Mark Scaling in Physics

In any system of assessment, particularly one where all candidates do not take the same modules, measures are often taken to ensure that, as far as is possible, equal assessment standards apply across all modules. Similarly measures should be taken to ensure that standards remain similar over the years so that a mark of 60% in year 2010 in a particular module means much the same as a mark of 60% in 2020. Internal and external checking of exam papers does this to some extent. In addition, module marks are often scaled in some way. As you probably know, the exam boards do this at A-level (by shifting the grade boundaries up or down).

In the Physics department we do not scale marks in year 1 as their overall contribution to final classification is small, and the degree of optionality is limited. All other examination marks are scaled using the procedure described below. Assessed work marks are moderated, but not scaled, since assessed work often tests different skills to examinations and there is therefore no obvious data set to scale with reference to.

There are two stages to the process we use, the first of which deals largely with comparability from year to year (and typically has little effect). The second deals with module to module fairness. In this second stage we take all candidates based in physics taking a particular module, and scale their marks with reference to their average performance in all examined physics modules. Students from other departments taking a physics module will have their marks scaled, but their performance will not be used to determine how the scaling will occur.

The scaling procedure determines what the average for a module should be. We then remap the dataset (students' marks) to obtain close to this average in what we believe to be a fair way. Rank order is of course maintained, whilst very high, or very low, marks will not be changed much (0 and 100 are fixed points).

Our approach has been extensively reviewed by a number of external examiners, who have all expressed their approval.

The Procedure

Examiners of all modules in years 2, 3 and 4 are given a guide average mark based on the performance in the previous year's examinations of the cohort of students registered for their module. Examiners are expected to return an average mark within 6% of this figure.

All examination papers taken by 2nd, 3rd and 4th year students are automatically scaled. The first stage of the procedure for a given year of study is to determine the global average (ignoring marks of less than 10%) of all physics-based students on physics examination papers. This value, A_p, is compared with the target figure A_T (63 for years 2 & 3 and 69 for year 4) and the quantity $A'_p = \frac{A_p + A_T}{2}$ is determined. A'_p becomes the new global target average and the appropriate global scaling factor is determined. This is expected to be (and always has been) close to unity.

A scaling factor for each module examination is defined as follows. We find the overall average on all physics papers, for the cohort taking a particular module, and define the

scaling factor to be the ratio between this and the average actually obtained on the examination for the module (both averages are computed ignoring marks of less than 10%). The scaled mark as a function of the raw mark should be a monotonic and invertible function and not favour any one student. We use piecewise linear functions which depend on the pass mark, P, where $P = 40$ for modules with PX2.. or PX3.. codes and $P = 50$ for modules with PX4.. codes. If the scaling factor for an individual module is S, we define $x = |S - 1|\times$ (raw average). For downscaling, we usually map raw to scaled marks as follows: 0 to 0, P to P, $60 + x$ to 60, $80 + x$ to 80 and 100 to 100 with linear interpolations between these points. For upscaling the mapping is: 0 to 0, $P - x$ to P, 70 to $70 + x$ and 100 to 100 with linear interpolations between these points. Under this scheme most students have their mark shifted by a constant amount, namely x. The change in mark then reduces to zero for students with raw marks close to 100. It also goes to zero for students with raw marks equal to P or less, when scaling down, and for raw marks equal to zero when scaling up.

Discussion

You may ask whether our approach is fair. We believe that it is fairer than using raw examination marks. Firstly, setting an appropriate target average mark for an exam means that the examination paper has to be designed to achieve this average, and hence should be a fair assessment of the module. The first stage of scaling, where we compare the global average of the marks on physics papers gained by physics students in a given year of study, results in a small change to marks (typically a few tenths of a percent). The second stage of scaling has more of an effect and it deals with the following issue. Suppose that there is an optional year 3 module which is taken largely by students who generally obtain first class marks. The exam paper is sat and marked and there is an average mark of say 68%, which is inside the target range. However it is a lower mark than the cohort of students taking it generally obtains, either because the exam paper was rather tough, or perhaps because the marking was tough. The effect of our procedure is that the marks on this paper will be shifted upwards so that the average matches that which this group of students obtains on all their papers, which might have been 73%.

February 2024