

Personal comments on the **SQA Technical Report: National Qualifications 2020 Awarding - Methodology Report, August 2020, Publication Code BA8262**. Referred to below as “The document” or “SQA document”

Inconsistent treatment of uncertainty.

The SQA know, and most agree, that teacher predicted grades are subject to uncertainty. They know this and couch their handling of grades in the language of statistics and statistical distributions. The teacher predicted grades are ‘corrected’ because centres might over- or underestimate predicted grades. However, the SQA treat teacher rankings of students differently. The rankings are treated by SQA as if they are fundamentally correct, fixed and not, as I have previously stated, subject to considerable uncertainty. For example:

“This approach ... is fundamental to our principle of treating the centres’ rank order as sacrosanct” (p34) and

“Ensuring that the relative ranking of learners as estimated by centres remained unchanged post-moderation and adjustment was of critical importance to SQA.” (p34)

Not accounting for the uncertainty in rankings means that you cannot be sure that you are moving the correct students between grades.

There is no consideration in the SQA’s document that the rankings will also be subject to considerable uncertainty (as mentioned in my Commons Education Select Committee evidence submission). There seems to be no technical analysis of ranking uncertainty.

The treatment of predicted grades and rankings is inconsistent and, in my view wrong. This does not necessarily have a bearing on the integrity of the system, but it can be of immense importance for individual students, who might be moved down when perhaps they should not be. In particular, this would seem a strong position from which to mount a challenge to a grade in any appeal.

Waterfall effect

The SQA document refers to the ‘waterfall effect’. This is where changing the estimated grade(s) of one (group of) student(s), e.g., from A to B, then means that there are “too many” students now in grade B. Some of those are then moved into grade C. Then, likewise, students have to be moved from C to D, and so on. Figure 6 shows a graphic depiction of this process. As far as I understand it, this operates within centres. As mentioned in my Commons Education Select Committee evidence a student’s final grade might be strongly influenced by other students in their centre and/or the teacher predictions for those. For example, suppose a centre has over-predicted Bs to be As, but less so for lower grades. Then, the process will reassign As to Bs, but this effect will then cascade down through the grades where, perhaps, it is not warranted. What evidence is there that the waterfall effect is not unfair?

The systemic point, again raised in my Select Committee evidence, was that the exams in each of the devolved nations are meant to be national exams. In a national exam a student might have their grade changed, but this would only be loosely influenced by other students as the whole national cohort is used to form grade boundaries. However, in SQA’s new standardisation process, a student might have their grade changed as a result of what students and teachers in their local centre had been doing. By contrast, if you take a driving test in the UK, it is a national test set to national standards. Your result should not depend on what has been happening in your local town.

Mathematical optimisation

The mathematical optimisation technique in 6.15 appears to move groups of students (and perhaps individual students, it is difficult to tell) between grade bands (including their refined bands). They claim that this technique “ensures” that the process is “objective”. However, part of this optimisation is to try and minimise the number of students that can be moved, but still achieve the aim of obtaining the correct grade profile. Minimising the number of students is achieved by using a penalty and one whose influence is manually adjusted by using a ‘weighting factor’, which requires a human user to choose a particular form of this weighting factor. For example, do you make fewer but more extreme adjustments, or make a larger number of minor adjustments? However, choosing the weighting factor and penalty is a subjective choice. Hence, the mathematical optimisation is not, overall, objective and it should not be advertised as such.

SQA realise that this is an issue in that they state, “*It was recognised that robust justification was required ...*” and, in the end, “*it was agreed*”, but it is not clear who or what agreed the final choice of weighting factor. If other weighting factors had been chosen, then the outcome, and the students who were moved between grades, might well have been different. Their process might be reasonable, but it is not objective.

Mathematical optimisation, equality and fairness (6.21)

I am concerned about the statement:

“Use of optimisation allowed SQA to explore the impact on the outcomes of the moderation process of applying slightly different constraints. Assessing the outputs of each set of different constraints (an optimisation run) against a number of measures and our guiding principles allowed us to make a judgement about which constraints generated the outcomes that best supported our principles for awarding” (p39)

It is difficult to fully understand what this means. Is it a statement about protected characteristics or social class or something similar? Is it saying anything about different types of Centre (e.g. independent, state or FE college)? What does “slightly” mean here? We know treatment of constraints is subjective (see above). Is this statement saying, “we tried different fudge factors to get the outcome we want”? It is essential that SQA publish full algorithmic details about what is going on here, preferably with anonymised sample data for us to even begin to understand what is being attempted, and then we can come to some assessment of whether it is suitable and fair.

New Centres

For a long while, it has not been clear how historical performance of centres would be incorporated. For new centres, I don’t think anyone knew what was going to happen. However, for new centres, those that do not have a history: “*estimates from these new centres were accepted unchanged*” (p39). The SQA and the Scottish Government have said today that, when adjustments have happened, 93.1% were adjusted down, which indicates that teacher-predicted grades were over-optimistic. So, presumably it is a distinct advantage for a student to be in a new centre. If your teachers are typical, then there will be a tendency to over-predict and, since you are in a new centre, your grades will not be modified. Similar questions about the treatment of centres with a less than complete history could also be asked.

It might be unavoidable, but the differing treatments of new, recent and established centres again means that we have to question whether these are ‘national exams’ with all students being treated in the same way?

Full Transparency

In due course, we would expect to see full publication of algorithms and sample data sets so that the community can come to a mature understanding of what has happened and, hopefully, feed positive alterations into future models. Reading about the SQA's process is enlightening, but the wordy explanation is sometimes confusing and ambiguous.

On the process as a whole, some in the community predicted some of these problems in consultations put before committees of the various governments of the UK, and, in some cases, suggested viable alternatives. In the aftermath, it would be interesting to see whether or how any of these alternatives were considered and/or how much was directed by government ministers.

The problem at the heart of the statistical standardisation is that it can be simultaneously unfair to individuals, but also maintain the integrity of the system. However, if system integrity damages the life chances of individuals, then it is not much of a system.

Professor Guy Nason
Statistics Section
Department of Mathematics
Imperial College, London
4th August 2020