AFRM Examinations Glossary

- **Borderline Regression Technique:** Borderline regression is an absolute, examinee-centred standard-setting method widely used to standard-set Objective Structured Clinical Examinations (OSCE). Candidates are awarded a "global score" for a station in a circuit based on the examiner's judgement of their ability.

  The actual mark that all candidates scored for a station is plotted on a graph against the global score they were awarded for that station, and a best-fit line (line of regression) is drawn. The point at which the line intersects with a vertical axis between the "borderline fail" and "borderline pass" rating indicates the performance standard on the station's checklist.

- **Circuit**: For the administration of an OSCE [such as the Fellowship Clinical Examination (FCE) and Entry Phase Examination (EPE)] a circuit is defined as a set of stations for the examination. An examination may have more than one circuit. As an example, an examination may have two or more concurrent circuits running during both the morning and afternoon at site 'A' and a single circuit at site 'B' during the morning and afternoon – so whilst there would be a total of six groups of candidates for this exam, there would only be three circuits (2 at site 'A' and 1 at site 'B').

- **Global Assessment**: The Global Assessment Ratings of 'Excellent', 'Clear Pass', 'Borderline Pass', 'Borderline Fail' or 'Clear Fail' reflect the candidate's overall performance in the station and are provided by consensus between Examiners on the station.

- **Item Analysis**: Once the scores have been determined, a process of item analysis is undertaken to monitor standard psychometric indices: Difficulty (D), the proportion of candidates who answer the item correctly; Discrimination (Di); correlation of candidates' performance on this question and overall score. Items where Di is negative (< 0) and/or those with extreme difficulty statistics (i.e. over 0.95 or under guess; 0.20 for 5 item questions and 0.25 for 4 item questions are reviewed by examination leads for accuracy and fairness. Modified Essay Question (MEQ) items or OSCE stations with mean scores below the pass mark are also reviewed by subject matter experts in collaboration with the Senior Lead, Assessment Quality and Development and Quality Assurance.

- **Modified Angoff standard-setting approach**: A Modified Angoff standard-setting approach is used to determine the minimum performance standard required on Multiple Choice Question (MCQ) and MEQ examinations and written components (static stations) included in clinical examinations.

  - For each examination a panel of AFRM Fellows who are subject matter experts in education and training across a diverse range of rehabilitation training at the level of the examination are convened for the standard-setting process. The minimum panel size for robust standards is five members. Members of the FAC and item writing group are permitted to join the panel but not more than one from each group. Standard-setting panel members rate and submit their estimate of the score the minimally competent candidate would achieve for each item in the examination. (Angoff score).

  - For MCQ examinations, examiners are asked to estimate a percentage of minimally competent candidates who would have answered each question correctly (i.e. percent chance the minimally competent candidate would answer the question correctly). These ratings are averaged per question, and the performance standard for the whole paper is determined by calculating the sum of all Angoff scores for each question in the paper.

Last Revised: February 2025

- For MEQ questions, the examiners are asked to determine a score the minimally competent candidate will achieve per sub-question. The final Modified Angoff score for each sub-question is the average of these scores. The pass mark for each question will be determined as the sum of the Modified Angoff scores per sub-question. The performance standard at the depth level is the aggregate Angoff score of all eight questions

- Before review and finalisation, interrater reliability indices are calculated for each question. Any question with an intra-class correlation coefficient less than 0.7 at the lower confidence interval boundary is discussed at the standard-setting meeting.

- **Standard error of measurement (SEM)** represents the estimation of measurement error in a candidate's observed score from an examination. All observed candidate scores include 'random error'. According to Classical Test Theory (CTT), a candidate's observed score is the sum of their 'true score', representing their true ability on the assessed domain and measurement error. I.e.

    Observed score = true score + measurement error.

The impact of measurement error in examinations is that when candidate scores fall within one standard error of the minimum performance standard score, there is uncertainty as to whether their true score lies above or below the performance standard. Thus, they may receive a false positive (designated a pass result when the true score is below standard) or a false negative (designated a fail result when the true score is above standard) outcome.

To mitigate false outcomes, it is accepted practice to create a confidence interval around the performance standard cut score by calculating the standard error of measurement for each examination and applying this as a confidence interval around the standard. Decision makers can have greater certainty that scores outside the confidence interval represent a true difference in performance against the expected standard. The overall pass mark for a successful outcome on the examination is set at the lower or the upper threshold of the examination, depending on the purpose and the stakes of the decision made at the point of assessment. E.g:

- To mitigate the risk of **false negative** outcomes (assigning a fail outcome to a true pass result) it is advisable to set the overall pass mark at the performance standard **minus** one SEM.

- To mitigate the risk of **false positive** outcomes (assigning a pass outcome to a true fail result) it is advisable to set the overall pass mark at the performance standard plus one SEM.

The Standard Error of Measurement is calculated by applying the following formula to all raw scores on the examination:

- SEM = Standard deviation of total scores * $\sqrt{1-\alpha}$

- Where α is Cronbach's alpha reliability coefficient calculated from all raw scores on the examination.

- A confidence interval of +/-1 SEM can be placed around the cut score for the examination.