

Edinburgh Medical School Knowledge-Based Assessments

Professor David Kluth
Director of Undergraduate Medical Teaching.

This document aims to outline the standard approach for construction and standard setting of multiple-choice questions for medical exams across all years. The objective is to achieve consistent style so that students will be familiar with exams and enable comparison across the years.

House style.

We have two current excepted questions styles:

- Single best answer questions (SBA) with normally 5 options
- Very short answer questions (VSAQ) which are the same in question design but the candidate must type the answer

Document outlining the “house-style” has been produced. This will be backed up with an online presentation on MCQ writing available via Media hopper (urls are at the end of this document). Some points to note are that negative questions (e.g. “which of the following does not occur..”) and the use of multiple true/false with one true answer should be avoided where possible. However, for basic science the latter may sometimes seem the only way to test certain knowledge. We should also aim to avoid questions that cue the correct answer. Examples are provided in attached guides: *Writing single best answer questions Edinburgh 2016*, *How to make better MCQs-A practical guide* and *Writing Very short answer question for the MBChB-A guide for staff*

Timetable for exam production and oversight

The quality and size of the question bank is key to producing reliable exams. This requires question writing and review meetings. Each assessment should aim to produce the following timetable of question writing, working back from the exam date.

- a) Identify question writers for each section/specialty and set a target of number of questions (usually 10-15 per author). One person should be appointed as having overall editorial control. The default would be the Head of the Year, but it would be more appropriate to delegate.
- b) Set deadline for question submission (3 months before exam).
- c) Organise review meeting. New questions are discussed, edited and approved as exam ready (~2 months before exam). This is best done as an away day. Year 6 organises their meeting at the Royal College of Physicians and we would recommend a similar arrangement. The University will fund this.
- d) Exam is constructed with ample time for standard setting (>6 weeks before the exam).

- e) Standard setting forms sent out, with deadline for return (~4 weeks before exam).
- f) Standard setting meeting conducted (~2 weeks before exam)

Exam paper construction

The aim is to produce a consistent, reliable paper. The exact length will be guided by previous experience and review of previous psychometric data but it is reasonable to aim for each paper to be 70-100 questions. These will be detailed for each year at the beginning of the year. An indicative blueprint should be produced outlining what areas are to be covered in the assessment. This will include both subject areas and skills/learning outcomes. An outline for this will be provided. The blueprint should be made available to students prior to the exam. The exam should cover all the important topics in that part of the course. There should be limited use of sequential linked questions (e.g. where one scenario is used to test more than one item of knowledge). The exams in years 1 and 2 will use SBAs. The exams in Years 4-6 will be a mixture of SBAs and VSAQs. The rough balance will be 90% SBA, 10% VSAQ, but this may be varied if thought to enhance the assessment. Each correct answer receive a single mark. There should be limited reuse of questions *already seen by that year*, but questions used in previous years assessments are essential in the exam design. Where previous year questions are used their performance should be assessed post-exam. Each section of the exam is signed off by the relevant module lead and finally approved by the editor.

Standard setting and post-exam analysis

Each exam should produce a summary of the method. A detailed policy is available (see standard setting appendix). The following is an outline of the key points:

- A modified Angoff approach should be used.
- The standard setters should come from question writers, tutors, lecturers and medical staff. In the clinical years, junior doctors should also be used where possible. The key is that standard setters are familiar with the course and student ability at the stage of the course being assessed. A chair of the standard setting group should be appointed (this is often the exam editor).
- The minimum number of standard setters is 4, but 6-12 is more ideal. A good range of opinion is essential.
- The year group should agree an anchor statement. An exemplar anchor statement would be “what proportion of borderline competent candidates would get the question correct?”.
- All standard setters review all questions independently and submit their standards. This will be a number between 0.2 and 1, using 0.05 integers. For most questions the standard will lie between 0.4 and 0.8.
- An average of the submitted standards is produced.
- A meeting should be held to discuss the average standard for each question. More discussion will be required for questions with divergent opinion. The final standard for each question is either achieved by a consensus opinion

or by individual standard setters revisiting their scores and a repeat average taken. The performance of previously used questions included in the exam should also be reviewed. This should use of classical test theory and item response theory data to calibrate the standard setting. This can help calibrate standards for other questions. Where a physical meeting is logistically difficult this process may be performed electronically.

- The average of the agreed standards is the pass score of the exam.
- All exams will be subject to a post-exam analysis. All poorly performing questions (e.g. negative discrimination or score less than chance) should be reviewed. Also all questions where the standard set is >10% greater than the difficulty of the question should be reviewed (i.e. the whole cohort has performed much worse than the standard). This will normally involve reviewing ~10% of questions. The post-hoc analysis will review both classical test theory and item response theory data to help ensure that the pass score identifies those who have not achieved the relevant knowledge competency. A pre-board panel should determine whether the question is removed, retained or standard modified.