UCL Academic Manual 2024-25

Chapter 4: Assessment Framework for Taught Programmes

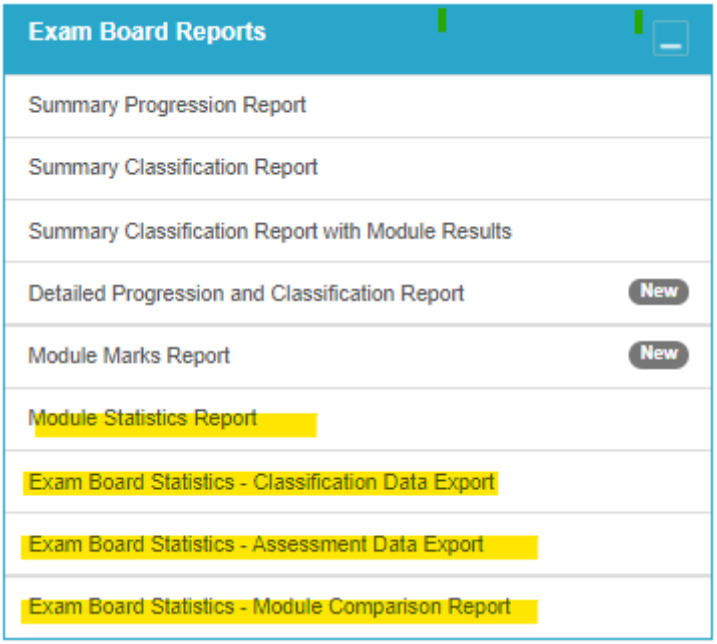
Annex 4.3.6

Guidance for Boards on using the Exam Board Statistics Reports

Introduction
This guidance has been drawn up in consultation with a group of Examination Board Chairs and also draws on existing guidance from the Faculties of Engineering, Arts & Humanities and Social & Historical Sciences and the Division of Psychology and Language Sciences.

Reports available and their uses
The following reports are available in the Exam Board reports container in Portico.

- Module Statistics Report: this provides detailed information for each module in the selected academic year, including a histogram showing mark distribution.
- Exam Board Statistics - Module Comparison Report: this provides a summary line for each module, enabling comparisons between modules and academic years.
- Exam Board Statistics - Classification Data Export: this provides a data file for more detailed analysis if required.
- Exam Board Statistics - Assessment Data Export: this provides a data file for more detailed analysis if required.

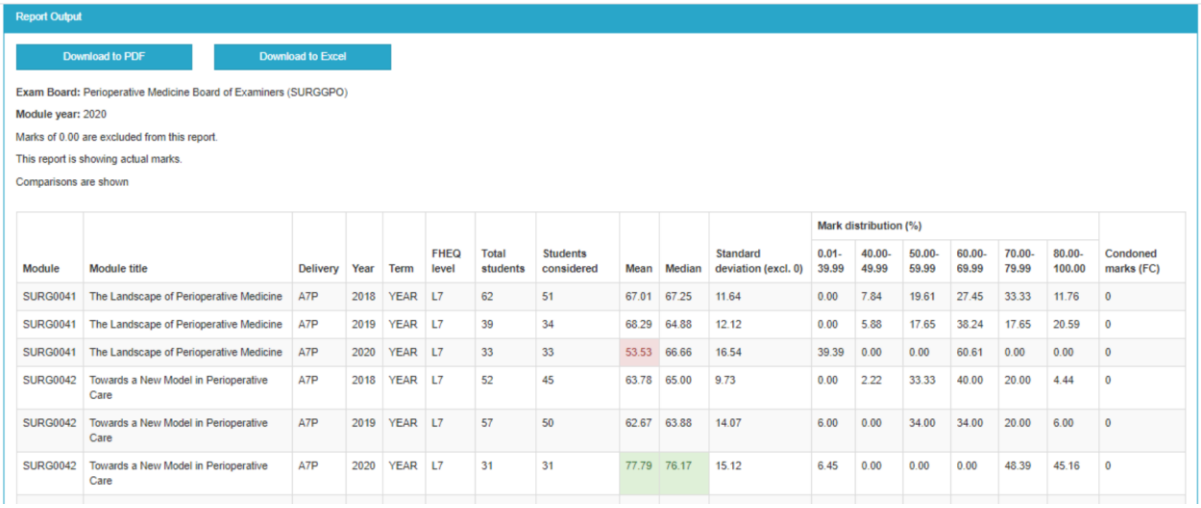


The **Exam Board Statistics - Module Comparison Report** is recommended as a minimum for use in Boards to enable them to carry out their responsibilities under regulation 13.3.7 to

- ensure that assessment, marking and moderation processes are appropriate, rigorous and fair.
- ensure equity of treatment for students.

Further details of reports and how to run them are available on the Exam Board Support website

Using the Exam Board Statistics - Module Comparison Report for Board Meetings



Report Output

Download to PDF | Download to Excel

Exam Board: Perioperative Medicine Board of Examiners (SURGGPO)

Module year: 2020

Marks of 0.00 are excluded from this report.

This report is showing actual marks.

Comparisons are shown

Module	Module title	Delivery	Year	Term	FHEQ level	Total students	Students considered	Mean	Median	Standard deviation (excl. 0)	Mark distribution (%)						Condoned marks (FC)
											0.01-39.99	40.00-49.99	50.00-59.99	60.00-69.99	70.00-79.99	80.00-100.00	
SURG0041	The Landscape of Perioperative Medicine	A7P	2018	YEAR	L7	62	51	67.01	67.25	11.64	0.00	7.84	19.61	27.45	33.33	11.76	0
SURG0041	The Landscape of Perioperative Medicine	A7P	2019	YEAR	L7	39	34	68.29	64.88	12.12	0.00	5.88	17.65	38.24	17.65	20.59	0
SURG0041	The Landscape of Perioperative Medicine	A7P	2020	YEAR	L7	33	33	53.53	66.66	16.54	39.39	0.00	0.00	60.61	0.00	0.00	0
SURG0042	Towards a New Model in Perioperative Care	A7P	2018	YEAR	L7	52	45	63.78	65.00	9.73	0.00	2.22	33.33	40.00	20.00	4.44	0
SURG0042	Towards a New Model in Perioperative Care	A7P	2019	YEAR	L7	57	50	62.67	63.88	14.07	6.00	0.00	34.00	34.00	20.00	6.00	0
SURG0042	Towards a New Model in Perioperative Care	A7P	2020	YEAR	L7	31	31	77.79	76.17	15.12	6.45	0.00	0.00	0.00	48.39	45.16	0

The report should be used as part of the initial moderation process (using 'actual' marks) to identify any modules with anomalous mark distributions that the Chair of Examiners and Board need to consider and at the Board of Examiners (using 'agreed' marks) to identify trends that may need follow up action. See 'Follow up Actions' below.

The report will enable you to review how the key statistics for a particular module compare

a) to those of previous years for the same module;[1] and

b) the corresponding programme or departmental averages.

Questions to ask

The baseline assumption is to expect comparable results (mean, distribution) for modules of the same level. In practice, some variation in cohort performance is to be expected and the same applies to some module groups. The objective of statistics is to be able, in the longer term, to quantify and rationalize this expected variation and also allow easy identification of anomalous sets of results.

Modules of potential concern are those that are significant outliers beyond the normal band of variation and/or modules that run consistently and significantly higher or lower than the average over a number of years. The colour coding on the report will help to identify those more easily. Attention should be paid to significant changes in the standard deviation of marks on a module, for example.

Modules with very small numbers of candidates will not provide a reliable basis for review – a minimum size of 20 is advised.

[1] Marks from 2018/19 onwards will be available. However, for some modules comparable 2019/20 marks may not be available owing to Covid mitigation measures (for example the capstone assessment replaced first year assessment.)

- What are the reasons for the results produced by these modules?
- Are these reasons acceptable?
- Has there been some drift in the distribution (mean, standard deviation) that needs to be corrected and brought back in line with the programme?

Follow-up Actions

If the module statistics and discussion identify modules in need of realignment, then the course of action depends on the extent and urgency of the matter.

Adjusting marks for the current year

It should be considered a fundamental principle of assessment that the assessment is set at a level appropriate to the module in question. Some variation in the mark distribution may occur from year-to -year for a number of acceptable reasons. However, a large-scale shift of the distribution either up or down the mark scale would generally be taken as an indication that the level of the assessment had been judged incorrectly by the examiner(s). **Scaling** *may* be used as part of the moderation process in such cases.

Scaling would normally be undertaken in advance of the Board meeting, in consultation with the external examiner and reported at the meeting for ratification. Approaches to scaling may vary according to discipline and type of assessment. Sometime, alternative scaling approaches may be trialled and the most appropriate selected – in consultation with externals. The results of such trials and the reasons for the final outcome should be reported to the Exam Board with opportunity for comment. See Annex A for examples of approaches to scaling and the process to follow.

Other actions

Other issues, such as modules that consistently have higher or lower than average marks will need further investigation. Any remedial action will depend on the root causes identified.

For example,

If performance in the module is consistently above average, then it might be appropriate to review the content and/or assessment of the module.

If performance in the module is consistently below average, it may also be necessary to provide more information to potential students about the prior level of knowledge required.

Note: there is an opportunity in the period immediately after Board meetings to request urgent changes to assessment patterns for the following academic year.

Annex A Guidance on Scaling

1. A simple translation of the mark distribution (e.g. “add five to all marks”) is only acceptable if it does not penalise the top performing students (i.e. if the upper limit of 100% does not result in these students failing to benefit equally from the scaling).

2. The average mark after scaling up should not exceed the average of the average marks over the past three years.

3. Scaling does not have to be uniform across the mark range but should not be “unreasonably” (dis)advantageous to any one group of students.

The Scaling Process

The need for scaling must be identified BEFORE the meeting of the Board of Examiners. It is a fundamental principle that scaling must be approved by a full, quorate meeting of the Board.

1. The module leader, programme leader or Chair of the Exam Board reviews the marks and identifies a possible need for scaling.

2. The Chair of the BoE, the module leader and the programme leader discuss possible scaling scenarios and agree on one or more such scenarios to present as options to the BoE.

3. The scaling scenarios will be sent to the External Examiner(s) for comment ahead of the BoE.

4. The module leader, Programme Leader, Chair of the Board and External Examiner(s) will agree on one scaling algorithm (or that no scaling should be applied). Where no overall agreement can be reached, the Chair of the BoE has the final say in this decision.

5. If (4) leads to scaling:

a. both the scaled and unscaled mark distributions are distributed to members of the BoE ahead of the meeting, with invitations to comment in advance of the meeting.

b. the unscaled marks are entered into Portico as the “actual” marks; the scaled marks are then entered as the “agreed” marks and used in the determination of progression and award results

c. A summary of the discussions related to the scaling, along with the scaled marks, details of the algorithm used are presented at the meeting of the Board of Examiners for ratification.

MCQ assessments

The following guidance from Psychology is provided as an example of an approach for MCQ assessments.

Although it is tempting to see rescaling as merely a statistical manipulation, rescaling is not a statistical process but a judgemental process whereby the examiner uses their expert knowledge of the content of questions and the expected achievement of candidates to set a mark corresponding to particular borderlines described in the marking criteria.

Rescaling is most easily carried out if, after looking at an MCQ paper, an examiner decides that, say, 42 items correct should correspond to a borderline first (‘70’) as described in the Criteria. Similar borderlines for other key points on the scale may be 35 for the II.ii/II.i borderline (‘60’), 25 for the II.ii/III borderline (‘50’), and 15 for the III/Fail borderline (‘40’). The latter should take into account that there is a one in four guessing rate for best-of-four MCQ

questions, which is 12.5 correct out of 50. Note also that the intervals do not have to be equally spaced, although they should be monotonic.

Marks on the raw scale are then linearly interpolated to the standard scale described in the Criteria. That is most easily carried out using a spreadsheet, and two spreadsheets are available, one aimed at MCQs (where raw marks are out of 50), and the other for exams with different marking schemes, and for moderation, where marks to be rescaled are already in the range 0 to100.