

EXAMINATION AVERAGES AND THE BALANCE OF PAPERS

Leslie H. Palmier, *Reader in Sociology, University of Bath*

.....

In essay-type examinations for degrees in British universities, marking customs tend to compress the candidates into a narrow band around the mean. This works especially adversely against those who, in examinations with several papers, achieve excellence in one or more. Often such candidates are awarded only the minimum distinction mark, the effect of which is nullified by aggregation. This is of some consequence in the British system, since the designation of a candidate's examination performance as "First Class," "Second Class, Upper Division," etc. is formally made by locating his examination average mark within specified intervals on a percentage scale. Awareness of the possible injustice which this method may cause has given rise to the practice of estimating the "balance of papers," or considering the "classes" attained by the individual papers in the examination (irrespective of the actual mark), and giving the candidate the benefit of the doubt should he, for example, have a certain number of papers in a class higher than that indicated by his examination average. A classification reached in this way may of course be at variance with that indicated by the candidate's average of marks, and thus introduce an arbitrary element. An alternative method is therefore proposed in this paper. The usual custom is to assume that the marks indicating the various classes should be the same for the degree as a whole and for the individual papers. It would seem preferable first to determine the balance of papers which would be accepted as a claim to a First Class degree (since it is unlikely that a candidate writing several papers will be able to attain a first class mark in all) and deduce therefrom the classification intervals for the individual papers. A method of doing so is elaborated.

.....

Probably the most important part of the British university examination procedure is the classification of the candidate. At least for the first few years after graduation the class of his degree has a considerable bearing on the range of opportunities open to him. The British Civil Service, for instance, usually insists on at least a "Second Class Honours, Upper Division" (usually abbreviated as "Upper Second") for its senior section, called the Administration Group, as well as for the Diplomatic Service. Grants for graduate study from official sources, such as the Social Science Research Council, are usually dependent on at least

the same level of attainment. In the teaching profession, the class of degree has a bearing on the salary paid. With the greatly increased number of graduates now on the market, there are signs that even business employers are tending to assume that the degree classifications are relevant for their purposes.

In most if not indeed in all cases the classification is reached by considering the candidate's performance in a number of courses measured by various instruments (unseen examinations with or without course work taken into account, exceptionally course work alone, etc.). It is common, of course, to give whatever weighting is considered necessary to the results on one or other course in reaching the aggregate total.

However, perhaps insufficient attention has been directed to the effect of multiplicity of papers on the process of aggregation. Too often there is the assumption, explicit or implicit, that the class limits prescribed for the aggregate totals also apply to the marks awarded in individual examinations. For example, if the university regulations prescribe that a candidate must reach a certain average to be awarded a First Class, then a performance on an individual course judged to be of that standard is awarded the same mark.

Where the examination is in "essay-type" subjects, this assumption may well prevent the candidate's quality being adequately reflected in the results. It is perhaps a general experience that the marks given in such examinations are concentrated in a narrow band, the boundaries of which are set by the pass mark below, and the minimum "first class" mark above. If we assume, as is often the case, that 40% is the pass mark, and 70% the minimum given for a "first," then virtually all marks will be found comprised by these two percentages. Indeed, for all practical purposes the limit of the percentage scale is not 100, but 80. There may be some defence for the infrequency of marks below 40%, in that weak candidates should perhaps have been excluded from courses long before they reached final examinations. However, there would seem no similarly persuasive reason for the reluctance of examiners to reward first class papers with more than the minimum mark for that category.

This situation reflects, of course, the fact that in essay-type examinations the mark represents a judgment; the assessment is not reached, as it is in certain numerical-type examinations, after allotting marks for correct answers to problems. And the judgments are necessarily related to some idealized conception of what a first class, or upper second, etc., answer should be. Accordingly, first class answers are so rare that the symbol of 70 is quite as expressive as 80 or 90; one is not likely to have, as one does lower down the scale, a number of answers differing only slightly in quality, so that one is compelled to consider which answer deserves, say, 64, and which 66.¹

The more papers a candidate has to answer, the less likely that he will attain distinction in any of them; the mere dissipation of his intellectual resources will see to that. When he does achieve such a distinction, his reward of a minimum

first class mark will probably be aggregated with lower marks in a fairly simple fashion, and any distinction it may imply is lost in averaging.

Awareness of this situation has led many university committees of examiners to weigh the "balance of papers" as a criterion for the class of degree to be awarded; often one or other part of the examination is given greater importance in estimating this balance.²

There are, of course, methods of assessment which obviate the difficulties above, principally by ranking and scaling. In the former case, candidates are not given marks, but are simply ranked in relation to one another. Their ranks for one or several examinations are then converted into "scores" and these are combined and averaged to give a final order of merit. However, this method has evidently not commended itself to university examiners. Understandably, they wish their assessments of candidates' papers to be incorporated in their original form into the final average. Scaling meets with different, but equally strong, objections. An examiner may well explain his higher average mark by arguing that the papers he marked were better written, and it would be an injustice to scale all marks down to a common mean.

The upshot is that, as already described, the unscaled examination marks are often simply averaged. This frequently gives rise to a conflict of evidence between the aggregate average and the balance of papers, which bears particularly hard on the more distinguished candidates. It would clearly be desirable to evolve a system of marking and aggregation which took into account the peculiarities of essay-type marking, as described above, and the averaging process, so as to produce a result similar to that obtained by weighing the balance of papers, while incorporating the original assessment in the form in which it was made.

In search for such a system, the marks awarded to a group of candidates for a Finals Honours Examination of twelve papers were taken for study. They are

¹ The inherent difference between the marking of "numerical" and "essay-type" examinations may help to explain the striking anomalies between the rates of conversion, as one might call them, into degree classes of the results of the General Certificate of Education "A" Level examinations taken on the termination of secondary schooling and a necessary qualification for university entry. The passing grades fall from A to E; of students who had passed in three subjects with an average performance better than a C in three subjects, in 1966 only 5% of those taking Social Studies obtained First Class degrees, against 17% of those who took a technological subject. UCCA, Statistical Supplement to the Fourth Report 1965-6, London, 1967, Table L.2.b.

² Cf Oxford University Examination Decrees 1972, p. 256, which states that in the examination for Philosophy, Politics, and Economics, "The highest honours can be obtained by excellence in *either* Philosophy *or* Politics *or* Economics provided that adequate knowledge is shown in the other subjects of examination."

reproduced in Table 1,³ which also shows the averages of each examination, as well as the mean for each candidate, and also his classification with the total number and percentage in each classification. Here, and throughout this paper, the minimum average percentages qualifying for the various degree classes are assumed to be as follows:

- I : 70+
- II_i : 60–69
- II_{ii} : 50–59
- III : 40–49

The distribution of means appears to follow the normal probability curve, namely, most fall into the “Lower Second” category, with smaller numbers above and below. It may also be noticed that the mean of all marks awarded is 55.9, which is close enough to mid-point of the lower second class category as not to demand any scaling procedures.

It will be seen from Table II that both the first two candidates, out of a total of twelve papers submitted, were awarded four “First Class” and five “Upper Second Class” assessments. We shall assume that on a balance of papers consideration, these two candidates would have been awarded First Class honours. If the averages they obtain are to reflect the balance of papers, it is clear that it will be necessary to raise the marks awarded to the individual papers by a sufficient amount to ensure that their average rises above the minimum for the First Class.

The obvious and easiest method is, of course, to raise the individual marks by the same proportion by which the average falls below the First Class minimum. Both candidates obtained an average of about 64%; they could be raised sufficiently if their individual paper marks were multiplied by 70/64. However, in justice the same measure would have to be applied to all candidates and would be equivalent to applying a marking scale for the individual papers of the kind below:

Class		Minimum Mark
I	$70 \times (70/64)$	76
II _i	$60 \times (70/64)$	66
II _{ii}	$50 \times (70/64)$	55
III	$40 \times (70/64)$	44

In other words, a paper judged of first class standard would be given a mark not less than 76, one of upper second quality not less than 66, and so forth; always assuming the degree class marks remain unchanged.

³Thanks are due to Mr. Brian Hart of the Computer Unit at Bath University for assistance with the calculations summarized in the tables of figures in this article.

Table I. Original Marks, Candidates in Order of Merit

Candidate Number	Papers													Average	Class	Total	%
	1	2	3	4	5	6	7	8	9	10	11	12					
1	61	57	65	57	61	59	60	69	73	70	70	70	64.3	II _i			
2	59	63	56	49	61	78	70	65	60	62	70	70	63.6	II _i			
3	57	61	56	62	69	55	67	70	67	64	62	62	62.7	II _i			
4	56	66	47	54	60	64	56	67	62	63	73	68	61.3	II _i			
5	62	56	55	59	63	74	58	63	58	59	60	60	60.6	II _i	5	21	
6	58	48	61	68	51	48	60	53	58	62	75	75	59.8	II _{ii}			
7	59	57	50	51	53	62	58	64	62	57	70	70	59.4	II _{ii}			
8	58	55	45	60	67	55	57	67	60	58	64	62	59.0	II _{ii}			
9	63	61	60	57	60	67	70	61	57	57	47	47	58.9	II _{ii}			
10	51	63	53	48	66	64	58	49	57	61	66	59	57.9	II _{ii}			
11	57	63	60	53	48	56	63	52	56	54	60	60	56.8	II _{ii}			
12	56	60	46	46	45	52	56	61	73	54	60	52	55.1	II _{ii}			
13	53	51	50	53	58	56	60	56	54	55	57	58	55.1	II _{ii}			
14	51	51	47	57	35	66	63	55	60	61	54	59	54.9	II _{ii}			
15	47	57	55	53	69	51	59	60	39	56	56	56	54.8	II _{ii}			
16	53	56	58	49	62	57	58	58	59	57	45	45	54.8	II _{ii}			
17	56	59	57	49	55	51	62	49	45	61	52	52	54.0	II _{ii}			
18	50	49	53	50	60	47	60	45	54	53	44	38	50.3	II _{ii}	13	56	
19	47	46	48	39	46	48	48	57	49	53	58	58	49.8	III _i			
20	47	46	44	35	54	47	35	61	54	57	50	60	49.2	III _i			
21	52	42	52	46	37	54	57	44	47	48	52	49	48.3	III _i			
22	50	46	52	46	45	43	47	54	51	58	42	43	48.1	III _i			
23	62	43	40	59	48	44	57	22	55	52	45	45	47.7	III _i	5	21	
Mean	55.0	54.6	52.6	52.2	55.3	56.4	58.2	56.6	57.0	57.9	57.9	57.3	55.9				

As will be seen below, the effect of this proportionate increase is simply to shift all candidates up the scale:

Class	Proportionately Raised Marks		Original Marks	
	No.	%	%	No.
I	2	9	0	0
II _i	14	61	21	5
II _{ii}	7	30	56	13
III ⁱⁱ	0	0	21	5

It is doubtful if this distribution would agree with balance of papers arguments, and it is certain it does not reflect the intentions of the examiners. What is required is not a measure which promotes all indiscriminately, but rather one which takes account of particularly meritorious achievement, which otherwise is submerged by the number of papers.

Let us assume, as before, that the examiners believe that a balance of papers such as that of the first two candidates should have been rewarded by a First Class, and that accordingly their more distinguished performances should have earned a premium to ensure that their average was not less than the minimum for that class. We may assume that "more distinguished" excludes Third Class marks, i.e., between 40 and 49. All higher marks are to be increased sufficiently to raise the averages of these two candidates to the First Class minimum. For each mark, the necessary increase may be determined as follows.

It will be observed that the weaker of the two candidates' average is 63.6; the portion of that mark to be rewarded is, of course, $(63.6 - 49)$ 14.6. The difference between his average and the First Class Minimum is $(70 - 63.6)$ 6.4; the relationship between this difference and the portion of his mark to be rewarded is $6.4/14.6$, or 0.438. It follows that if one were to add to each of his marks an amount bearing the same proportion (0.438) to that part of it that merits reward, the average would be raised sufficiently to place him in the First Class.

The self-evident nature of this proposition may be seen when expressed in the following formula. Let RM stand for the rewarded mark, CM for the classification minimum, and M for the mark as awarded, with the portion to be rewarded being, of course, $(M - 49)$. Then

$$\begin{aligned} \text{RM} &= M + (M-49) \frac{(CM - M)}{(M - 49)} \\ &= \text{CM} \end{aligned}$$

As we have seen, the calculation $\frac{CM - M}{M - 49}$ for the outstanding candidates yields the figure 0.438. To ensure justice to all candidates, all marks are adjusted by the

Table II

Candidate Number	Number of Marks in Groups			
	I	II _i	II _{ii}	III
1	4	5	3	0
2	4	5	2	1
3	1	8	3	0
4	1	7	3	1
5	1	5	6	0
6	2	4	4	2
7	1	6	3	2
8	0	6	5	1
9	2	3	6	1
10	0	5	5	2
11	0	5	6	1
12	1	3	5	3
13	0	1	11	0
14	0	4	6	1
15	0	2	8	1
16	0	1	8	3
17	0	2	7	3
18	0	2	5	4
19	0	0	4	7
20	0	2	4	4
21	0	0	5	7
22	0	0	5	6
23	0	1	4	6

formula $RM = M + 0.438 (M - 49)$. The detailed calculations are shown in Table III.

The distribution of classes may be compared with that of the original marks, as follows:

Class	Original Marks Candidates		Converted Marks Candidates	
	No.	%	%	No.
I	0	0	8	2
II _i	5	21	39	9
II _{ii}	13	56	39	9
III	5	21	13	3

It will be observed that converting the marks in this manner has had the effect of distributing the candidates in less bunched manner across the classifications, with some movement upwards.

The above results were obtained by manipulating marks already awarded. Many examiners may, however, have objections to such a procedure, and would prefer not to tamper with the individual marks. Their wishes may be met by setting the Paper Classification Minima, or minimum marks indicating the thresholds for the various classes of answers in the individual papers, at such levels compared with the degree Classification Minima, or minimum means required for the various classes in the degree, that the same objective is achieved.

This may be done by first deciding the balance of papers that would qualify for a first class degree (e.g., so many firsts, so many upper seconds, and so forth), then assigning numerical values to the individual paper classifications (medians of the classes may be used) and obtaining the average of such a distribution. The figure $\frac{(CM - M)}{(M - X)}$ may then be calculated, where CM is the classification minimum for the degree (e.g., 70 for a first), M is the average of the papers, and X is the limit above which marks are to be rewarded (49 in the example above). If, for ease of exposition, we assume that, as in the example above, the average is 63.6, then the calculation of course yields the figure 0.438. The Paper Classification Minima may be calculated from the formula:

$$PCM = CM + (0.438) (CM - X)$$

Applied to the Classification Minima we have assumed in this paper, the formula yields the following:

Class	Paper Classification Minima	Degree Classification Minima
I	79	70
II _i	65	60
II _{ii}	50	50
III	40	40

This means, of course, that an answer paper judged First Class would be given a mark of at least 79, and not 70. Similarly, Upper Second papers would be indicated by marks not from 60 to 69, but from 65 to 78, while a Lower Second answer would be marked between 50 and 64, not between 50 and 60. A Third Class or Pass paper, however, would still be given only between 40 and 49. In other words, under the conditions specified, the more meritorious the paper is judged, the higher up the scale the paper mark is pushed relative to the degree classification brackets.

Table III. Converted Marks, Candidates in Order of Merit

Candidate Number	Papers												Average	Class	Total	%	
	1	2	3	4	5	6	7	8	9	10	11	12					
1	66.3	60.5	72.0	60.5	66.3	63.4	64.8	77.8	83.5	79.2	79.2	79.2	79.2	71.0	I	2	8
2	63.4	69.1	59.1	49.0	66.3	90.7	79.2	72.0	64.8	67.7	79.2	79.2	79.2	70.0	I		
3	60.5	66.3	59.1	67.7	77.8	57.6	74.9	79.2	74.9	70.6	67.7	67.7	68.7	68.7	II _i		
4	59.1	73.4	47.0	56.2	64.8	70.6	59.1	74.9	67.7	69.1	83.5	76.3	66.8	66.8	II _i		
5	67.7	59.1	57.6	63.4	69.1	84.9	61.9	69.1	61.9	63.4	64.8	64.8	65.7	65.7	II _i		
6	61.9	48.0	66.3	76.3	51.9	48.0	64.8	54.8	61.9	67.7	86.4	86.4	64.5	64.5	II _i		
9	63.4	60.5	50.4	51.9	54.8	67.7	61.9	70.6	67.7	60.5	79.2	79.2	64.0	64.0	II _i		
7	61.9	57.6	45.0	64.8	74.9	57.6	60.5	74.9	64.8	61.9	70.6	67.7	63.5	63.5	II _i		
8	69.1	66.3	64.8	60.5	64.8	74.9	79.2	66.3	60.5	60.5	47.0	47.0	63.4	63.4	II _i		
10	51.9	69.1	54.8	48.0	73.4	70.6	61.9	49.0	60.5	66.3	73.4	63.4	61.9	61.9	II _i	9	39
11	60.5	69.1	64.8	54.8	48.0	59.1	69.1	53.3	59.1	56.2	64.8	64.8	60.3	60.3	II _i		
12	59.1	64.8	46.0	46.0	45.0	53.3	59.1	66.3	83.5	56.2	64.8	53.3	58.1	58.1	II _{ii}		
14	51.9	51.9	47.0	60.5	35.0	73.4	69.1	57.6	64.8	66.3	56.2	63.4	58.1	58.1	II _{ii}		
15	47.0	60.5	57.6	54.8	77.8	51.9	63.4	64.8	39.0	59.1	59.1	59.1	57.8	57.8	II _{ii}		
13	54.8	51.9	50.4	54.8	61.9	59.1	64.8	59.1	56.2	57.6	60.5	61.9	57.7	57.7	II _{ii}		
16	54.8	59.1	61.9	49.0	67.7	60.5	61.9	61.9	63.4	60.5	45.0	45.0	57.6	57.6	II _{ii}		
17	59.1	63.4	60.5	49.0	57.6	51.9	67.7	49.0	45.0	66.3	53.3	53.3	56.3	56.3	II _{ii}		
18	50.4	49.0	54.8	50.4	64.8	47.0	64.8	45.0	56.2	54.8	44.0	38.0	51.6	51.6	II _{ii}		
19	47.0	46.0	48.0	39.0	46.0	48.0	48.0	60.5	49.0	54.8	61.9	61.9	50.8	50.8	II _{ii}	9	39
20	47.0	46.0	44.0	35.0	56.2	47.0	35.0	66.3	56.2	60.5	50.4	64.8	50.7	50.7	II _{ii}		
22	53.3	42.0	53.3	46.0	37.0	56.2	60.5	44.0	47.0	48.0	53.3	49.0	49.1	49.1	III		
23	67.7	43.0	40.0	63.4	48.0	44.0	60.5	22.0	57.6	53.3	45.0	45.0	49.1	49.1	III	3	13
21	50.4	46.0	53.3	46.0	45.0	43.0	47.0	56.2	51.9	61.9	42.0	43.0	48.8	48.8	III		

Mean: 59.4

To elaborate further, if the examiners rule that performances below Upper Second level are not to be rewarded, X then is 59, and the reward fraction

$\frac{CM - M}{M - 59}$. Worked out on the basis of the average scored by the weaker of the

two outstanding candidates, the calculation becomes:

$$\frac{70 - 63.6}{63.6 - 59} = \frac{6.4}{4.6} = 1.39$$

Applying the formula

$$RM = M + 1.39 (M - 59)$$

to all the marks in the examination under discussion, one obtains the following distribution of classifications:

Class	Candidates	
	No.	%
I	2	8
II _i	8	34
II _{ii}	8	34
III	5	21

It will be observed that, as is implicit in the formula, the movement upwards is less than in the preceding calculations while still achieving a more even distribution of candidates in the Second Class and First categories.

The Paper Classification Minima for use in the individual examination papers at the time of marking would appear as follows, based on the formula

$$CM = CM + 1.39 (CM - 59):$$

Class	Minimum Mark to be Awarded
I	85
II _i	61
II _{ii}	50
III	40

The various distributions of classifications of the candidates under study, using the methods described above, are repeated below for comparative purposes.

Class	Original %	Proportionately Raised Marks %	Marks Above 49 Rewarded %	Marks Above 59 Rewarded %
I	0	9	8	8
II _i	21	61	39	34
II _{ii}	56	30	39	34
III	21	0	13	21

Perhaps the principal point underlying the above discussion has been that examination results follow from the decisions of the examiners concerning the number of papers to be averaged and the balance of performances considered worthy of allocating to one or other class. Most examining boards, knowing well the limitations of examination marking, are unwilling to accept the mean of a candidate's marks as the sole criterion of the class into which he is to be placed. The suggestions above may be helpful in encouraging examiners to use methods of marking which do not create conflicts between the candidate's mean and that of the balance of his papers.