UNIVERSITY OF Southampton
Mathematical Sciences

The Vice-Chancellor,
University of Oxford

August 29, 2018

Dear Vice-Chancellor,

External examiner's report 2017/18:
Part C, Mathematics, Mathematics & Statistics, Mathematics & Computer Science

I have pleasure in enclosing my report on the above examination.

B1. Academic Standards

(a) The portfolio offered by Oxford Part C continues to be one of the most extensive and challenging in the UK in the areas of pure and applied mathematics and statistics. The academic standards achieved by the students continue to be at least as high as anywhere else in the UK.

(b) The performance on the project work continues to demonstrate an extremely high standard with several that were certainly publishable.

B2. Rigour and conduct of the assessment process

(a) The examinations in part C are rigorous and all of (at least) a level 7 FHEQ standard.

(b) The quality of preparation of the papers continues to be extremely high. I thank the assessors for their considered responses this year to my comments on their papers.

(c) From the viewpoint of the external examiner, the administration process is extremely efficient. WebLearn continues to provide and extremely streamlined and secure way to handle the scrutiny of the process. I again single out the dedication of the administrative staff for special praise.

(d) I am grateful for the detailed response from the Head of Policy (Taught Courses) on 12th April 2018 concerning the points I raised last year about historical norms, high scaling gradients and scaling of project marks.

B3. Issues

(a) I am grateful for and satisfied by the response from the Head of Policy (Taught Courses) to my comments raised in 2016/17 regarding the scaling of project marks.

(b) I am grateful for the response from the Head of Policy (Taught Courses) regarding scaling to historical norms, and to high scaling gradients. I acknowledge the guidance that has been given to examiners. However, this year the final exam board for Part C was again dominated by the questions of scaling, some repeating my concerns from last year.

i. Last year I referred to the University guidance that:

'Scaling should not be used mechanistically to fit the spread of classes on a paper to historical norms (i.e. norm referencing).'

I note the official response to my concerns over scaling last year that,

Mathematical Sciences, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 8059 5145 Fax:+44 (0)23 8059 5147 http://www.southampton.ac.uk/maths/

"The division does therefore agree that, when considering the marks of students on a fourth year option paper, it is valid to take into account the marks of the same students in the previous year. This is why the University guidance emphasises that scaling should not be used mechanistically …. Whilst there is understandable broad consistency in the class percentages there is sufficient fluctuation to suggest that Examiners do employ their discretion, and such is occasionally commented on in the Examiners Reports."

The Final Examination Board for Mathematics Part C commenced this year as if nothing had changed. The internal examiners had already spent a day prior to the arrival of the external examiners scaling the marks to obtain similar distributions of classifications for incoming and exiting cohorts on each paper. From the very existence of this prior internal scaling meeting, the presumption in Mathematics appears to be that there should be **systematic** (if not mechanistic) norm referencing on individual papers. The external examiners then have an uphill task to convince the internal examiners to move away from the prior scaling assumptions, as evidenced by the next point.

ii. The result of the internal scaling meeting was again significantly high scaling gradients at the top and bottom of the marks distribution. For example, at the low end, initially we were presented with mappings such as
   - C1.3: 2.7 raw $\to$ 37 USM ,
   - C1.4: 2.85 raw $\to$ 37 USM ,
   - C2.4: 6 raw $\to$ 37 USM , etc.

   Eventually after some extended "frank" discussions with the internal examiners, the following "working hypothesis" was agreed: if a student scores x at the lower end of the raw mark scale, this should be scaled to no more than $2x + 15$ on the USM scale.

   This means that a student on Mathematics Part C at Oxford has only to write their name on the paper and they will receive 15 USM marks.

   That external examiners are placed in the position that they must haggle with the internal examiners to arrive at this compromise is extremely regrettable and is entirely due to the prior assumption of cohort-based scaling for norm referencing. For several candidates, this variation in scaling made a difference to rank orderings and, in at least one case, affected their classification.

   I was, and remain strongly against even having the +15 systematic shift, especially at the lower end. If not "mechanistic" scaling, it is a "mechanistic" solution to the issues caused by the presumption of systematic scaling.

iii. The net effect of the systematic use of scaling at module level has been a monotonic increase since 2014 of the proportion of firsts from around 45% to 55%.

   There is no evidence to suggest that the Part C examinations have become easier, that the means of delivery has changed materially to boost outcomes, or that Oxford students have become as more able as the Part C results suggest.

   The corresponding percentage of first class classifications of Part C students at Part B over the same period has only risen from 40% to 44%. Hence the (also scaled) Part B marks suggest some potential improvement in the cohort at the top end, but only

at half the rate achieved at part C.

Over the same period, the percentage of firsts and 2is combined has almost monotonically decreased from around 89% to 84%.

Hence the net effect of systematically scaling at individual paper level appears to be a boosting of candidates at the 2i/first borderline. Only after all the individual papers have been scaled is the overall distribution of classifications within the cohort considered, by which time it is extremely difficult to unpick this boosting.

iv. When one paper was discussed that required a significant amount of scaling to replicate the ingoing to outgoing proportion of classifications, one of the internal examiners stated that "this paper is always difficult and needs a lot of scaling".

I appreciate the natural (perception of) variance in the difficulty between mathematical subjects, especially at Part C level. However, if certain papers are known to be consistently more "difficult", there should be more robust formal mechanisms (exam board direction, vetting, change of assessor, etc.) for ensuring that future papers are set at an appropriate standard, or not at all.

Allowing such outlier papers to continue at best restricts student choice, arguably means that those taking these papers are exposed to the potential lottery of scaling and, in view of the comments above, at worst is potentially giving some of these students an unfair boost merely for taking the paper, not necessarily linked to their actual abilities.

The systematic level of scaling at Oxford is the greatest I have seen in the six universities (including Cambridge and Imperial) at which I have been involved in examinations as an internal or external examiner. This need for this might be minimised with more concerted action to ensure more homogeneity in the difficulty of the papers.

v. Given that Oxford is one of the most selective of Universities, and Mathematics as a subject especially so, the argument as to whether a cohort is weaker or stronger than previous ones should be less relevant, especially if classification descriptors are the guiding principle as the guidance states. The corollary is that the quality of the Oxford degree is variable, depending for a large part on the particular cohort that a candidate is in.

The default process should be that prior to the final exam board, there is a simple multiplication of raw marks by 2 (maximum of 50 raw marks on the paper to the 100 mark USM scale). The papers should then be scrutinised at the borderlines by the final exam board (including the external examiners) informed by the assessor comments, to see if the marks at borderlines are consistent with the classification descriptors. Only then should scaling be considered.

Papers that, over a number of years, consistently need significant scaling can then be identified and appropriate (non-scaling) action taken to address this.

(c) I was informed that Statistics use Turnitin for plagiarism detection on all projects, but that Maths apparently do not consistently do so. I suggest that there should be a consistent approach, especially when compulsory projects are introduced for 2020.

Mathematical Sciences, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 8059 5145 Fax:+44 (0)23 8059 5147 http://www.southampton.ac.uk/maths/

(d) I was surprised to learn that a candidate taking an integrated masters in Mathematics with Computer Science is not required to take any mathematics (or indeed any computing) modules at level 7. Hence at best the two candidates who chose this route (one either way) this year have only achieved at level 6 in just one of the named subject areas in the level 7 integrated masters degree title.

(e) It appears that candidates in one module taken by both the Part C and MSc Computing were marked by different examiners. To "correct" this Computer Science initially decided to scale the Part C candidate (according to my notes, down), despite both cohorts being level 7. Fortunately both the external examiners at the meeting objected to this and this was changed. Whilst errors in process can always occur, I am surprised that the decision was taken to correct via down-scaling Part C candidates, rather than getting one examiner to remark the other sub-cohort. This is further evidence of a bias towards "mechanistic" scaling as a first tool of resort.

(f) I am (pleasantly) surprised by the low numbers (only a few) of medical and special circumstances that were brought to the attention of the board. In my experience at several universities, I would have expected a far higher proportion of students to submit such cases, especially when the stakes are so high. Whilst the processes that students need to go through to have special considerations taken into account may be rightly blocking the frivolous, might they also be dissuading some students with genuine cases of merit?

B4. Good practice and enhancement opportunities

The following items continue to be examples of good practice.

- I commend the extensive range of examination topics offered.
- I commend the care with which the examination papers were produced and administered
- I commend the high quality and challenging nature of the questions that have been set.
- I commend the comprehensive comments from the examiners based on examination performance.
- I commend the generally high level of achievement by the cohort of candidates.

I believe this is my last year as an External for Part C. I wish to record my thanks for the care with which the department has listened and responded to my concerns and the hospitality which has been shown to me.

Yours faithfully,

C.J.Howls
Professor of Mathematics

Mathematical Sciences, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 8059 5145 Fax:+44 (0)23 8059 5147 http://www.southampton.ac.uk/maths/