****

**Assessment workshop:**

**The complexities of examiners’ judgements of performance**

**Instructions:**

When a learner is awarded a mark and a corresponding grade, it is easy to see this outcome as a straightforward ‘measurement’ of the learner’s knowledge and understanding. However, for most assessments this is an oversimplification.

Examiners marking assessments must make a judgement of quality based on the work in front of them. These judgements very often vary between examiners. Also, the same examiner may judge a piece of work differently depending on the context. For example, they may subconsciously compare an essay to the last one that they marked, or they may judge a piece of work differently depending on whether they are fatigued or well rested. While the consistency of marking is very high in objectively assessed subjects such as mathematics, it is lower in more subjectively assessed subjects such as English literature.

In GCSEs and A levels, marks are turned into grades. This is also a challenging process as grading standards need to be comparable over time. Awarding committees must judge whether the learners’ work they are seeing at a particular mark is worthy of a grade by comparing it to learners’ work from previous years. This often means making difficult, expert judgements about performance when comparing learners who have taken different papers and answered different questions.

These judgements are also complicated by the fact that the route to any mark varies. For example, a mark of 50 out of 70 could be achieved by a consistently high standard of performance or by a mixed performance in which weak responses to some questions are compensated for by excellent responses to other questions. Two pieces of work might both score 50 out of 70 but look quite different.

In this exercise we will explore some of the challenges that are inherent in examiners’ judgements of performance. You will be provided with authentic, anonymised learner work for each of two GCSE subjects (mathematics and business studies) and we would like you to make judgements about the standard of the work.

Both GCSEs comprise two examined assessments (unit 1 and unit 2). The key grade boundaries are A, C and F. To maintain standards from one year to the next, examiners set these key grade boundaries by aligning the quality of student performance across the two years. They look holistically at the student performance in the context of the difficulty of the examination. They ask themselves - What mark would a student on the boundary last year achieve on this year’s exam? The other grade boundaries (e.g., grade B) are calculated arithmetically (i.e., halfway between the boundaries for A and C).

**Task 1:**

Consider the 3 scripts from unit 1 GCSE business studies. These were awarded a grade C.

* What are the different ways a grade C was achieved?

These 3 scripts were awarded different marks.

* Which script do you think achieved the highest mark?
* Which script do you think achieved the lowest mark?

In setting grade boundaries one mark needs to be chosen to definitively represent the standard, and examiners do not mark the work as part of this process.

* How easy is it for you to decide which is the best work?
* How hard would it be if you were trying to compare work across years (and so with different examination questions)?
* How could we help examiners with this task?

**Task 2:**

GCSE mathematics is a tiered subject. Three different examination papers are set – Foundation tier, Intermediate tier and Higher tier. Students who find mathematics challenging take the Foundation tier exams. Students who are capable mathematicians take the Higher tier exams. The rest take the intermediate exams. This is to ensure that all students can engage with the exam questions they are set, and the best students can demonstrate their knowledge and understanding. Some grades can be achieved on more than one tier, and it is essential that standards are aligned between the tiers so that it isn’t easier to get a grade on one tier than the other.

Consider the 3 higher tier grade C scripts from unit 1 GCSE mathematics and the intermediate tier 3 grade C scripts from the same unit. Each of the 3 scripts within a tier was awarded a different mark.

* Can you rank order the 6 grade C scripts?
* How confident are you in your rank ordering?
* What are the barriers to aligning standards between the two tiers?
* How would you help examiners to decide the marks upon which the standards between the tiers are aligned?

**Task 3:**

The setting of grade boundaries to maintain comparable standards between tiers and over time is very challenging.

* To what extent do you think issues with the consistency of marking make this task harder in GCSE business studies and mathematics?
* To what extent does the format of the questions in the two subjects help or hinder the comparisons?

There is often a desire among some stakeholders that standards across GCSE and A level subjects are aligned. For example, they would like to be assured that the grade C standards in GCSE mathematics and GCSE business studies are equivalent.

* What do you think is meant by ‘aligned’ in terms of inter-subject standards?
* Why might stakeholders desire alignment?
* What are the barriers to achieving alignment?

**Information for tutors:**

The examples of student work have been selected to highlight some of the inherent challenges and ambiguities around examiners’ judgements of performance. The main learning objectives of the workshop are around the challenges of achieving comparable standards whether that be over time, between tiers or between subjects. By seeing for themselves how hard it is to rank order work from the same question paper and to align performance standards across tiers, students will be well positioned to understand the operational challenges around standard setting.

The students are invited to consider what additional evidence examiners may need to consider when setting grade boundaries. This evidence includes statistical information (e.g., about outcomes over time or between tiers) and information about how hard the examinations proved to be for the students (e.g., mean and standard deviation of marks). In the third task they are invited to think about what it means to talk of ‘aligned standards’ across subjects.

In the folder you will find scripts on the grade boundary mark, one mark above and one mark below the grade boundary. There are 3 scripts at each of these mark points, for each boundary.

For **Task 1** give students3 grade C scripts (1 for each mark – 36, 37, 38) from unit 1 GCSE business studies.

For **Task 2** give students3 higher tier grade C scripts from unit 1 GCSE mathematics (1 for each mark – 12, 13, 14) and the intermediate tier 3 grade C scripts from the same unit (1 for each mark – 27, 28, 29).

**Additional Reading:**

We have included some references below, should your students like to learn more.

Baird, J. A. (2000). Are examination standards all in the head? Experiments with examiners' judgements of standards in A level examinations. Research in Education, 64(1), 91-100.

He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. Oxford Review of Education, 44(4), 494-513.

Newton, P. E. (2022). Demythologising A level exam standards. Research Papers in Education, 37(6), 875-906.

Taylor, R., & Opposs, D. (2018). Standard setting in England: A levels. Examination standards: How measures and meanings differ around the world, 100-113.