a regulatory document that is published on our website each year¹⁶. During the awarding period, we collect data from the exam boards offering GCSEs, AS and A levels in England.

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 summer¹⁷. 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.

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

¹⁵ [Board paper for new GCSEs](#), September 2014
[Awarding reformed AS and A levels](#), board paper September 2015

¹⁶ The summer 2017 data exchange procedure is [available here](#)

¹⁷ Summer 2017 out of tolerance information is [included here](#).

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

Grade A*

The reporting tolerance at A* is fixed at +/-2 percentage points¹⁸. Because the A* grade is at the very top end of the mark distribution, 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.

¹⁸ 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. There are separate tables for AS, A level and GCSE at the key grades that we review. The figures for the outcomes at each grade are cumulative, so, for example, 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.



© Crown Copyright 2018

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this licence, visit

www.nationalarchives.gov.uk/doc/open-government-licence/

or write to

Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:



Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344
public.enquiries@ofqual.gov.uk
www.gov.uk/ofqual