

## Standard setting

The Bar Standards Board sets the education and training requirements to ensure that newly qualified barristers have the knowledge, skills and understanding they need to provide quality advocacy to the public. The aim of Bar Professional Training Course (BPTC) and Bar Transfer Test (BTT) examinations is to assess competence and therefore to separate the competent candidates from the not yet competent candidates. Centralised examinations in three key areas: Civil Litigation, Criminal Litigation and Professional Ethics ensure consistency across course providers; standard setting those examinations helps us to ensure consistency of outcomes, assuring the public about the minimum standard of competence that is expected of those who are able to progress to the next stage of training – pupillage.

### *Why the Centralised Examination Board (CEB) undertakes standard setting*

There are challenges in ensuring that examination papers in different sittings are of the same level of difficulty, therefore a fixed pass mark which does not change between sittings is not appropriate. To ensure parity between cohorts, and therefore fairness to candidates, a standard setting exercise is undertaken in respect of each of the centrally assessed BPTC/BTT subjects<sup>1</sup>.

### *The standard setting methods used by the CEB*

The Angoff method is used as the standard setting process for the Civil Litigation and Criminal Litigation examinations. Standard setters ('judges') estimate how a group of *borderline* candidates would perform on each question in the examination. This method is one of the most widely used standard setting methods; it is used to determine the pass standard for all postgraduate medical multiple choice examinations (including those at the Royal Colleges of Anaesthetists, General Practitioners, Physicians, Paediatrics, Radiologists, and Emergency Medicine) as it has a vast body of research supporting its use.

The Contrasting Groups method is used as the standard setting process for the Professional Ethics examination. The pass standard is determined by using the scores of candidates who were judged as belonging to one of three groups (fail, borderline and pass) in relation to the perceived minimal level of competence required for a pass<sup>2</sup>. This method is also commonly used; it is used to determine the pass standard for postgraduate medical examinations including those at the Royal Colleges of General Practitioners, Physicians, and Psychiatrists.

### *The rationale for the standard setting methods chosen by the CEB*

The Angoff method of standard setting is used for the Civil Litigation and Criminal Litigation examinations because it is more suited to multiple-choice question (MCQ) examinations. By contrast, the Contrasting Groups method is more suited to short answer question (SAQ) assessments such as the Professional Ethics assessment. It should be noted that in neither method are test-takers compared to their peers but

---

<sup>1</sup> BPTC and BTT candidates sit the same centrally assessed examination papers on the same day. Therefore the pass standard is the same for both cohorts of candidates.

<sup>2</sup> This is deemed to be 'pupillage ready' (irrespective of the number of pupillages available).

## Standard setting

only to the pass standard. As a result, neither method results in quotas of passing and failing candidates (or in quotas of candidates allocated to a particular grading).

### *Explanation of the methods used by the CEB*

#### *Civil Litigation and Criminal Litigation*

Standard setting for the Litigation subjects takes place before the examinations are taken because the process is based on the level of difficulty of questions, rather than on candidate performance.

The Angoff method computes the passing score from an estimate of the probability of a hypothetical borderline candidate answering each question correctly. After a discussion and agreement of the characteristics of a borderline candidate, each judge (a subject matter expert) makes an independent assessment of the probability that a borderline candidate will answer each question correctly. The judges' assessments for each question are summed to give the probable total score of the borderline candidate. The median value across all judges is used to identify the recommended pass standard for the assessment as a whole.

The following sequence of steps is applied:

1. The judges are selected and a meeting is held.
2. The Chief Examiner pre-selects three questions which he deems to be of varying levels of difficulty (relatively hard, relatively easy and another which has a 50:50 probability of a borderline candidate answering correctly). The judges are not told the Chief Examiner's judgements on the questions, nor his reasoning.
3. The judges make preliminary judgements on these three questions. These are put into a spreadsheet visible to all judges.
4. The judges deemed to be outliers are asked to justify their judgements, and are given the opportunity to change their judgements. There is a group discussion about the perceived difficulty of each question and why it is thought to be difficult or easy.
5. The judges then make preliminary judgements on the first ten questions on the examination paper. These are put into a spreadsheet visible to all judges.
6. A brief discussion is conducted on each of these questions.
  - a. Each judge's choice of probability for each question is shown. If the numbers are similar, discussion moves on to the next question.
  - b. If the numbers are not similar, a judge who chose one of the highest numbers is asked to explain the reasons for choosing a high probability of borderline passing candidates getting the question right. Then a judge who chose one of the lowest numbers is asked to explain the reasons why it was thought to be a more difficult question. There is then a wider discussion regarding the group's judgements.
  - c. Having listened to the discussion, the judges are then told that they can change their judgements if they want. It is reiterated that the judgements are to describe the performance of *borderline* candidates.
  - d. After discussing the first ten questions, the judges make preliminary judgements on the next ten.

## Standard setting

- e. The next ten questions are discussed as per step (b), and judges are given the chance to change their judgements if they so desire.
- f. After discussing the questions 11 to 20, the judges make preliminary judgements on questions 21 to 30.
- g. These processes are repeated for questions 31 to 40, and the judges are asked to complete their remaining judgements after the meeting concludes.
- h. All judgements are sent to the BSB where they are collated.
- i. The psychometrician reviews the judgements in respect of all questions; any judgements that are found to have an unacceptable level of deviation are referred back to the judges in question to justify or amend their judgements.
- j. The 75 judgements of each judge are added up in order to arrive at their estimate of the borderline candidate's expected score for the whole examination.
- k. The scores for all judges are combined to give the median. This consensus judgement of the score that a borderline candidate would be expected to get becomes the recommended pass standard.

A translation process is required to express candidates' results as a percentage of the marks available. Thus, if the standard setting process identifies a score of 43/75 as representing the pass standard, any candidate achieving 43/75 will be reported as achieving the pass mark of 60%. All other marks above and below this figure have an appropriate factor applied to them so that they too can be reported as percentages. This is done because the other components of the BPTC and BTT which are set and assessed by Providers, rather than the CEB, have a pass mark of 60% so it is easier for Providers to combine these with CEB marks reported with a pass mark of 60% to give candidates' overall marks for the BPTC and BTT.

### *Professional Ethics*

Standard setting for Professional Ethics takes place after the examination is marked. An appropriate sample of scripts is selected by the psychometrician. The standard setters review the answers to each SAQ sub-part on each sample script to assess whether to allocate the answer to one of three categories. These are coded as:

- 1 for Fail
- 2 for Borderline
- 3 for Pass.

In carrying out this exercise the standard setters are not aware of the marks awarded to any of the questions. A linear regression methodology<sup>3</sup> is then used to predict the numerical score which, on the evidence of the sampled scripts, best represents the borderline score (or pass standard) for each SAQ sub-part.

---

<sup>3</sup> Regression analysis is a statistical technique that attempts to explore and model the relationship between two or more variables – in this case the marks awarded to a candidate in respect of their SAQ answer on the one hand, and the classification of the answer by the standard setters on the other.

## Standard setting

### *An example*

There are 10 marks available for each SAQ, giving a total of 60 marks for the examination overall.

The psychometrician derives the contribution to the pass standard for each sub-part by using the numerical mark awarded by the marker, and the judgement code of the standard setter (representing Fail, Borderline or Pass). He does this by applying a linear regression model to predict the numerical score which, on the evidence of the sampled scripts, best represents the borderline. This process is repeated for each sub-part. The value derived for each sub-part is applied to all candidates.

This methodology might identify the percentage of candidates meeting the standard for each sub-part as being:

SAQ 1A: 62%	SAQ 3A: 61%	SAQ 5A: 43%
SAQ 1B: 63%	SAQ 3B: 32%	SAQ 5B: 68%
SAQ 2A: 29%	SAQ 4A: 51%	SAQ 6A: 60%
SAQ 2B: 44%	SAQ 4B: 53%	SAQ 6B: 75%

The percentage meeting the standard for each sub-part is averaged to give the statistically equivalent mark for the examination overall. In this example, the average is 53.5%.

It should be noted that, as marks are allocated as whole or half marks, it may not be possible to achieve a pass rate of exactly 53.5%. It is necessary to identify the score at which the pass rate is closest to (in this example) 53.5%.

As with the Litigation subjects, a translation process is required to express candidates' results as a percentage of the marks available. Thus, if the standard setting process identifies a score of 24/60 as representing the pass standard, any candidate achieving 24/60 will be reported as achieving the pass mark of 60%. All other marks above and below this figure will have an appropriate factor applied to them so that they too can be reported as percentages.

### *How we select and train standard setters who are best placed to determine the minimum level of competence for a pass*

The panel of standard setters is required to conceptualise the minimum level of performance required for a pass in the examination. In order to do so, they must be knowledgeable about the candidate population, the content of the examination (i.e. they must have legal expertise and also be familiar with the BPTC syllabus and the Professional Statement for Barristers), and importantly, the reference group (e.g. the borderline candidate) which forms the basis of the performance standard.

As such, the panel of standard setters for each of the examinations includes examiners, BPTC and BTT tutors, and practising barristers (including pupil supervisors).

## Standard setting

For the Angoff method (used for the Litigation subjects), careful consideration is given to the number of standard setters required; our panels have a minimum of ten standard setters per subject as this is considered best practice for this method. For the Contrasting Group method (used for Ethics), it is not necessary to have such large panels.

Standard setters receive training, so that they can provide their judgements in an informed manner. We are able to use real data from previous examinations in training exercises; the independent psychometrician can isolate a group of 'just passing' candidates and compare their performance on questions with the judges' ratings.

### *Making the standard as reproducible as possible*

Variability between standard setters is inevitable, but, in addition to training the standard setters, the way in which it is implemented helps to ensure that, if the standard setting process for the examination was repeated under the same conditions, the same results would be achieved. The minimum level required for a pass is clearly described and discussed by the panel at the start of the standard setting process.

Steps are taken to minimise the statistical impact of outliers. For the Litigation subjects we do this by ensuring that the median of the judges' ratings is used, rather than an average. For all three subjects, the impact is also minimised by group discussion, and allowing judges to alter their judgements after the discussion.

### *Carrying out a sense check on the outcomes*

As part of our quality assurance processes, we carry out a sense check on the standard setting outcomes at both Subject Boards and the Final Examination Board.

As the borderline candidate is adequately described and discussed before standard setting, (and training exercises use real data from past examinations) and due process is followed, there is no need to revisit the standard once it has been set. The ultimate responsibility for setting the pass standard resides with the Chair and Final Examination Board, and the pass standards are provisional until agreed by the Final Examination Board. This is why the pass standards are not made public before examinations take place.