FHS Physics 2007

Chairman's Report

Summary In addition to the usual detail on procedures and marks, this report contains (i) the description of a new algorithm for normalizing optional papers, and (ii) comparisons between the distributions of classes in Physics and other schools that leads to the recommendation that we fundamentally change the way we associate marks with class boundaries.

New papers This year saw the introduction of two new papers: *B4, Mathematical Physics* and *C7, Biophysics*. It was also the first year in which practical marks contributed to the classification.

Although the majority of candidates sitting B4 were physicists, the setting of the paper was in the hands of the Mathematics Examiners. Although the overall student feedback on this course is positive, the June paper attracted strong criticism, and it must be said that the distribution of marks is not satisfactory: two questions attracted only one attempt each and the least popular four questions attracted a total of 13/112 attempts. It is not clear that scrutiny of Mathematics papers is as careful as that given to Physics papers, and in future years it would be good if a member of the Physics faculty were given an opportunity to comment on draft questions for the B4 paper.

The introduction of B4 has a significant effect on the structure of the third year of Physics, since now not all MPhys students sit all B papers; most students sit 3/4 B papers, while a handful sit all four B papers. This situation makes it necessary to reconsider the scaling of the B papers, and next year changes will have to be made to the scaling of the C papers, which this year, as previously, were scaled by reference to the B papers, which everyone had sat.

The introduction of C7, which is set and marked within Physics, raises no issues.

The introduction of practical marks to the classification process proved a major headache even though this year it only affected the BA list; next year it will also affect the longer MPhys list. The use of these marks adds considerable complexity because (a) they are made up of many components, and (b) for legitimate reasons some candidates have no practical marks, while others have incomplete sets. More consideration should be given to keeping things simple when policy changes are considered.

Paper Preparation The procedures introduced last year for setting and checking papers were again employed. Despite strenuous efforts by the Examiners the following mistakes slipped through

a. A1: Minus sign missing in $F = -kT \ln Z$;

b. B1: In Q1 it was intended to ask candidates about Ca ($Z = 20$) but the question specified Mg ($Z = 12$) while giving that the ground state is $4s^2$. The mistake was announced 35 minutes into the exam and great care was taken in the marking to mark the relevant section of the question generously and to provide compensation in cases where a candidate had lost time by virtue of the conflicting information.

c. B2: A cross section had dimensions cm^{-3} rather than cm^2. This error was announced in the exam.

d. C3: Q4 had $/2b$ instead of $/4b$. This error came to light only after the exam.

e. C6: In the last part of Q3, the radius of an orbit was given as $\frac{9}{5}r_s$ rather than $\frac{9}{2}r_s$, which made it impossible to get the specified frequency range. This error only came to light during marking. Only a minority of students reached this part of the question and they were awarded full marks if they showed that they understood how to calculate the frequency range. There was no evidence that any student lost significant time trying to get the required answer.

Table 1 shows the means and standard deviations of individual papers. We are still having difficulty setting papers that achieve scores as high as that desired (65 for compulsory papers) but we did better than last year: the raw means of A1 and B1 were slightly above 65, those of C1, C3 and C7 were extremely close to 65 and only B3, C4, C5 and C6 have a raw means significantly below 60. The target means for the B and C papers differs from 65 for reasons explained below under the heading "scaling". The problem of underscoring is exacerbated by the fact that the target means for B4, C4 and C6 are above 70.

Scaling As last year, the A papers were quadratically scaled to a mean of 65 (0 and 100 invariant), while the C papers were quadratically scaled to means calculated by looking at the performance of each paper's candidates in the B papers. At the marks meeting it was resolved to change this algorithm so that

Table 1: Statistics of individual papers. All papers out of 100 except S papers, which are out of 50

	A1	A2	A3	B1	B2	B3	B4	C1	C2	C3	C4	C5	C6	C7
Cands	171	157	171	132	137	151	28	34	25	44	43	27	39	30
Target	65	65	65	64.54	65.90	63.16	72.69	64.27	63.89	64.30	70.25	63.99	72.78	64.70
Raw mean	66.19	59.77	59.72	67.01	60.36	52.87	64.21	64.35	59.92	64.98	55.67	53.57	57.09	63.13
Raw SD	13.14	14.78	13.59	12.30	13.36	14.29	12.72	15.45	16.45	14.01	15.07	12.94	18.56	10.33

	S1	S2	S4	S6	S7	S8	S10	S12	S15	S16	S17	S18	S20	S21	S22
Cands	59	54	88	14	21	5	18	28	30	12	4	4	3	8	15
Mean	28.42	27.76	34.61	34.36	37.38	41.80	27.56	28.00	27.50	37.33	40.25	31.75	30.67	31.25	33.60
SD	11.64	6.76	5.97	2.27	6.42	4.44	10.61	8.44	7.46	8.42	8.50	12.55	5.86	2.71	5.07

Note: S1 is Functions of a Complex Variable, S2 is Astrophysics, S4 is Energy Studies, S6 is History of Electromagnetism, S7 is Classical Mechanics, S8 is Covariant Electromagnetism, S10 is Medical & Environmental Physics, S12 is Biological Physics, S15 is Chaos, Chance & Predictability, S16 is Plasma Physics, S17 is Numerical Methods, S18 is Advanced Quantum Mechanics, S20 is the History of Science, S21 is & is the Philosophy of Science, S22 is Teaching and Learning Physics in Schools

the mean of the scaled C-paper marks was equal to the mean of the same candidates' scaled marks across the B papers $\overline{X}_{\rm B}$; last year the C-paper mean was set to $\overline{X}_{\rm B}$ times 65 divided by the mean of the scaled B-paper marks of *all* MPhys candidates. The new scaling slightly raises the C-paper marks because on average the MPhys candidates scored 66.58 in the B-papers.

Now that very few candidates offer all the B-papers it is inappropriate to set the mean scaled mark on every B-paper to 65, for the same reasons that the means of scaled C-paper marks have always been individually set. The algorithm just described for the C-papers could have been applied to the B-papers by reference to the A-papers, but then next year the C-papers would have to be scaled with respect to the A-papers, sat two years earlier. This procedure is intrinsically undesirable, and also error-prone because candidates who fall behind for a year or enter later than Part A make it difficult to get a complete set of marks for a given cohort an earlier stage.

For these reasons a *self-scaling algorithm* was developed that calculates the means of a series of papers by examining the differences $X_i^{(\alpha)} - X_i^{(\beta)}$ between the scaled marks that candidate i scored in papers α and β: specifically, the means $\overline{X}^{(\alpha)}$ of the scaled marks on paper $\alpha = 1, 2, \ldots$ are those that minimize the squared discrepancy $S = \sum_{\alpha<\beta}\sum_i (X_i^{(\alpha)} - X_i^{(\beta)})^2$ subject to the constraint that the mean of all scaled marks is some given number, in this case 65. The idea behind this algorithm that a strong candidate tends to perform well in any paper s/he sits, so any systematic tendency for given candidates to score less highly in paper α than in paper β will reflect paper α being tougher than paper β. The mathematical details of self-scaling are given in an appendix.

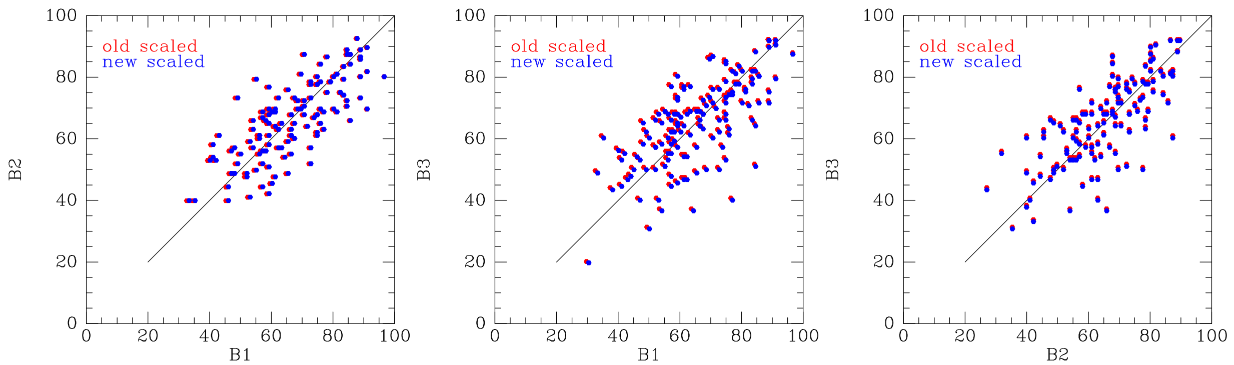

Fig. 1 Scaled marks on the 2006 B papers as used last year (red) and as calculated by self-scaling (blue).

The algorithm was first tested on last year's data. When used to scale the B papers the scaled means for B1–B3 became 65.68, 65.05, 64.28; these means differ from 65 only because BA candidates took only 2/3 B papers, and the stronger candidates were more likely to drop B3 while the weaker candidates were

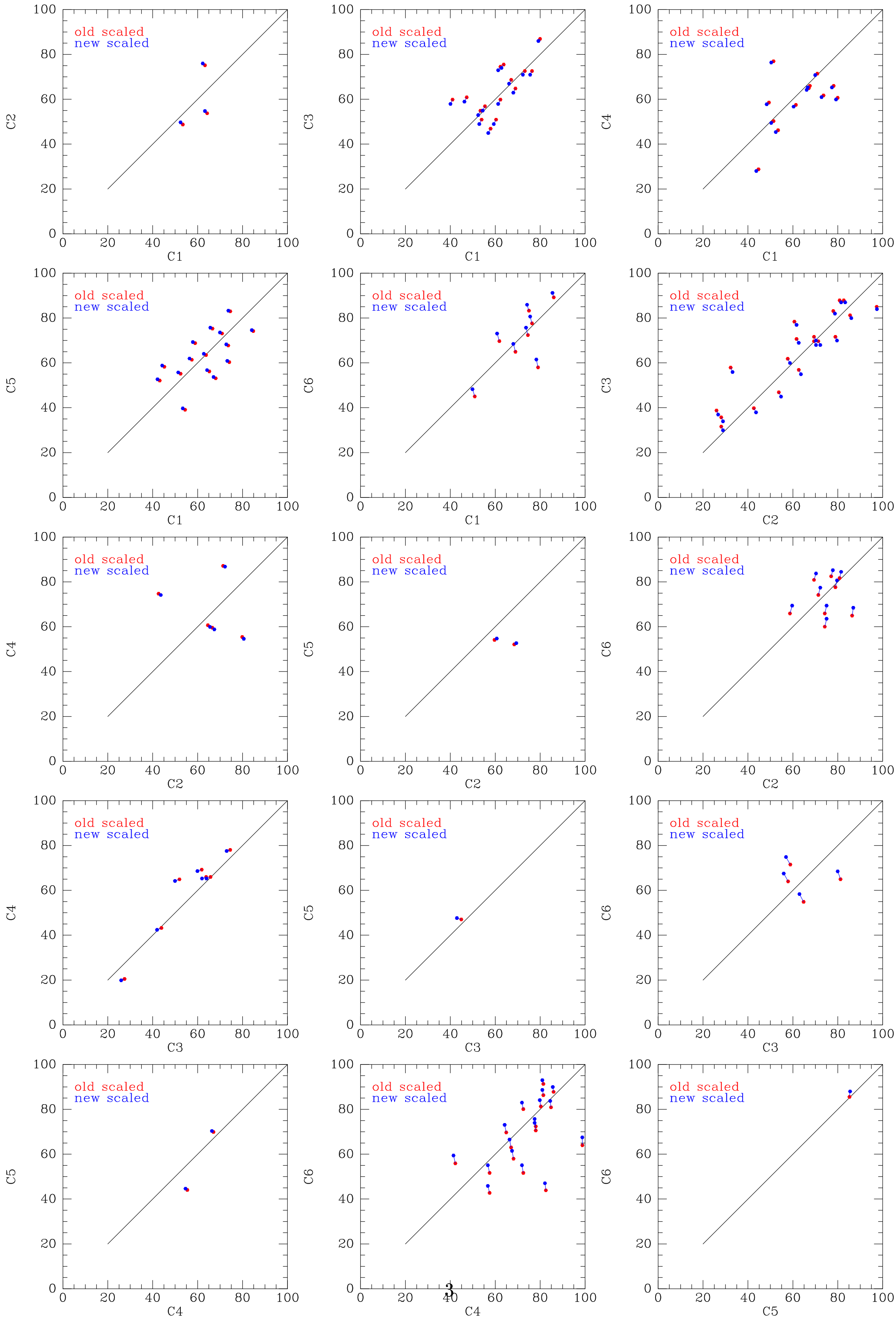

Fig. 2 Scaled marks on the 2006 C papers as used last year (red) and as calculated by self-scaling (blue).

Table 2: Scaled means of 2006 C-papers

	C1	C2	C3	C4	C5	C6
self-scaling mean	62.62	65.79	60.91	65.29	62.12	73.27
mean from B-papers	63.56	64.97	62.60	65.93	61.61	70.19

more likely to drop B1. Fig. 1 shows the correlations between marks on different papers. It also shows that individual marks change negligibly when self-scaling is used rather than setting all means to 65.

Fig. 2 shows the much more stringent test that is provided by last year's C papers. Table 2 shows the scaled means of the C papers from self-scaling and from consideration of the B-papers. The biggest effect of self-scaling is to increase the mean of C6 by 3.08 marks, and the next biggest effect is to lower the mean of C3 by 1.7 marks. The two algorithms agree about which options attract the strongest candidates. The fact that self-scaling assigns a higher mean to C6 than that obtained from the B-papers should indicate that the strong candidates who choose this option have in the fourth year improved their understanding in other areas of physics more rapidly than the rest of the cohort.

The Examinations The Examinations went off smoothly. We did better than last year in two tiresome areas: splitting scripts into sections that are marked in different sub-departments, and gathering in scripts from candidates who do not sit in the main examination hall. However, I repeat the recommendation made in last year's report, that Schools should cease to assign different serial numbers to a given paper depending on the School (Physics, Physics & Philosophy, Mathematics, etc.) within which it is offered. This multiplicity of serial numbers unnecessarily complicates script tracking. I also report that although the invigilators were uniformly helpful, on one occasion a permanent member of staff in Schools made the Examiners' work unnecessarily difficult, and it was appropriate to complain about this to the Clerk of Schools.

Marking Marks for all 29 papers were returned by the set date. All scripts were centrally checked to ensure that (a) every page had been scanned by an Examiner, (b) that the marks noted in the margin were correctly summed to question totals on the cover sheet, and (c) that these totals had been correctly entered on the spreadsheets. This checking process threw up a fair number of errors, overwhelmingly by one or two marks only.

Projects The higher standard of MPhys reports set last year after the introduction of new guidance to candidates was maintained. In the light of last year's exercise the grading forms were slightly revised and a strenuous effort made to improve the homogeneity of the readers. Two Examiners specialized on BA projects and essays, leaving five Examiners to work exclusively on MPhys reports. Each sub-department was asked to nominate readers committed to reading 10 reports; one sub-department provided only one reader committed to reading 20 reports, and one provided three readers. The smaller number of readers improved quality control in three ways: (i) all readers attended an induction session when the Examiners explained what we were trying to do in assessing material, (ii) by reading many reports readers were better able to gauge the appropriate level to look for, and (iii) it was possible to analyze the statistics of all readers and calibrate their efforts.

This year there was an anomalously large volume of work to read on account of the raising of the MPhys hurdle. Consequently it was necessary to ask most readers to read more reports than they were committed to. I an grateful to everyone for being understanding about this.

As last year, each essay or report was read by an examiner from another sub-department and a more expert Assessor, both grading on the same form. When the two marks differed by more than 14, the Examiner and Assessor discussed the report in order to understand each other's position, following which they could revise their marks. The marks of one Assessor were consistently much above those assigned by (various) Examiners. Resolving these discrepancies was hard because the Assessor left the country soon after returning his marks. It was decided to set these marks aside and use marks assigned by a fresh Assessor. In a few other cases marks were obtained from third readings, and these were averaged with those of the first Assessor or Examiner as appropriate.

The left-hand panels of Figs. 3 and 4 shows the correlations between the two marks on each report; this is gratifyingly tight for the BA material, but subject to considerable scatter for the MPhys reports.

In an attempt to reduce this scatter, the raw marks of each Examiner and Assessor were quadratically

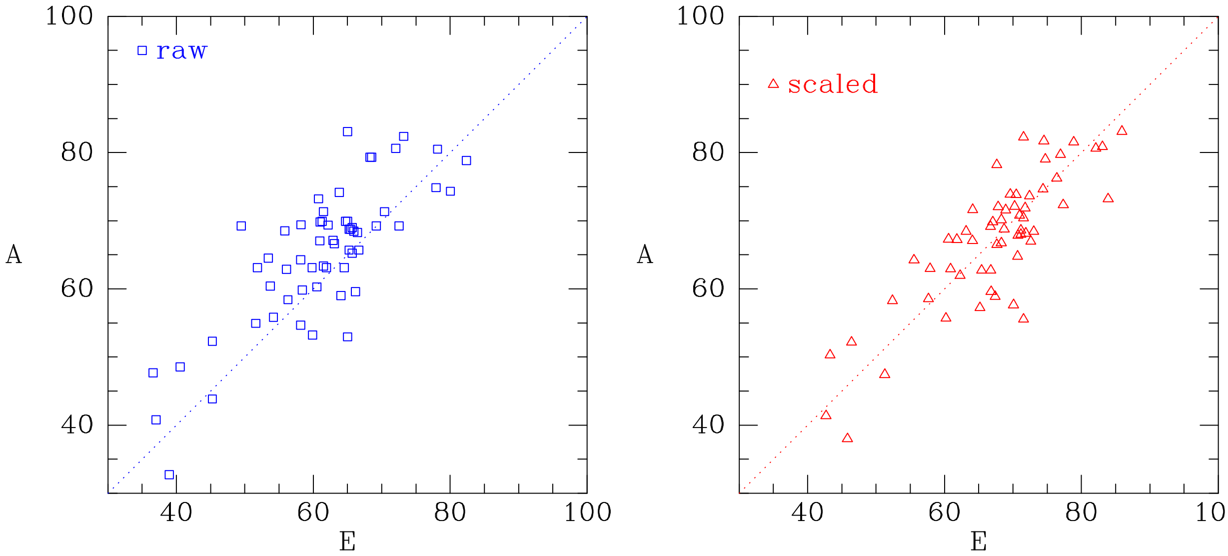

Fig. 3 Marks of BA projects and essays from Assessor and Examiner before (left) and after (right) scaling.

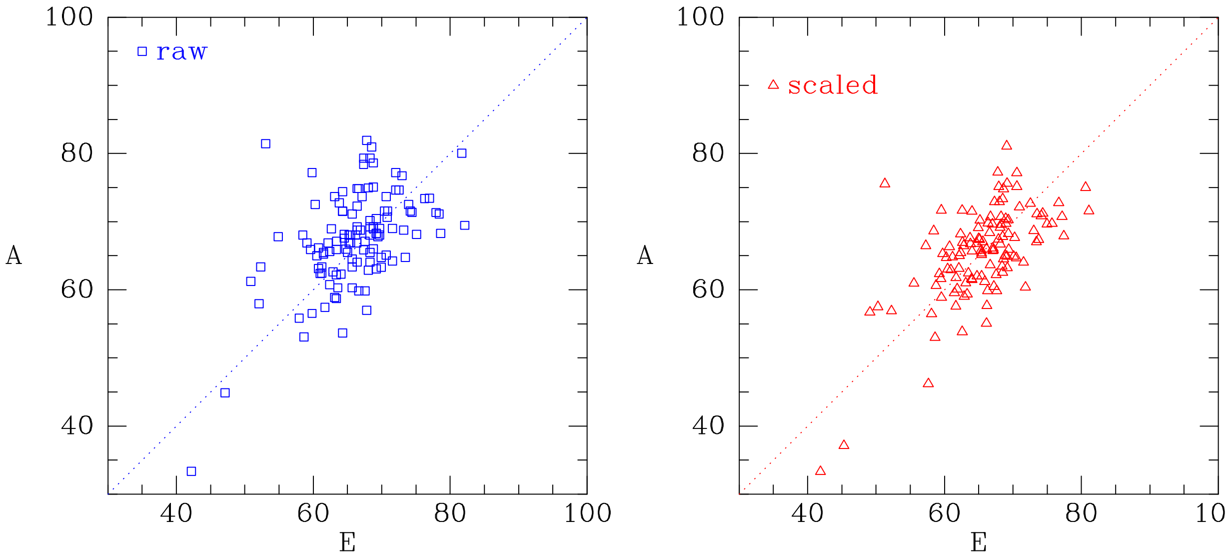

Fig. 4 Marks of MPhys projects from Assessor and Examiner before (left) and after (right) scaling.

scaled to a mean that was determined by the self-calibration algorithm described above: from the perspective of that algorithm, each Examiner and Assessor is a paper, so each candidate has sat two papers and the algorithm assigns mean marks to each Examiner and Assessor such that the sum of the squares of the differences between the marks given to each report are minimized subject to the global mean mark being 65 in the case of the BA and the mean B-paper mark of the candidates in the case of the MPhys. The right-hand panels of Figs 3 and 4 show the scatter between the scaled marks.

Supervisors have always had indirect input to the marks via the questionnaires they fill in, which are read by the Examiner and Assessor before they produce their marks. This year for the first time we tried to obtain a mark directly from Supervisors by asking them in what percentile of the relevant cohort they would place the candidate in the light of their project or essay work. This is a well defined question. Unfortunately, it can only be answered by someone who has some experience of physics candidates, and one supervisor, based in another department legitimately declined to answer the question on the grounds that s/he had no previous experience of MPhys physicists. However, Fig. 5 reveals a different problem with Supervisors' assessments: the students are systematically over-rated, particularly in the MPhys (if the Supervisors' returns were valid, the distributions in Fig. 5 would be uniform by definition). If there were reason to believe that all students

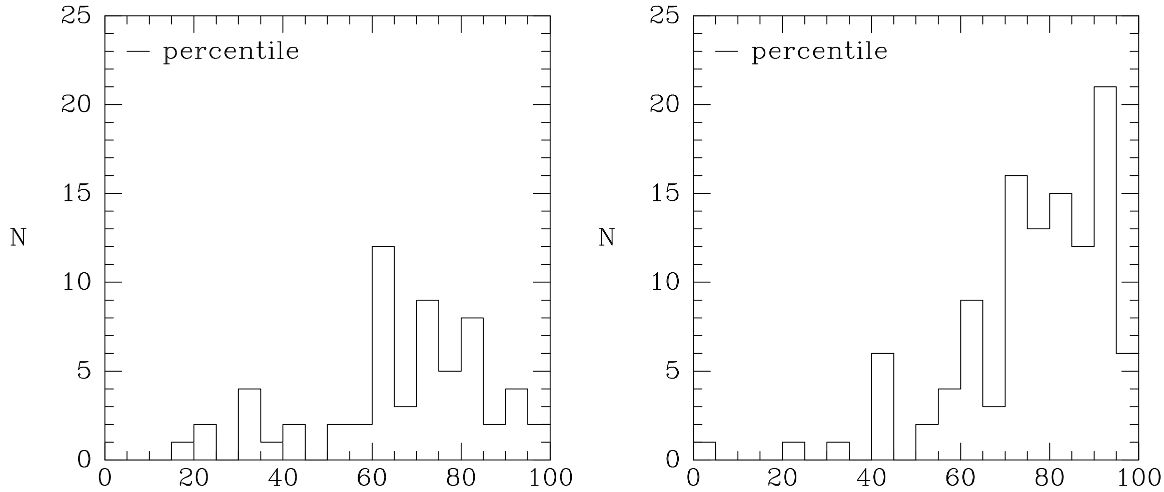

Fig. 5 Supervisors' percentile estimates of candidates for BA (left) and MPhys (right).

benefited equally from Supervisors' rosy spectacles, it would be safe to use a rescaled version of the data. However, it is likely that the percentile assignments of experienced Supervisors are realistic because they can compare the current student with the many others that they have supervised, while inexperienced supervisors systematically over-rate students because they have yet to do a project with a really good student.

In view of Fig. 5 the Examiners came close to discarding Supervisors' percentiles, but decided in the end that they had to use them because in the Examination Conventions they had said they would. The percentiles were converted to marks according to the scale published in the Conventions, and then these marks were *linearly* scaled to a mean of 65 and a dispersion of 5. Fig. 6 shows histograms of the raw (full) and scaled (dashed) Supervisor marks.

The final BA project/essay marks had a mean of 66.4 and a dispersion 9.64.

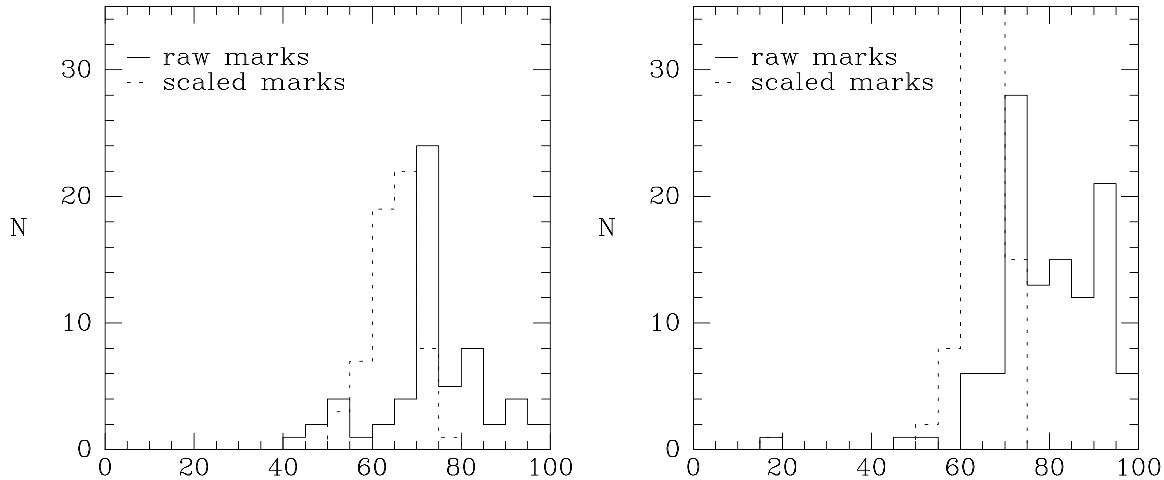

Fig. 6 Raw (full) and scaled (dashed) Supervisors' marks for BA (left) and MPhys (right).

As specified in the Conventions, the total project mark was $0.1S + 0.6E + 0.3A$, where S, E and A are the scaled marks from Supervisor, Examiner and Assessor.

Data Handling Since the end of the 2006 season, a considerable effort had been made to upgrade the data handling system, and on the whole the system worked well this year. Spreadsheets were only used to gather marks into comma-separated files, which were then read into the database. No calculations were performed on spreadsheets. Programs written in C++ that can directly read and write the database were used for scaling, weighting and statistical analysis. A much improved web interface produced tables and graphs for the use of Examiners. The new system for building the database was verified by rebuilding the 2006 database from the original marksheets. The upgraded system resolves all the issues raised in the 2006 Chairman's report with the exception of the automatic production of class lists, which Schools still requires in a format that is hard to produce and intrinsically confusing. Last year the Junior Proctor agreed that this specification should change and it is distressing to note that no progress has been made with this.

Table 3. Percentages in each class of the BA

Year	1	2.1	2.2	3	Pass	Fail
2007	11.7	36.7	38.3	10.0	3.3	0
2006	23.7	23.7	26.3	18.4	7.9	0
2005	8.8	26.5	35.3	23.5	5.9	0
2004	20.5	46.2	25.6	7.7	0	0
2003	10	66,7	6.7	10	6.7	0
2002	12.5	40	30	12.5	5.0	0
2001	14.6	53.7	26.8	4.9	0	0
1996	9.8	28.8	21.6	7.8	2.0	0

Table 4. Percentages in each class of the MPhys

Year	1	2.1	2.2	3	Pass	Fail
2007	37.6	38.5	19.7	4.3	0	0
2006	39.5	41.9	15.3	3.2	0	0
2005	35.7	47.6	11.9	4.8	0	0
2004	36.7	46.7	12.5	4.2	0	0
2003	34.9	53.2	6.3	5.6	0	0
2002	31.1	51.6	15.6	1.6	0	0
2001	35.8	53.3	8.3	2.5	0	0

Table 5. Five-year averages 2001–2006 of various Schools

School	1	2.1	2.2	3	Pass	Fail
Physics 3yr	15.6	41.1	24.4	13.3	5	0
Physics 4yr	36.4	48.0	10.8	3.7	0	0
Maths 3yr	25.6	50.0	17.4	5.5	1.4	0
Maths 4yr	51.2	37.2	3.2	2.2	0.3	0
Chemistry	37.1	48.1	12.6	2.3	0	0
Engineering	32.3	38.1	22.2	6.2	0.6	0
Whole University	24.6	63.8	10.0	1.3	0.3	

Classification Candidates were classified as previously announced in the *Examination Conventions*. Overall percentages were calculated by dividing each candidate's weighted total mark by the maximum mark achievable by this candidate. Then candidates were ranked by decreasing value of this percentage, and lines were pencilled in near 70%, 60%, 50%, 40% and 30%, avoiding locations where adjacent scores were extremely close.

All letters and certificates forwarded by the Proctors over the last three years were reviewed. Cases in which candidates had received compensation for disabilities, such as extra time or use of a computer, were disregarded, as were cases in which the candidate lay far from a class boundary. In other cases when an event had undermined a candidate's performance on a paper, the mark from that paper was dropped from the total and the candidate's potential maximum was accordingly revised downwards in the calculation of a revised overall percentage. Medical certificates caused two candidates who lay just below boundaries to move up a class. One candidate, who had initially been at the very bottom of a class, moved down because s/he had submitted a report late because the Examination Conventions required the mark on the report to be reduced.

These procedures resulted in the class percentages given in Table 3 for the 60 BA candidates classified and in Table 4 for the 117 MPhys candidates classified. The 2001–6 five-year rolling averages for Physics and the other large MPLS schools are given in Table 5. The following comments are in order:

(i) Fluctuations in the statistics for the BA are expected to be significant, especially at the top end. However, it is clear that many fewer firsts are given in the Physics BA than in the Maths BA, and this year our number was at the low end of our low range.

(ii) Physics stands out as awarding many more pass and third-class degrees than other MPLS schools and the University as a whole; over five years the fraction of pass degrees is more than an order-of-magnitude higher than in the University as a whole, while the fraction of third-class degrees in Physics is roughly double that in other MPLS Schools, and larger than that of the whole University by an order of magnitude. This year's proportion of pass degrees is in line with our past practice, while the number of third-class degrees is on the low side, at least in the BA list.

(iii) In both the BA and the MPhys the proportion of students awarded 2.2 degrees shows clear signs of a secular increase. In the case of the MPhys this increase is clearly at the expense of the 2.1 class.

(iv) The sum of the numbers of first and 2.1-degrees is significant because all such degrees are considered "good". This year 66.7% of our graduates have good degrees, while the five-year averages of Chemistry, Engineering and the University as a whole are 85.2%, 70.4%, 88.4%, respectively. This year even the MPhys candidates, who are the elite of an elite, received good degrees in only 76% of cases.

Physics is a rigorous discipline and a good Physics degree is widely regarded as proof that someone can perform at a high level, especially when quantitative thinking is required. We do not want to lead the way in cheapening academic qualifications by unjustly flattering our graduates. Nonetheless, the evidence above that at Oxford it is harder to get a good Physics degree than even a degree in another respected MPLS subject, never mind the humanities and social sciences, is worrying, not least because it may discourage ambitious children from continuing with physics.

The driver behind recent increases in the fraction of poor degrees is easy to identify: the insistence by the EPSC that (i) papers be scaled to a mean mark of 65% and (ii) that class boundaries be aligned with the 70, 60, 50, and 40 percentage marks. When these policies are implemented, as they have been in the last few years, the fraction of students in each class is determined by the dispersion in the marks around the mean. In a rigorous, quantitative subject such as Physics, things are either right (100%) or wrong (0%) and the dispersion in marks is bound to be higher than in biology or jurisprudence, where essays involving nuances, judgments and more or less complete lists of points to mention play the dominant role. On the positive side the higher dispersion in Physics marks increases the fraction of candidates receiving first-class rather than 2.1 degrees. But it inevitably produces a long tail of students with marks below 60%, who are then denied good degrees.

The EPSC wants to standardize marks scales for a respectable reason: students should be equipped with transcripts that include marks. This development will enable employers and grant-givers to distinguish a top 2.1, practically a first, from a bottom 2.1, frankly rather a weak performance, and minimize the impact of arbitrary decisions made at class boundaries. However, the current scheme for standardizing marks scales is deeply flawed and should cease.

I see two possible ways forward: (i) we standardize the dispersion of marks in addition to their mean, or (ii) we decouple the marks actually awarded on a paper from the marks that appear in transcripts, for example by scaling the marks actually gained by first-class students to the interval 100–70, and the marks awarded by 2.1 students to 70–60, and so on. In case (i) a suitable scaling algorithm would have to be selected. I strongly favour case (ii) because (a) it would be easy to implement, and (b) it removes the constraint on paper-setters of aiming for an inconveniently high mean.

It must be recognized that examination papers are, *de facto*, important teaching tools. It is vital that they are stimulating and searching because students will refine their grasp of the subject by working old papers. A paper cannot be stimulating if you require 85% of students to score $> 60\%$ on it. In fact, for our more challenging options, such as Mathematical Physics, the requirement is tougher: as Table 1 shows, we should be setting papers on which the mean mark is $\sim 73\%$ rather than 65%. This requirement is blatantly incompatible with the need to stimulate and challenge the students who choose these options, and we do not fulfill it, despite every effort to push up the mean mark by, for example, moving as many marks as possible to the easy first parts of questions. To be clear, this is not an argument in favour of excessively hard questions. One recognizes such a question easily enough in the distribution of marks – when no student scores nearly full marks, the question is too hard. But a question on which a few students score essentially full marks and there are insignificant numbers of attempts attracting less than 20% is a good question even if the mean mark is as low as 50% because to score 50% a student has to know what the question is about and to show considerable ability to work with the relevant material. Outside the examination room, most students will tease out more than 80% of the question, and the remaining 20% will provide good material for tutorial or class discussion.

James Binney
September 12, 2007

Appendix: the scaling algorithm

First recall the algorithm used in 2006 to scale A and B papers: if x_i is the raw mark of candidate i on a paper, his scaled mark is

$$X_i = x_i + \gamma x_i(x_i - 100)$$

where γ is set by the criterion that $\overline{X} = T = 65$, the target:

$$\gamma = \frac{T - \overline{x}}{\overline{x^2} - 100\overline{x}}.$$

Given a set of N papers called $\alpha, \beta, \ldots$ we want to choose N normalization constants $\gamma^{(\alpha)}$ etc such that the cross-subject discrepancy

$$S \equiv \tfrac{1}{4} \sum_{\alpha,\beta \leq N} \sum_i (X_i^{(\alpha)} - X_i^{(\beta)})^2$$

is minimized. Here the sum over i includes all candidates with marks in the pair of papers (α, β). The idea behind minimization of S is that if $S = 0$, the papers tell perfectly consistent stories about the quality of each candidate.

In principle simple minimization of S fixes the γ factors, but the problem is likely to be ill-conditioned in practice. In any event we want to achieve something like $\overline{X}^{(\alpha)} = T$ for all α. The whole point of the exercise it to allow the means of individual papers to move up or down with the quality of the cohort that sits each paper, but when minimizing S it is sensible to impose the constraint

$$\frac{1}{N_{\rm tot}} \sum_{\alpha=1}^{N} n^{(\alpha)} \overline{X}^{(\alpha)} = T, \qquad (\dagger)$$

where $n^{(\alpha)}$ is the number of candidates for paper α and $N_{\rm tot} = \sum_\alpha^N n^{(\alpha)}$ is the total number of scripts.

We can do this with a Lagrange multiplier: that is, we solve for the $\gamma^{(\delta)}$ the linear system

$$\frac{\partial S}{\partial \gamma^{(\delta)}} - \lambda \frac{\partial T}{\partial \gamma^{(\delta)}} = 0 \quad (\delta = 1, \ldots, N)$$

simultaneously with ($\dagger$). We have

$$\frac{\partial T}{\partial \gamma^{(\delta)}} = \frac{n^{(\delta)}}{N_{\rm tot}} \left(\overline{x^{(\delta)\,2}} - 100\overline{x^{(\delta)}} \right).$$

$$\begin{aligned}
\frac{\partial S}{\partial \gamma^{(\delta)}} &= \sum_{\alpha i} (X_i^{(\delta)} - X_i^{(\alpha)}) x_i^\delta (x_i^{(\delta)} - 100) \\
&= \sum_{\alpha i} \left[x_i^{(\delta)} + \gamma^{(\delta)} x_i^{(\delta)} (x_i^{(\delta)} - 100) - x_i^{(\alpha)} - \gamma^{(\alpha)} x_i^{(\alpha)} (x_i^{(\alpha)} - 100) \right] x_i^\delta (x_i^{(\delta)} - 100) \\
&= \sum_{\alpha i} \left\{ \left[x_i^{(\delta)} (x_i^{(\delta)} - 100) \right]^2 \gamma^{(\delta)} - x_i^{(\delta)} (x_i^{(\delta)} - 100) x_i^{(\alpha)} (x_i^{(\alpha)} - 100) \gamma^{(\alpha)} + (x_i^{(\delta)} - x_i^{(\alpha)}) x_i^{(\delta)} (x_i^{(\delta)} - 100) \right\}
\end{aligned}$$

So the system to be solved is

$$\begin{aligned}
&\sum_{\alpha i} \left\{ \left[x_i^{(\delta)} (x_i^{(\delta)} - 100) \right]^2 \gamma^{(\delta)} - x_i^{(\delta)} (x_i^{(\delta)} - 100) x_i^{(\alpha)} (x_i^{(\alpha)} - 100) \gamma^{(\alpha)} \right\} - \frac{n^{(\alpha)}}{N_{\rm tot}} \left(\overline{x^{(\delta)\,2}} - 100\overline{x^{(\delta)}} \right) \lambda \\
&\quad = \sum_{\alpha i} (x_i^{(\delta)} - x_i^{(\alpha)}) x_i^{(\delta)} (x_i^{(\delta)} - 100) \\
&\sum_\alpha n^{(\alpha)} \left(\overline{x^{(\alpha)\,2}} - 100\overline{x^{(\alpha)}} \right) \gamma^{(\alpha)} = N_{\rm tot} T - \sum_\alpha n^{(\alpha)} \overline{x^{(\alpha)}}
\end{aligned}$$