

Teaching Committee report to the Faculty Board

Examinations for Parts IA, IB and II of the Tripos, 2016

1 Introduction

The Committee met three times, for two hours each, to consider the 2016 undergraduate Tripos examinations. We looked, for each part of the Tripos, at:

- the examiners' report;
- the external examiners' reports;
- the examiners' comments on their questions;
- the examination statistics;
- the examination papers;
- the analysis of the paper lecture questionnaires;
- the responses to the on-line questionnaires;
- the report from the CATAM assessors (Parts IB and II).

We noted with pleasure that as usual the external examiners, without exception, commented favourably both on the examination process and also on the performance of the candidates. The six external examiners' reports include comments such as:

The Mathematical Tripos in Cambridge remains one of the most challenging Mathematics courses in the country, and the examination reflects this. (Part IA)

Overall, I am very satisfied with the standards of the examination and of the qualifications. The processes for assessment, examination and the determination of awards are sound and fairly conducted. (Part IB)

Cambridge's system of compendium examination papers is appropriate and commendable, and on the whole worked very well. (Part IB)

The standard of the examination papers remained gratifyingly high. The examinations have been conducted with scrupulous fairness. (Part II)

The overall standards exceedingly high – I saw many excellent answers to some very difficult questions on some very advanced material. In my view the standard is significantly higher than any other British University except, possibly, Oxford. (Part II)

Intellectual standards on Part II of the Mathematical Tripos are very high, comparing favourably both with other mathematics departments in the UK and with other departments in Cambridge. The Mathematical Tripos Part II is a course that is of great intellectual depth and rigour. (Part II)

There follows a summary of the points raised in the examiners' reports which the Committee believe need the attention of the Faculty Board. We have not generally highlighted points of a purely administrative nature: that is for the Chairs of this year's examiners and the Undergraduate Office to pick up rather than the Faculty Board.

2 General Matters

2.1 Errors

This year only two errors were reported, in a total of about 280 questions: no errors in Part IA; no errors in Part IB; two errors in Part II. In addition, two questions in Part IB were the subject of complaints after the examination (see §4.3). Overall, these numbers are significantly smaller than in recent years, and we

congratulate the examiners, particularly the three chairs, for their diligent and meticulous checking of the questions.

Factors which may have contributed to the low error count are: (i) circulation to examiners of a letter from the General Board complaining about the number of errors on mathematics papers in the previous year; (ii) introduction of a list of typesetting conventions, which allowed examiners to focus more on the content of questions in their meetings; (iii) a possible reluctance of examiners to announce clarifications if not absolutely necessary in order to satisfy the General Board's concerns about the workload for the Registry.

We **recommend** continued use of the typesetting conventions. We also **recommend** that all questions are produced using the LaTeX template from now on (rather than TeX), if necessary with assistance from the Undergraduate Office.

2.2 Rubric

The rubric on Part IA examination papers states:

*“The examination paper is divided into two sections. Each question in Section II carries twice the number of marks of each question in Section I. Candidates may attempt **all four** questions from Section I and **at most five** questions from Section II. In Section II, **no more than three** questions on each course may be attempted.*

Complete answers are preferred to fragments.”

Similar wording is used in Parts IB and II, though the exact rules are different.

Prof Szendroi (Part IA external examiner) comments that “*Candidates may attempt at most N questions from Section X* ” is potentially misleading if some candidates are aware that the actual policy is for examiners to use the best N answers while other candidates might believe that there is some sort of penalty for violating the rubric.

This traditional wording has been debated many times, the issue being a desire to discourage almost all candidates from wasting time on excess questions that will get no marks, while not penalising the few who really do have time to attempt an $(N+1)$ th question as insurance. (Indeed Prof. Szendroi comments that some weaker candidates do hand in many fragments, to their disadvantage.) The wording is stated in the Schedules, and the Faculty policy to excess questions is clearly explained. We **recommend** that candidates are explicitly reminded of this policy this year in the email sent out at the start of the Easter Term that describes the rubric and cover sheets. We **recommend** that the sentence “The policy towards excess attempts is described on p.2.” is added to the IA section of the Schedules.

Of course, “*Candidates may attempt at most N questions from Section X* ” is not strictly true, because candidates may attempt more, and they will be marked, and then the best N marks will be used. It was suggested that “*Candidates may obtain credit on at most N questions from Section X* ” would be a truer statement. We also noted that the statements about Section II questions carrying twice as many marks as Section I questions and complete answers being preferred to fragments, while strictly true, are not a very clear representation of the actual system involving alphas and betas. A wording which reflects these concerns is:

“The examination paper is divided into two sections. Each question in Section II carries twice the number of marks of each question in Section I. Section II questions also carry an alpha or beta quality mark and Section I questions a beta quality mark.

*Candidates may obtain credit from attempts on **all four** questions from Section I and **at most five** questions from Section II. In Section II, candidates may obtain credit from attempts on **no more than three** questions on each course.*

Part III has moved to the direct style “Attempt N questions” with no explanation about what happens if one does more! Does Faculty Board have views on how to balance accuracy, brevity, clarity, and direction?

2.3 Short questions

The beta threshold for short questions in 2015/16 was reduced from 8/10 to 7/10, and the target success rate for betas was been raised from 55–60% to 65–70%. The intention was to make short questions more attractive, particularly to weaker students who struggle to complete long questions. The outcome is shown in the following table.

	beta rate (%)			average mark			takeup in II	
	IA	IB	II	IA	IB	II	short	long
2016	59	59	57	6.8	6.8	6.5	17.4	17.1
2015	44	52	51	6.7	7.0	6.8	12.4	15.8
2014	60	51	48	7.2	7.0	6.6	12.9	15.5
2013	59	52	47	7.5	6.9	6.6	12.2	15.3

The success rate did increase, but not by as much as hoped. More encouragingly, in Part II the take-up of short questions increased by 50%, reflecting greater use of C courses.

We again **recommend** that pressure is maintained to keep short questions sufficiently short and easy.

2.4 Administrative and Computer Officer support

The examiners' reports remarked on the excellent support given by Ms Amy Dittrich and Mr John Sutton, in particular, and the administrative staff more generally. We **recommend** that the Chair of the Faculty Board expresses the appreciation of the Board for their hard work.

3 Part IA

3.1 Short questions

The examiners' report states incorrectly that their 59% beta rate (on average) on short questions was at the upper end of the Faculty recommended range. In fact, the recommended range is 65–70%, which was almost achieved by Pure, but significantly undershot by Applied.

3.2 Difficulty of questions

The 17% alpha rate for the long questions on Groups was well outside the Faculty recommended range of 40–45%. We ask this year's examiners to pay careful attention to previous years' statistics and the guidelines on question setting in the letter from the Chair of the Faculty Board.

4 Part IB

4.1 Class descriptors

Prof. Anderson and Prof. MacKay (external examiners) make a number of comments about the verbal class descriptors in the Schedules, which they seem to have understood to be the '*other factors besides marks and quality marks that may be taken into account*' and then tried to use them as secondary criteria to make decisions about individual cases near the boundaries. They found phrases such as 'a fair number of correct answers to straightforward and challenging questions' too vague to be useful and, if interpreted as a threshold on alphas and short betas, it was not clear where CATAM fitted in.

We agree that the verbal descriptors are too vague to be used to make detailed decisions about individual cases, and we believe this is inevitable because any class descriptor has to cover the considerable range of achievement within a given class, the variety of possibilities for accruing credit from a mixture of long and short questions and CATAM, and the possible variations in difficulty (and hence of borderlines) from one year to the next. Indeed, it was for this reason that the Faculty adopted the merit mark as the primary, and quantitative, classification criterion.

As the Schedules say, the intention is that *after* applying the classification criteria (merit mark, other factors, approximate percentages) to make detailed decisions, the classes can be characterised by the verbal descriptors. We **recommend** changing the wording to 'broadly characterised' to make this clearer. The Schedules also say that '*stronger performance on the Computational Projects may compensate for weaker performance on the written questions or vice versa*', which applies to all the descriptors.

What then are the 'other factors'? They are secondary factors, *other than* the values of m , α or β , which might tip the balance in marginal cases or which, in exceptional cases, might need to be considered in the interests of fairness. Examples might include whether the alphas are mainly 20s or 15s (or if there are a lot of 14s), whether the layout of the answers shows particularly clear or unclear thinking, or whether the questions attempted were on courses that were generally found difficult or easy (this is particularly relevant for fairness in Part II where candidates take very different selections of courses). It is anticipated

that such factors are more likely to affect the location of the borderline than to cause a departure from merit-mark order. Only in exceptional cases should there be a significant departure from merit-mark order.

We are reluctant to add too much prescriptive detail to the Schedules in case this ties the examiners' hands in unusual situations and prevents them from exercising appropriate judgement. On the other hand, we **recommend** that the instructions to examiners in the Chair of Faculty Board's letter be expanded to describe, as above, the nature and role of other factors.

Prof. Anderson and Prof. MacKay also commented that the verbal descriptor for the 3rd class describes a lower level of achievement than would be usual for a 3rd class at other leading universities. We note that questions on Tripos examinations are probably more demanding than those elsewhere, but agree that the description is overly negative. The current wording is

Candidates placed in the third class will have demonstrated some knowledge but little understanding of the examinable material. They will have made reasonable attempts at a small number of questions, but will have lacked the skills to complete many of them.

We **recommend** that it be amended to

Candidates placed in the third class will have demonstrated some knowledge of the examinable material. They will have made reasonable attempts at a small number of questions, but will not have shown the skills needed to complete many of them.

4.2 Number of meetings

By force of circumstances, the IB examiners again only had two meetings to consider the draft questions, though they had intended to add a third meeting as in Parts IA and II. We strongly **recommend** that they do have a third meeting this year to improve the robustness of the setting process against divergences in difficulty and glitches in content.

4.3 Complaints about questions

Shortly following the examination, two complaints were received via the student feedback email about the content of two questions. One question on Metric & Topological Spaces started by asking for a definition of the p -adic metric on Q , though this had not been defined as an example in lectures in 2015. One question on Geometry asked for a definition of cross-ratio, which was not lectured in Geometry, but had been defined in IA Groups (where it is mentioned in the Schedules). Unfortunately, the last part of the question depended on using a different convention from the one used in Groups.

We are satisfied that the examiners took appropriate action to deal generously with those candidates who had got stuck on the last part of the second question by using the wrong convention. We are also satisfied that the questions could have been entirely suitable and on Schedule if either the question or the lecturer had given the relevant definitions. However, we are very sympathetic to the student complaint, and regard this as a failure of the examiner and the lecturer to communicate sufficiently about what students should be expected to know. Of the three undergraduate years, Part IB is the most vulnerable to a gap between what the examiner expects and what students have actually been taught. We **recommend** that the IB examiners take particular care to discuss their questions with lecturers, especially for the Easter Term courses where **both** years of possible lecturing should always be taken into account.

4.4 Syllabi and content of courses

Prof. Anderson (external examiner) comments that the syllabi provided in the Schedules are a bare list of topics and results, and did not give him sufficient indication what students are expected to be able to do. (This comment may well be connected to the complaints about the questions above.)

The Teaching Committee believes that students should be familiar with all the material lectured that is on the Schedules, and with the content of the example sheets, and should be able to use this information to attempt examination questions, which may contain both bookwork and novel problems, of the type found on past papers. Both example sheets and 16 years of past papers are available online for examiners and students to use as a guide, and students should have worked through at least 3 or 4 years worth of past papers. (It is **not** expected that students should be familiar with any of the recommended text books.)

4.5 Availability of paperwork

Professor Anderson (external examiner) asks if some of the paperwork, such as the statistics on questions, could be made available by email on the evening that the external examiners arrive in Cambridge. The

timetable is very tight, and sometimes careful cross-checking between the scripts, the cover sheets and the mark list is still in progress. Moreover, the support staff are very busy processing results for all Parts of the Tripos almost concurrently. Nevertheless, we **recommend** that, if possible, the papers be sent electronically to the external examiners before their day of examining scripts.

He also asks about statistics on questions from previous years. These are sent to all examiners in November, and serve an essential purpose at the time of setting the questions. It was not clear to us why they might be useful at the time of classifying the candidates.

4.6 Rubbish sacks

The Part IB examiners recommend reconsideration of the policy of searching the rubbish sacks if a candidate claims a question on the master cover sheets but there is no attempt in the bundled script. They had five such cases, and search of the sacks produced three fragments with matching handwriting all worth zero marks. The Part II examiners had three such cases, and judged that two were clearly mistakes in filling in the cover sheets, while the third candidate was so far from a border that it would make no difference. A fourth candidate in Part II had failed to fill in their cover sheet, but subsequently appealed because their mark breakdown didn't show a question they claimed to have done. A search of the sacks did discover the candidate's answer which was marked 18/20, though this did not change the candidate's class. In recent years, there have been cases where candidates have claimed questions on the cover sheet, and a search of the sacks did produce substantive answers.

While some of us felt that it is the candidates' fault if they don't hand in an answer, the University policy of retaining the sacks suggests that we are expected to be more sympathetic. We **recommend** that examiners should (as in Part II) search the sacks if it is possible that full marks on a missing attempt would change the candidate's class, but otherwise wait to see if the candidate makes an appeal. The examiners should, of course, always check that the missing attempt has not simply been overlooked or put in the wrong bundle.

4.7 CATAM: plagiarism

New guidelines on plagiarism allowing for stricter penalties came into force this year, which were very much welcomed by the Part II external examiners who had been arguing strongly for this for several years. These guidelines were clearly advertised to the students, and it is thus very disappointing that a case came to light in Part IB for which it was appropriate to impose the maximum penalty of removing all marks for CATAM. Plagiarism is a serious offence and an assault on the integrity of the examination, and we **recommend** that examiners impose the maximum penalty appropriate whenever it is discovered.

(We note, as an aside, that the submission rates in Part IB (98%) and Part II (96%) continue to rise.)

5 Part II

5.1 Mark schemes

The examiners' reports contain conflicting comments on the level of detail appropriate for mark schemes. We do not wish to be overly prescriptive and simply state two principles: the draft mark scheme must be sufficiently detailed for the external examiner to judge the level of the draft questions; the examiners are responsible for the final mark scheme, and are at liberty to modify, refine or coarsen the suggestions made to them by the lecturers.

5.2 Difficulty of questions

Most of the courses identified last year as having an anomalously large number of alphas were moved back towards the average, the exception being Number Fields (still up at 79%). It was, however, again the case that more of the lowest scoring courses, as revealed by the bar charts of marks *after adjustment* for student effect, were Applied and more of the highest scoring courses were Pure.

Prof. Walters (external examiner) was concerned about the effect of disparity between courses, and recommends consideration of adjusted marks during classification. There are various reasons why this should be approached with great caution, in particular because good students will choose easier questions and should be rewarded for that, and also because it is far from clear how statistically reliable the separation of course effect and student effect actually is. We suggest that an exploratory analysis of past results might be considered if the computer-officer time and statistical expertise are available.

Meanwhile, we again **recommend** that this year’s examiners make every effort to reduce differences in difficulty between questions on various courses, and that they inform the relevant lecturers of last year’s issues and the desired direction of travel towards the middle of the range. The letter from the Chair of the Faculty Board to the 2017 examiners has been strengthened to emphasise this point.

5.3 C-course long questions

Prof. Walters (external examiner) observes that he understood the guidance in the Schedules that ‘*no distinction be made, for classification purposes, between quality marks obtained on the Section II questions for C courses and those obtained on D courses*’ to mean that these questions should be of the same standard. We agree. On the other hand, he has twice encountered examiners who thought that even the long questions on C courses ought to be easier. We disagree with these examiners. We **recommend** either that the phrase ‘and that these should thus be of comparable difficulty’ be added to the Schedules, or that something be added to the instructions to examiners.

5.4 Athena SWAN

The examiners’ report did not include a breakdown of results by gender as requested by Faculty Board. We remind this year’s Chair of Examiners to do so.

6 Summary of recommendations

(The exact recommendation is described in the section indicated.)

- 2.1 Use typesetting conventions. Use LaTeX rather than TeX.
- 2.2 Email and Schedule additions concerning marking policy towards excess questions.
- 2.3 Pressure maintained to keep short questions genuinely short.
- 2.4 Thank Ms Amy Dittrich and Mr John Sutton for their efforts last year.
- 3.2, 4.5, 5.2 Examiners and lecturers to actively consider previous years’ statistics and adjustments necessary to achieve a more uniform level of difficulty.
- 4.1 Two changes to Schedules regarding class descriptors. Addition to Chair’s letter regarding ‘other factors’.
- 4.2 IB examiners to have an extra meeting to discuss questions.
- 4.3 Discuss draft questions carefully with all relevant lecturers.
- 4.5 Paperwork to be sent out electronically if available in time. Statistics to be sent to externals in November.
- 4.6 Search rubbish sacks for missing claimed questions only if it might make a difference to a class, or on appeal.
- 4.7 Impose the maximum appropriate penalty for plagiarism in CATAM.
- 5.3 C course long questions to be of comparable difficulty to D course questions.

Nilanjana Datta	Claudia Feng	Henry Wilton	Berry Groisman
Richard Jozsa	Peter O’Donnell	George Robinson	Henrik Latter
András Zsák	John Lister (Chair)		

November 9, 2016