

Awarding and comparable outcomes

Ofqual's role

- Maintaining standards over time
- Aligning standards between boards in a subject (and between tiers)
- Our powers:
 - General Condition H3 – requires an AO to monitor the standards of its qualifications, and similar qualifications offered by other AOs, and to use this monitoring to inform its standard-setting
 - General Condition H4 – allows us to direct an AO to adjust its standard in order to come into line with previously specified requirements

Awarding principles – comparable outcomes is not new

- Similar cohort adage (Newton, 2011) - if the cohort hasn't changed much, then we wouldn't expect the proportions at each grade to change much either – a principle used by exam boards for decades
- Cresswell (2003) – prioritise 'comparable outcomes' over 'comparable performance' when new A levels were introduced in 2002, to be fair to those candidates in the first cohort
- Statistics have always been used to support awarding, to counter possible biases in examiner judgment

References

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

Awarding grades in linear qualifications

- Awarding committees for each specification
- Script evidence guided by statistical evidence
- Grade boundaries set at overall subject level
- 'Key' grade boundaries set by awarders:
 - 9, 7, 4 and 1 for GCSE
 - A*, A and E for A level

Ofqual's approach since 2009 – for first awards in...

- 2009 – new (mostly 2-unit) AS awards
- 2010 – new (mostly 4-unit) A level awards with new A* grade
- 2011 – new GCSEs in all subjects except English, maths, sciences and ICT
- 2012 – new GCSEs in English, maths, ICT, Science
- 2013 – new GCSEs in Additional science, biology, chemistry, physics
- 2014 – linear GCSEs (and English/English language with separate S&L)

- Prioritised comparable outcomes to protect students from being unfairly disadvantaged by being the first to sit new/changed qualifications

Predictions

- Statistical approach to predict likely percentage of students achieving each grade
 - Predicted outcome contextualised by the prior attainment profile of cohort (KS2 or GCSE)
 - Based on the relationship between the prior attainment and outcomes in a reference year
- Why...?
- Overcomes limitations of examiner judgement
 - Protects students in the early years of a new specification
 - Protects against 'benefit of the doubt' in the system leading to grade inflation

So does that mean...

- A fixed quota of grades each year?
 - Students are unfairly disadvantaged?
 - My school can't improve?
 - What students have to do to achieve a C (or other grade) will be very different each year?
-
- No

Standard KS2 Prediction

■ Reference Year:

			GCSE Grade								
			*	A	B	C	D	E	F	G	U
Candidate 34567	3.333	C									
Candidate 34568	4.000	C									
Candidate 34569	5.000	A	37%	46%	63%	74%	83%	90%	97%	99%	100%
Candidate 34570	2.667	D	33%	41%	52%	63%	73%	81%	92%	97%	100%
Candidate 34571	1.333	F									
Candidate 34572	4.333	*									
Candidate 34573	3.667	A	26%	32%	41%	52%	66%	71%	85%	91%	100%
Candidate 34574	3.667	*									
Candidate 34575	2.667	C	13%	19%	31%	44%	54%	67%	81%	88%	100%
Candidate 34576	5.000	B	5	10%	14%	24%	40%	52%	62%	75%	85%
Candidate 34577	1.667	G									
Candidate 34578	5.000	A	6	7%	12%	20%	34%	42%	59%	71%	80%
Candidate 34579	5.000	A									
Candidate 34580	2.333	C									
Candidate 34581	3.000	B									
Candidate 34582	4.000	C	8	3%	6%	12%	20%	34%	55%	66%	76%

So how does that actually work?

The exam boards start with the reference year. For every student, they know their mean KS2 score. They also know what GCSE grade they got in the subject.

They divide all those students into eight equally sized categories – or octiles – from 1 highest to 8 lowest. For each student, they map their KS2 category to their achieved grade. So this one is category 3 and got a C. This one is category 2 and got a C. And they keep doing that until they have a matrix that shows, for every category, the cumulative percentage of students that got each GCSE grade.

Standard KS2 Prediction

■ Current Year:

			GCSE Grade										
			*	A	B	C	D	E	F	G	U		
Candidate 34567	3.333	??	<div>KS2 Category</div>	1	37%	45%	63%	74%	83%	90%	97%	99%	100%
Candidate 34568	4.000	??		2	33%	41%	52%	63%	73%	81%	92%	97%	100%
Candidate 34569	5.000	??		3	26%	32%	41%	52%	66%	71%	85%	91%	100%
Candidate 34570	2.667	??		4	13%	19%	31%	44%	54%	67%	81%	88%	100%
Candidate 34571	1.333	??		5	10%	14%	24%	40%	52%	62%	75%	85%	100%
Candidate 34572	4.333	??		6	7%	12%	20%	34%	42%	59%	71%	80%	100%
Candidate 34573	3.667	??		7	5%	9%	17%	25%	35%	60%	68%	79%	100%
Candidate 34574	3.667	??		8	3%	6%	12%	20%	34%	55%	66%	76%	100%
Candidate 34575	2.667	??											
Candidate 34576	5.000	??											
Candidate 34577	1.667	??											
Candidate 34578	5.000	??											
Candidate 34579	5.000	??											
Candidate 34580	2.333	??											
Candidate 34581	3.000	??											
Candidate 34582	4.000	??											

Prediction

They then turn to the current year.

Again, for each student they know the mean KS2 score. But they don't know their GCSE grade for the subject – that's what they are going to try and predict.

Again they divide the students into eight categories – but these are not (necessarily) evenly sized. They use the same cut-offs in terms of mean KS2 score that were used in the reference year.

So, based on their mean KS2 score, this one would be in category 3. This one would be in category 2.

And they keep doing that until they have divided all the students into the eight categories.

And then – based on how the percentage of students in each category in the current year compares to the percentage of students in each category in the reference year – they predict the cumulative percentage of students that will get each GCSE grade for each category.

This forms the national prediction matrix that all the exam boards then use in making the individual prediction for their specification. Those individual predictions vary based on the mix of students taking the specification, in terms of their mean KS2 score.

10

Tiering in reformed GCSE maths

Foundation	Higher
	9
	8
	7
	6
5	5
4	4
3	(3)
2	
1	
U	

Reformed GCSE maths is available at foundation tier (targeted at grades 5-1) and higher tier (targeted at grades 9-4 with an allowed grade 3).

Tiering in reformed GCSE maths

Foundation	Higher
	9
	8
	7
	6
5	5
4	4
3	(3)
2	
1	
U	

The tiers overlap at grades 4 and 5 – students can achieve these grades on both tiers.

Students can also achieve a grade 3 on both tiers. On higher tier this is an allowed grade though so the higher tier isn't actively targeted towards grade 3.

Subject conditions

- GCSE subject level conditions and requirements for mathematics
 - Foundation and higher tier
 - Overlapping tiers model
 - Foundation 5-1 and higher 9-4 (with allowed grade 3)
 - Overlap at grades 4 and 5 (grade 3 not actively targeted on the higher tier)
 - No mixed tier entry
 - At least 20% common marks between tiers, targeted at grades 4 and 5

Requirements for tiering are set out in the subject level conditions for GCSE maths.

Comparability of tiers

- AOs required to document their approach to ensuring comparability between tiers, including:
 - the use of marks from common questions;
 - the proportion of marks for each tier that are targeted at the overlapping grades; and
 - the level of attainment that Learners achieving these grades at the different tiers must demonstrate in relation to the areas of content, content domains and assessment objectives.
- AOs required to report how they have approached ensuring comparability between tiers each exam series.

Exam boards are required to document their approach to ensuring comparability between tiers and report how they have ensured comparability each exam series.

Tier equating - example

- For each total score on the Higher tier, find the percentile associated with that score

Mark (max = 100)	Cum %
50	55.00
49	57.00
48	59.00
47	61.00
46	63.00
45	65.00

Test equating will be used by each exam board to ensure comparability between tiers. The method is called 'chained equipercentile equating' and this uses performance on the common items to generate equated pairs of marks that represent the same standard on each tier.

Here's a fictitious example for a test with 100 marks on each tier and 20 common marks between tiers. There are 4 steps to the equating.

Step 1 is to generate a mark distribution for the higher tier that shows the cumulative percentage of students achieving each mark.

Tier equating - example

- For each total score on the Higher tier, find the percentile associated with that score

Mark (max = 100)	Cum %
50	55.00
49	57.00
48	59.00
47	61.00
46	63.00
45	65.00

From this you can see the cumulative percentage of students at each mark. So if we assume our grade boundary on the higher tier is 48, you can see that the cumulative % of students achieving that mark is 59%.

We then want to know what the equivalent mark is on the foundation tier.

Tier equating - example

- Find the score on the sub-test comprising the common items for each percentile identified above for the higher tier cohort.

Common items mark (max = 20)	Cum %
12	55.00
10	57.00
8	59.00
6	61.00
4	63.00
2	65.00

Step 2 is to generate another mark distribution for just the common items for higher tier students (max = 20 marks).

Tier equating - example

- Find the score on the sub-test comprising the common items for each percentile identified above for the higher tier cohort.

Common items mark (max = 20)	Cum %
12	55.00
10	57.00
8 ←	59.00
6	61.00
4	63.00
2	65.00

From this we can see the score on the common items that would give the same percentage outcomes at our grade boundary of 48. In this examples this is 8 marks.

Tier equating - example


- Find the percentile associated with that score on the sub-test comprising the common items for the foundation tier cohort.

Common items mark (max = 20)	Cum %
12	35.00
10	37.00
8	39.00
6	41.00
4	43.00
2	45.00

Step 3 is to generate a mark distribution for the foundation tier students on the common items.

Tier equating - example

- Find the percentile associated with that score on the sub-test comprising the common items for the foundation tier cohort.

Common items mark (max = 20)	Cum %
12	35.00
10	37.00
8 	39.00
6	41.00
4	43.00
2	45.00

You can then see that at 8 marks (our higher tier mark on the common items) the percentage of students achieving this on the foundation tier is 39%.

Tier equating - example

- Find the total score associated with that percentile on the Foundation tier whole test.

Mark (max = 100)	Cum %
80	35.00
79	37.00
78	39.00
77	41.00
76	43.00
75	45.00

Finally you generate an overall mark distribution for the foundation tier (max = 100).

Tier equating - example

- Find the total score associated with that percentile on the Foundation tier whole test.

Mark (max = 100)	Cum %
80	35.00
79	37.00
78	39.00
77	41.00
76	43.00
75	45.00

You then find the mark that corresponds to 39% - in this case 78. And this is the grade boundary for foundation that equates with the higher tier boundary of 48.