

## Introduction to Criterion Referencing

### Background

Until 2002, the MRCP(UK) Written Examination papers were norm-referenced – candidates were placed in rank order and a chosen top percentage were passed from each diet. In recent diets, around 40% of Part 1 candidates and 65% of Part 2 candidates have passed the written examinations and each time the pass mark may have been higher or lower than at the previous diet.

During the major review of the MRCP(UK) in 1997, it was suggested that this method did not take into account the candidates' actual performance in the exam. For example, if a group of candidates happened to be less good in a certain diet of the Part 1 Examination, the top 35% of them were still passed. Similarly if the quality of candidates was unusually high, 35% of them were passed regardless. It was obvious that the quality of candidates being passed on different diets could have varied.

A new method of assessing candidate performance on each diet of the Written Examination was needed. Setting a permanent pass *mark* (as opposed to the pass *rate* method above) was not acceptable, since the difficulty level of different exam papers might vary just as much as the quality of candidates, resulting in the same inconsistent quality of successful candidates.

The Colleges have decided that the best way to ensure that the right candidates pass the exam is to judge them against a measurable external standard set by experienced examiners. In this way, the pass mark may be set according to the difficulty level of the exam paper, and the performance of each candidate is compared to this standard instead of that of the other candidates. The method chosen for setting the standard of the MRCP(UK) is called the Angoff method of Criterion Referencing.

### Angoff's method

The purpose of the Angoff method is to use the knowledge of a group of experts to set a reasonable pass mark for each individual examination paper. There are a few key concepts that must be held in each examiner's mind as they are making judgements on the difficulty of the questions:

The 'just-passing' candidate – This is the group of candidates that will be most affected by the chosen pass mark. Those who do very well will always pass, and those who do very poorly will always fail. But small changes in the pass mark will affect large numbers of 'borderline' candidates.

The curriculum and the aims and objectives of the MRCP(UK) – It is important for the examiners to know what is expected of the candidates writing each exam. The aims and objectives as they appear in the regulations remind examiners of **why** these candidates are sitting the exam. The curriculum lays out the specific areas of knowledge with which the candidates are expected to be familiar.

Level of knowledge expected from an MRCP(UK) Candidate – Examiners must consider the knowledge expected from a doctor at the stage of training of the typical MRCP(UK) candidate.

With the above concepts in mind, examiners can then start the process of Criterion-Referencing.

## Methodology

The Part 1 and Part 2 Standard Setting Groups (SSG) are not set up in the same way, although their methodology is very similar. The Part 1 group consists of approximately seven examiners, who each analyse every question. The Part 2 group consists of approximately 14 members who divide into 2 smaller groups, and each smaller group considers a selection of questions rather than the whole paper. This is done for practical reasons to ensure the examiners can get through the required questions in the time available for the meeting.

The examination papers to be considered are circulated in advance of the SSG meeting. Each examiner must consider their allocated questions (all questions in the case of Part 1) and assign a percentage score to each. The assigned score represents the percentage of **just-passing candidates** that the examiner feels should get the question correct. In coming to this decision, examiners must keep in mind the above **key concepts**, which allow them to decide if a candidate should or should not be able to answer the question. All of the examiners' assigned questions must be judged this way, independently, prior to the meeting of the group. Examiners are encouraged to score questions, in the first instance, under test conditions (i.e. no answers, no textbooks). They may then go back through the paper with the answers and any other necessary materials and adjust scores as necessary if they did not get the correct answer. These marks are then recorded as 'pre-discussion' judgements.

At the meeting of the SSG, examiners go through the questions one by one and discuss them with the group. The examiners with the highest and lowest judgements may be asked to explain their reasons for their score. After discussion, each examiner then re-scores the question (again, independently) and these marks are recorded as 'post-discussion' judgements.

### Optional step – Modified Angoff

If the question being judged has been used before, there may be statistical information (called 'normative' data, or just norm data) available on its previous performance. Examiners can look back at how the **just-passing candidates** performed on the question for its previous use. The examiners then re-score the question based on this norm data. The result of this optional step, if used, is called the 'Modified Angoff Method' of criterion referencing.

It is important to note that if the question has been changed at all since the previous use, norm data cannot be used. Groups of candidates may react entirely differently to a question that has been changed even slightly, and there is no way to predict this. The norm data is therefore invalid.

## Results

After the SSG meeting the final set of numbers (whether Angoff or modified Angoff), as agreed by the examiners and the group chair, are used to calculate the pass mark for the examination. There are several methods that can be used to set the final pass mark. A few options are set out below.

### Mean

The simplest method begins with taking the mean for each examiner of all final judgements for all questions. This will give a value, expressed as a percentage, for each examiner. The overall mean of these values can then be used as the pass mark for the examination.

### Trimmed Mean

If there is concern about extreme judgements from 'rogue' examiners, the lowest and highest examiner scores can be removed before the mean is taken. This results in a 'trimmed mean.' Again, the overall mean of these trimmed means can then be used as the pass mark for the examination.

### Hofstee Method

The Hofstee method is currently used to set the pass mark for the Part 2 Written Examinations.

After equating was introduced for MRCP(UK) Part1 examination in 2008 September diet, the Part 1 standard setting group will meet once a year to monitor candidate performance and decide if the pass score will stay the same or needs to be adjusted.

### **Example - Criterion Referencing Judgements**

Below is an example of the spreadsheet used to enter the examiners' judgements. The blue section would be filled in prior to the meeting. The pink section would be filled in after group discussion. The green section would be filled in only if there is norm data (Modified Angoff Method)

Paper	Question	Pre-discussion judgements				Post-discussion judgements				Post norm-data judgements			
		Ex1	Ex2	Ex3	Ex4	Ex1	Ex2	Ex3	Ex4	Ex1	Ex2	Ex3	Ex4
(Ex1 = examiner 1, ..etc)													
1	1	75	60	70	55	70	65	70	55	75	70	70	65
1	2	50	45	45	50	50	45	45	50	40	40	45	50
1	3	70	70	60	70	75	70	70	75	80	80	70	75
1	4	85	75	80	60	75	75	80	70	This can be left empty if there is no norm data			

May 2009

MRCP(UK) Central Office