

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

FFLM Guidelines on Standard Setting (Modified Angoff) for Examination Questions

Introduction

Any examination must be a fair and balanced assessment, that is, it assesses that which the candidate is required to know or demonstrate, based on the curriculum. The starting point of this process is 'blueprinting' which determines the content of the examination: what must be assessed, which modes of assessment are available, and lastly which are the most appropriate to use. Some parts of the curriculum could be assessed using any mode, whilst for example a clinical task is likely only to be adequately assessed by an objective structured clinical examination, (OSCE). Blueprinting ensures broad and representative sampling of the whole testable domain, increasing reliability.

In a postgraduate examination, *'the aim of examinations within this context is to separate the competent candidates from the not competent candidates, according to how they compare to a minimum standard.'*¹

The FFLM's examinations include the following modes of assessment:

- The single best answer question (SBA)
- The short answer question (SAQ)
- The Objective (Observed) Structured Clinical Examination (OSCE)
- The Objective (Observed) Structured Practical Examination (OSPE)

These must all be standard set, that is, undergo an assessment by experienced examiners who then assess each question on the basis of what **'the borderline candidate' would** know/do, which in turn determines how difficult or easy a question is, and, the overall pass mark. Therefore, with each paper the pass mark will vary, i.e. it is not a fixed pass mark e.g. always 50%.

Examiners, when writing or reviewing questions, will have ensured that the answers have been checked and are supported by appropriate references or evidence, e.g. current guidance, legislation etc.

Examiners are reminded of the importance of confidentiality and that details of questions must not be revealed or discussed in any other forum.

A. Single Best Answer (SBA) Questions

1. Before the meeting

- 1.1 The questions must have been written, adequately reviewed and be confirmed as suitable, before the standard setting process begins.
- 1.2 It is not appropriate to identify that a question is not clear, or to be correcting the wording, grammar, or the answers, during the standard setting process. Moreover, none of the questions should be asked in the negative. These should have been removed before the paper is deemed ready along with any question that inadvertently gives away the answer to another.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

- 1.3 A question which has been used before may not need to be standard set again; however, it may need to be if:
- It has been substantially re-written, for example because it was misunderstood in an earlier examination and performed badly
 - It has performed unduly badly or well in a previous examination (suggesting it was more difficult or easier, than had been thought)
- 1.4 The single best answer is exactly that, there should be a single best answer; some of the options (the distractors) may be 'nearly right', but there must be one answer, which in the circumstances/scenario/context, is the correct answer to the question posed.
- 1.5 The question paper (which is password-protected) is sent out by the Examinations Manager, along with a spreadsheet. The date by which this task must be completed will be included in the covering email. Examiners must have at least a week to undertake this task
- 1.6 Examiners must acknowledge receipt of the email, by return.
- 1.7 The examiner undertakes the assessment 'blind', that is not in collaboration with other examiners. The examiner must consider what might characterise a borderline performance or a **borderline candidate**. For example, such a candidate may have patchy knowledge: good in some areas, absent in others, or they have some superficial knowledge, but it lacks depth or detail, or they have limited ability to apply it.
- 1.8 For each question, the examiner must look at the question, carefully and decide what proportion (percentage) of 100 borderline candidates **would** identify the correct single best answer, i.e. they **would** get the question right.
- 1.9 It is essential the examiner considers what the candidate **would** do, not what they '**should**' do. This may appear counter-intuitive, but if '**should**' is used, the pass mark would become unreasonably high.
- 1.10 Scores must be entered as a percentage ending in 0 or 5, i.e. 5%,10%,15%, etc. It is unlikely that a score will be 0% (i.e. no borderline candidate would give the correct answer) or 100%, (i.e. all borderline candidates give the correct answer).
- 1.11 If the examiner thinks that 75% of borderline candidates would get the correct answer, then the figure 0.75 is entered onto the spreadsheet. If, however, the examiner thinks only 35% of borderline candidates would get the correct answer, then 0.35 is entered onto the spreadsheet.
- 1.12 If the paper being standard set is that for the DLM and the MFFLM Part 1, examiners must be aware that whilst the same questions are used, the MFFLM Part 1 is marked at a higher standard, than the DLM, so each question has to be standard set twice, at:
- A lower standard for the DLM
 - A higher standard for the MFFLM part 1
- 1.13 It may be useful for the examiner to make notes on why they have made their decision, to help inform any discussions at the meeting.
- 1.14 Once all the questions have been assessed in this way, the spreadsheet must be sent back to the Examinations Manager, by the date required.

2. At the Standard Setting (Modified Angoff) Meeting

- 2.1 The date of the meeting will be set beforehand, and a minimum of six (6) examiners must take part, but it is preferable to have more.
- 2.2 It is usually held as a teleconference meeting.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

- 2.3 The Examinations Manager will have prepared a spreadsheet of all the examiners 'scores', which will also highlight where the scores diverge **by 30% (0.30)** or more; this is based on advice from the FFLM's examination's advisor.
- 2.4 The meeting will be chaired, by the Chief Examiner, the Deputy Chief examiner, the Lead for SBAs or an experienced examiner.
- 2.5 In the course of the meeting, it is advisable to make notes about any question that has proved particularly contentious. This would demonstrate, e.g. to candidates or others that the process of standard setting involves careful discussion and that the performance of all questions is considered post results e.g. any question where on reflection the wording or options were ambiguous should be removed from the results.
- 2.6 Each question highlighted where the range of scores (that is the difference between the lowest and highest score) **is 30% or more** will be discussed. This discussion may be informed by the notes which an examiner may have made (see 1.13 above).
- 2.7 As a result of listening to the discussion an examiner may choose to alter their score, but there is no requirement to do so, therefore this difference of 30% or more, may remain. An examiner should not be asked to alter their score, to bring the difference down to less than 30%.
- 2.8 The pass mark of the examination is determined as a result of this process.

B. Short Answer Questions (SAQ)

1. Before the meeting

- 1.1 The questions must have been written, adequately reviewed and be confirmed as suitable, before the standard setting process begins.
- 1.2 They must have been written in such a way, answers to one part of a question should not be dependent on knowing the correct answer to another part.
- 1.3 It is not appropriate to identify that a question is not clear, or to be correcting the wording, grammar, or the answers, during the standard setting process. All these issues should have been resolved before the paper is deemed ready.
- 1.4 A question which has been used before may not need to be standard set again; however, it may need to be if, for instance:
 - It has been substantially re-written, for example because it was misunderstood in an earlier examination and performed badly
 - It has performed unduly badly or well in a previous examination (suggesting it was more difficult or easier, than had been thought)
- 1.5 The question paper (which is password-protected) is sent out by the Examinations Manager, along with a spreadsheet. The date by which this task must be completed will be included in the covering email. Examiners must have at least a week to undertake this task.
- 1.6 Examiners must acknowledge receipt of the email, by return.
- 1.7 The examiner undertakes the assessment 'blind', that is not in collaboration with other examiners. The examiner must consider what might characterise a borderline performance or a **borderline candidate**. For example, such a candidate may have patchy knowledge: good in some areas, absent in others, or they have some superficial knowledge, but it lacks depth or detail, or they have limited ability to apply it.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

1.8 For each question/subsection, the examiner must read the questions, carefully, as well as the outline of a 'model' answer, and must then decide what **mark** the candidate **would** get for that section of the question. N.B.

This is a different approach to that used in the SBAs and the OSCEs.

1.9 It is essential the examiner considers what the candidate **would** do, not what they '*should*' do. This may appear counter-intuitive, but if '**should**' is used, the pass mark would become unreasonably high.

1.10 The examiner then enters the **marks** into the spread sheet; i.e. what mark a borderline candidate **would** get. For example, an SAQ has 2 sections where the maximum marks are as follows:

a) 10

b) 6

The examiner decides that a borderline candidate **would** get:

a) 5.5 marks

b) 4 marks

Therefore, the examiner would enter 5.5 and 4 into the spreadsheet.

1.11 Scores, i.e. the **marks** a borderline candidate **would** obtain, must be entered as a whole number of marks, or to a decimal point of .5 (**not** as a percentage). For example:

- A section has a maximum of 15 marks, a borderline candidate would get 7.5 marks, so 7.5 is entered into the spreadsheet
- The next section has 7 marks, and the borderline candidate would get 4 marks, so 4 is entered into the spreadsheet
- And so on, for all sections of every question

1.12 It may be useful for the examiner to make notes on why they have made their decision, to help inform any discussions at the meeting.

1.13 Once all the questions have been assessed in this way, the spreadsheet must be sent back to the Examinations Manager, by the date required.

2. At the Standard Setting (Modified Angoff) Meeting

2.1 The date of the meeting will be set beforehand, and a minimum of six (6) examiners must take part, but it is preferable to have more.

2.2 It is usually held as a teleconference meeting.

2.3 The Examinations Manager will have prepared a spreadsheet of all the examiners 'scores', which will also highlight where the scores diverge **by 30% (0.30) or more**.

2.4 The meeting will be chaired, by the Chief Examiner, the Deputy Chief examiner, the Lead for SAQs, or an experienced examiner.

2.5 It is useful to have a discussion about what constitutes a borderline performance or a borderline candidate', e.g. patchy knowledge or lacking in depth or application.

2.6 In the course of the meeting, it is advisable to make notes about any question that has proved particularly contentious. This would demonstrate, e.g. to candidates or others that the process of standard setting involves careful discussion and that the performance of all questions is considered post results e.g. any question where, on reflection the wording was not clear or ambiguous, or the answers were very different from those expected or provided in the 'model' answer and should be removed from the results.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

- 2.7 Each question highlighted where the range of scores is 30% or more will be discussed. This discussion may be informed by the notes which an examiner may have made (see 1.12 above).
- 2.8 As a result of listening to the discussion an examiner may choose to alter their score, but there is no requirement to do so, therefore this difference of 30% or more, may remain. An examiner should not be asked to alter their score, to bring the difference down to less than 30%.
- 2.9 The pass mark of the examination is determined as a result of this process.

C. Objective Structured Clinical Examinations (OSCEs and OSPEs) and Weightings

1. Before the meeting

- 1.1 The stations must have been written, adequately reviewed, including any review which is deemed necessary, after the practice/rehearsal day, and be confirmed as suitable, (including achievable timings) before the standard setting process begins.
- 1.2 They must be written in such a way that the instructions to the candidate are clear and unambiguous. It is not appropriate to identify that a question is not clear, or to be correcting the wording, grammar, or the answers, during the standard setting process. All these issues should have been resolved before the station is deemed ready.
- 1.3 A station which has been used before may not need to be standard set again; however, it may need to be if:
 - It has been substantially re-written, for example because it was misunderstood in an earlier examination and performed badly
 - It has performed unduly badly or well in a previous examination (suggesting it was more difficult or easier, than had been thought) often based on examiner feedback
 - Practice has changed significantly or recently
- 1.4 The question paper with all the stations (which is password-protected) is sent out by the Examinations Manager, along with a spreadsheet. The date by which this task must be completed will be included in the covering email. Examiners must have at least a week to undertake this task.
- 1.5 Examiners must acknowledge receipt of the email, by return.
- 1.6 For the OSCEs and the OSPEs, the examiner has two tasks: **standard setting and weighting**
- 1.7 The examiner undertakes both these assessments 'blind', that is not in collaboration with other examiners.
- 1.8 The examiner must carefully read all parts of the station: the candidate's instruction, and in particular the task, the construct, the simulated patient (SP, in the OSCE)/simulated doctor (SD in the OSPE) information, and then consider what constitutes a borderline performance or the **borderline candidate**. For example, such a candidate may be disorganised in their approach, have patchy knowledge or limited experience, so they may be competent and capable in some areas, but not (or less so) in others. Or they show variable ability to communicate well, or demonstrate some practical skills, but they may lack the ability to complete the task.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

- 1.9 For each station, there will be an outline of what the candidate is expected to demonstrate, and the examiner must decide what proportion of borderline candidates **would** perform competently in demonstrating what is required in the task given in the candidate's instructions.
- 1.10 The skills to be demonstrated are covered by domains, and the content of each is considered individually, as different stations will have different 'demands' within each domain. Therefore, the score for a particular domain will vary from station to station. For example, a station where the SP is distressed, and so the initial approach is important, might be one where the borderline candidate would not perform competently; whereas, a station where the initial approach is straightforward, may be one where the borderline candidate may be more likely perform competently.
- 1.11 Ignore the 'A to E' rating, the task for the examiner is to determine what percentage of borderline candidates would be deemed competent for that particular domain.
- 1.12 It is essential the examiner considers what the candidate **would** do, not what they **should** do. This may appear counter-intuitive, but if '**should**' is used the pass mark would be unreasonably high.
- 1.13 Scores must be entered as a percentage ending in 0 or 5, i.e. 5%,10%,15%, etc.
- 1.14 It is unlikely that a score will be 0% (i.e. no borderline candidate would demonstrate what was required) or 100%, (i.e. all borderline candidates demonstrated what was required).
- 1.15 If the examiner thinks that 75% of borderline candidates would demonstrate competence on a particular domain, then the figure 0.75 is entered onto the spreadsheet. If, however, the examiner thinks only 35% of borderline candidates would demonstrate competence, then 0.35 is entered onto the spreadsheet.
- 1.16 In the OSCE and OSPE examination there is also the simulated patient/simulated doctor (i.e. the actor) score to Angoff.
- 1.17 If the OSCEs being standard set are GFM and SOM examinations, (LFFLM and the MFFLM Part 2), examiners must be aware that whilst the same stations may be used, the MFFLM Part 2 is marked at a higher standard, so each question has to be standard set twice, at:
- A lower standard for the LFFLM
 - A higher standard for the MFFLM part 2
- 1.18 For both OSCEs and OSPEs, it is also necessary to address the weighting of domains.
- 1.19 The **weightings** reflect the relative importance of each domain to the station as a whole, with reference to the construct. It is advisable to do these at the same time as the standard setting, whilst the station is 'fresh' in one's mind.
- 1.20 The weightings must add up to 100% and are entered into the spreadsheet and should be entered as the % e.g. if the weighting for a domain is 25% it is entered on the spreadsheet as 25, (not 0.25)
- 1.21 It may be useful for the examiner to make notes on why they have made their decision, both on the standard setting and the weighting, to help inform the discussions at the meeting.
- 1.22 Once all the questions have been assessed in this way, the spreadsheet must be sent back to the Examinations Manager, by the date required.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

2. At the Standard Setting (Modified Angoff) Meeting

- 2.1 The date of the meeting will be set beforehand, and a minimum of six (6) examiners must take part, but it is preferable to have more.
- 2.2 It is usually held as a teleconference meeting.
- 2.3 The examinations manager will have prepared a spreadsheet of all the examiners 'scores', which will also highlight where the scores diverge **by 30% or more** (0.30).
- 2.4 The meeting will be chaired, by the Chief Examiner, the Deputy Chief examiner, the Lead for OSCEs or OSPEs, or an experienced examiner.
- 2.5 Initially, it is helpful to discuss what constitutes a borderline performance or the **borderline candidate**, (see 1.8 above).
- 2.6 In the course of the meeting, it is advisable to make notes about any station that has proved particularly contentious. This would demonstrate, e.g. to candidates or others that the process of standard setting involves careful discussion and that the performance of all stations is considered post results e.g. where, on reflection the wording was not clear or ambiguous, or the answers were very different from those expected or provided in the 'model' answer and should be removed from the results.
- 2.7 Each station highlighted as having a range of scores which differ **by 30% (0.30) or more**, will be discussed, as advised by the Faculty's Educational Adviser. This discussion may be informed by the notes which an examiner may have made (see 1.21 above).
- 2.8 As a result of the discussion an examiner may choose to alter their score, but there is no requirement to do so, therefore this difference of 30% or more, may remain.
- 2.9 The pass mark of the examination is determined as a result of this process.
- 2.10 Weightings for the domains within each station must be discussed; this is essential as a consensus must be reached.
- 2.11 It is not acceptable to 'average' the weightings; for example, ten examiners have given a particular domain the following weightings: 10,15,10,15, 20, 25,15,10, 5 and 25. It might appear that the next step should be to average these: $150 \div 10 = 15$. This is not correct, as clearly there is a significant variation in how each examiner sees the importance of the domain. Thus, the examiners' discussion is very important.
- 2.12 The examiners' discussion must identify the most important domain(s), which will then have the greater weighting(s). This discussion is assisted by paying particular attention to the construct of the station. Once this is achieved, the examiners can turn their attention to the other (less important) domain(s). The weightings total must be 100%, but the starting point must be which is/are the most important domain/domain(s), not making the sum equal to 100.

D. After the Examination

The performance of every question is reviewed, along with any feedback from examiners and candidates. Where appropriate, action may be taken where problems have been identified with a question e.g. the model answer may be amended, a question might be removed and therefore no longer be part of the final mark.

FACULTY OF FORENSIC & LEGAL MEDICINE

of the Royal College of Physicians of London



Registered Charity No 1119599

Developed by Dr Bernadette Butler, former Academic Dean, based on the original guidance provided by:

Ms Angela Hall, Emeritus Reader in Clinical Communication,
Institute of Medical and Biomedical Education St. George's, University of London,
for which the FFLM is very grateful.

February 2020; review date: February 2023.

Reference

1. The Academy of Medical Royal Colleges (AoMRC)
Guidance for Standard Setting: A framework for high stakes postgraduate competency-based examinations (2015).
http://www.aomrc.org.uk/wp-content/uploads/2016/05/Standard_setting_framework_postgrad_exams_1015.pdf

Useful reading and Resources

1. Baird AS. The Part 1 MFSRH Examination: the new Single Best Answer paper
J Fam Plann Reprod Health Care 2016;42:222–224.
<https://srh.bmj.com/content/42/3/222>
2. Queens University Belfast Examiner Training: OSCE standard setting
https://www.med.qub.ac.uk/osce/background_Standard.html
3. Academy of Medical Educators
<https://www.medicaleducators.org>
4. Association of Medical Education in Europe (AMEE)
<https://amee.org/getattachment/AMEE-Initiatives/MedEdWorld/Multiple-Choice-leaflet-v3.pdf>