Marking and grading

Assigning grades to students' work in the UK context is a complicated and large scale process. We've put together a simple guide to the marking process, which can be found below. Overleaf, Beth Black from the OCR Research and Technical Standards team sets out to explain the grading process.

A simple guide: Assuring OCR's Marking Accuracy

Our marking accuracy is founded on four elements:

① our mark schemes,
② our markers,
③ our standardisation and
④ our marking quality assurance processes.

1. Mark schemes are the instructional documents which specify how marks must be awarded. They are carefully reviewed and revised for clarity and accuracy before an examination, and are only finalised once senior assessors have reviewed samples of candidates' work and checked that the range of candidates' answers is covered.
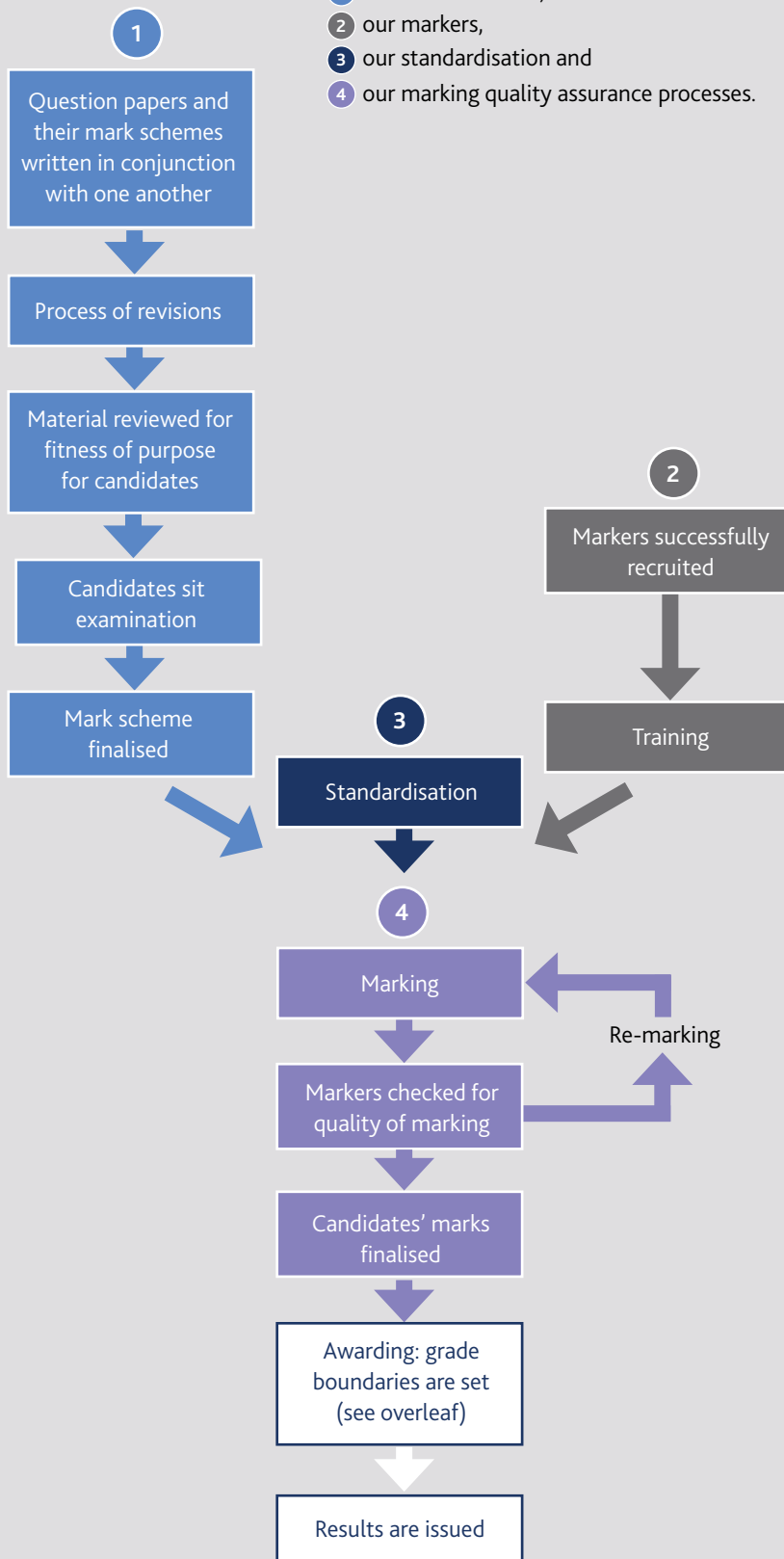
2. Markers are carefully selected and trained, and often have a number of years' teaching experience. Each marker is assigned to a marking team under the supervision of a team leader, who is supervised in turn by the examination's principal examiner.

3. Examiners are **'standardised'** to ensure they all apply the mark scheme fairly and consistently. This is achieved by marking common scripts, either at a meeting or remotely online, which have already been marked by the Principal Examiner. Team leaders then carefully review markers' performance on these common scripts, and provide guidance and feedback to ensure examiners can apply the mark scheme accurately and consistently.

4. Those markers who pass standardisation proceed to **marking** candidates' scripts. Marking accuracy is monitored by:

- Interspersing test scripts (previously marked question papers) amongst each marker's online marking (known as 'seeding')
- Team leaders reviewing a sample of each marker's paper-based marking
- Principal Examiners checking team leaders' marking and their supervision of the markers
- Team leaders checking a sample of the Principal Examiner's marking.

If a marker or team leader is found to be inaccurate, all their scripts are re-marked.

① Question papers and their mark schemes written in conjunction with one another

→ Process of revisions

→ Material reviewed for fitness of purpose for candidates

→ Candidates sit examination

→ Mark scheme finalised

② Markers successfully recruited

→ Training

③ Standardisation

④ Marking

→ Markers checked for quality of marking

Re-marking

→ Candidates' marks finalised

→ Awarding: grade boundaries are set (see overleaf)

→ Results are issued

How do candidates get their grades?

Assigning grades to students' work in the UK context is a complicated and large scale process. Beth Black from the OCR Research and Technical Standards team sets out to explain the process.

Grading versus Marking

Grading (usually referred to as 'awarding') is a separate process from that of marking and takes place after all (or nearly all) the marking is completed. For A Level and GCSE, marks are awarded using a detailed mark scheme and the marks are assigned to scripts independently of any consideration of grade-worthiness. The process of categorising marks into grades takes place later and is conducted by an awarding committee.

The Task of the Awarding Committee

For each specification or syllabus an awarding committee is convened. The awarding committee usually consists of:

(a) a Chair of Examiners who has overall responsibility for maintaining standards across different specifications in a subject area

(b) the Chief Examiner who is responsible to the Chair for the specification as a whole and

(c) Principal Examiners who are responsible for the marking and/or question setting for the individual papers (units) that make up the specification.

The committee's task is to recommend grade boundaries for each of the units within the qualification so that there is comparability of standards both at unit level and qualification level from one year to the next.

This is achieved by combining:

(i) **Expert judgement** on the performance standards exhibited on the candidate work ('scripts').

(ii) **Statistical information** in the form of a range of data.

Expert judgement

Unit by unit, the awarding committee seeks to determine the location on the mark scale where the standard of work is equivalent to a previous year. The difficulty of the two examinations is often different and the boundary marks should be adjusted to reflect this.

To ensure the standard is carried over between years, examiners familiarise themselves with archive scripts which were on the boundary marks for 'key grades' (grades A, C and F for GCSE units; grades A and E for A Level units) in the previous year. By doing this they gain a sense of any difference in the demand of the question paper for the current year and the benchmark year. They also refer to Grade Descriptors which denote the qualities of candidates at each of the key grades. Examiners' judgements are then collated and facilitate discussion (see table 1).

Table 1: Awarding committee examiners' judgements about grade worthiness – E/U boundary

MARK	EXAMINER 1	EXAMINER 2	EXAMINER 3	EXAMINER 4	EXAMINER 5
24	✓✓	✓✓	✓✓	✓✓	✓✓
23	✓✓	✓?	✓?	✓✓	✓✓
22	✓✗	✓?	✓?	✓✓	✓?
21	✓✗	✓✗	✓✗	✓✗	✓?
20	✗✓	✗✓	✗✓	✗✓	✗✓
19	✗✗	✗✗	✗✗	✗✗	✗✗

In this example, the lowest mark of 19 would be excluded from further consideration as there is no judgemental evidence supporting the idea that work on this mark meets the performance standards of a grade E. The decision as to the actual boundary would hinge on further consideration of statistical data (see below) and candidate performance.

Once the 'key grade' boundaries have been determined, the boundaries of the intervening grade are assigned arithmetically.

Statistical Information

A substantial amount of statistical evidence is considered, for example:

(i) **Mark distribution** information such as the mean mark and standard deviation, of the current session as well the previous sessions.

(ii) **Teacher forecast** grade information. This is usually presented as cumulative percentages.

(iii) **Prior attainment** of the candidates who have taken the unit. For A Level, the measure of prior attainment is the profile of mean GCSE scores. For GCSE, Key Stage 2 test results are used. This methodology is commonly referred to as the *Comparable Outcomes Approach*.

Comparable Outcomes Approach

The aim of this approach is to fix into the system the same amount of value added, year-on-year, for a subject at the relevant qualification level. This approach ensures that two successive cohorts with the same profile of prior attainment will receive the same distribution of grades.

A prediction of the grade profile for the specification is produced using the following method:

A matrix is created using national data from a benchmark year for the subject in question. This shows prior attainment and outcome in the particular A Level (example at table 2).

Table 2: Example prediction matrix for an A Level specification

GCSE mean category	% of matched candidate entry in each category	Number of candidates	CUMULATIVE % AT GRADE						
			A*	A	B	C	D	E	U
1	4.0	1099	37.22	83.72	98.54	100.00	100.00	100.00	100.00
2	7.4	2021	14.79	60.36	93.32	99.46	99.95	99.95	100.00
3	7.9	2163	8.69	40.55	85.30	98.89	99.91	100.00	100.00
4	10.8	2947	4.38	24.16	69.49	96.13	99.59	100.00	100.00
5	12.9	3532	2.18	13.62	53.54	90.23	99.29	99.97	100.00
6	11.6	3176	1.07	8.53	43.35	84.57	98.80	100.00	100.00
7	8.8	2414	0.50	4.68	30.28	76.76	97.01	99.84	100.00
8	15.2	4148	0.58	2.96	22.56	66.87	94.86	99.73	100.00
9	12.2	3343	0.24	1.26	11.91	50.91	90.37	99.49	100.00
10	9.1	2496	0.12	0.88	7.49	33.21	78.13	97.95	100.00
Total	100.0	27339	4.33	17.49	45.29	77.22	95.50	99.69	100.00

The left hand column shows each GCSE mean category (numbered one to ten). This represents the deciles of mean GCSE results of 18 year-old A Level candidates who had taken GCSEs two years previously. Category 1 represents the top, most able decile; and category 10 the bottom or least able decile according to prior attainment at GCSE.

Next, we look at the entry for the current year in terms of prior attainment (example at table 3).

For GCSE the process is essentially the same, the only difference being that the prior attainment measure is Key Stage 2 results, divided into 8 categories, for those GCSE candidates who are 16 years old at the point of GCSE certification.[1]

---

1 One exception to this is GCSE Science in the years 2012 and 2013, for which the prediction matrix is based upon 15 year-old candidates. GCSE Science is an exception on the basis that in 2012 and 2013 the majority of entrants were 15 year-olds.

Table 3: Example entry for current year and the predicted grade distribution

GCSE mean category	% of matched candidate entry in each category	Number of candidates	A*	A	B	C	D	E	U
1	4.9	700	260.5	586.0	689.8	700.0	700.0	700.0	700.0
2	7.0	1003	148.3	605.4	936.0	997.6	1002.5	1002.5	1003.0
3	10.8	1546	134.3	626.9	1318.7	1528.8	1544.6	1546.0	1546.0
4	13.3	1898	83.1	458.6	1318.9	1824.5	1890.2	1898.0	1898.0
5	14.3	2044	44.6	278.4	1094.4	1844.3	2029.5	2043.4	2044.0
6	12.4	1765	18.9	150.6	765.1	1492.7	1743.8	1765.0	1765.0
7	9.7	1376	6.9	64.4	416.7	1056.2	1334.9	1373.8	1376.0
8	7.3	1043	6.0	30.9	235.3	697.5	989.4	1040.2	1043.0
9	10.5	1490	3.6	18.8	177.5	758.6	1346.5	1482.4	1490.0
10	9.7	1388	1.7	12.2	104.0	461.0	1084.4	1359.5	1388.0
Total, cumulative	100.0	14243	708.0	2832.1	7056.3	11361.1	13665.8	14210.8	14253
Cumulative % at grade			5.0	19.9	49.5	79.7	95.9	99.7	100.0

The above table shows an example of the predicted grade distribution for the current year through the application of the outcomes in table 2 to the candidates. For example, this method predicts that there are 14243 matched candidates[2] entering this specification and of the 700 candidates entering this particular specification who are in mean GCSE category 1, we would expect approximately 261 (i.e. 37.22% of them – see corresponding part of Table 2) to achieve an A*, and approximately 586 (i.e. 83.2% of them) to achieve an A* or A.

All English exam boards use the Comparable Outcomes Approach, implementing the same prediction matrix within each subject (at GCSE or A Level). Boards are expected by Ofqual to reach predictions as closely as possible although the regulator allows a degree of flexibility – commonly known as 'reporting tolerances'. These take a numeric form (e.g. +/- 2%), and relate to specific grades.[3] Generally speaking, the greater the number of candidates for a specification, the smaller the reporting tolerance. For example, A Level specifications with more than 3000 matched candidates have a reporting tolerance of +/– 1% for grade A. So, for the prediction in Table 2 above, an awarding organisation would be permitted to award between 18.9% and 20.9% for grade A and between 78.7% and 80.7% at grade C.

On occasions there is some difficulty in reaching the predicted outcome. It may be that the consensus of examiner judgement of candidate performance varies significantly from the grade boundaries required to fit the comparable outcomes prediction; or that there is some knowledge of the current year's cohort being of a markedly different ability profile. In such cases, a further detailed review of evidence would take place and the awarding organisation would inform the regulator of an 'out of tolerance' award. In turn, Ofqual might require further information in order to be satisfied that an out of tolerance award would indeed legitimately maintain standards.[4]

Because of the stringency of such reporting tolerances and the conditions surrounding any potential out of tolerance award, there is little or no mechanism by which awarding bodies can compete on standards, despite the common media usage of phrases such as *"competition on standards"* and *"race to the bottom"*.

Summary and Final Comments

Awarding is a complex activity, the purpose of which is to maintain standards year on year. This is achieved by arriving at judgements about grade boundary thresholds by synthesising a range of indicators of 'standards'. The purpose of expert judgement of scripts is *maintaining content and performance standards* (i.e. what students demonstrate they know and can do), while the comparable outcomes approach is designed to *maintain the value added* of successive cohorts.

---

2   Matched candidates are those candidates who, for A Level, are 18 years old and who have a mean GCSE result for 2 years prior to A Level certification.

3   Reporting tolerances apply to grade A only for A Level, and grades A and C for GCSE.

4   For further details, see Ofqual document http://ofqual.gov.uk/files/2013-06-26-130624-OOT-outcome-data-review-process.pdf