

RACGP Assessment handbook for candidates

Assessment performance reporting



THE ROYAL AUSTRALIAN
COLLEGE OF
GENERAL PRACTITIONERS

There are two main areas that are traditionally featured in the reporting that follows the conduct of any assessment process: candidate performance (reported to the candidates who undertook the assessment) and assessment performance (reported to the parties responsible for the administration of the assessment).

Certain terms are frequently used in the assessment literature and need to be understood and addressed in any assessment process.

Validity

Validity refers to the accuracy with which a test measures what it purports to measure.

Content validity refers to the adequacy of sampling of the specified universe of the discipline. For example, to test competence for general practice, content validity would be achieved only if the comprehensive content of general practice was appropriately sampled, avoiding over emphasis on some areas and under emphasis on others. This is achieved by using a matrix/blueprint when developing an assessment process.

Reliability

Reliability refers to the consistency with which a set of test scores measures whatever it is meant to measure. In other words, reliability is the extent to which results obtained in an assessment process are consistently reproducible.

Three factors influence the reliability of a set of scores:

- first, the test will be unreliable if items are too hard, too easy or ambiguous, as such items fail to discriminate between the good and the average candidates
- second, individuals vary from hour to hour and day to day in their energy, alertness, emotional state and other personal characteristics. If these personal variables affect the candidate's performance appreciably, reliability will suffer
- third, the consistency and objectivity of the examiners will affect reliability. The extent to which the scores examiners assign depends upon their feelings and impressions of the moment, rather than on the consistent application of predetermined criteria and standards, will determine the reliability or otherwise of the scores given by them.

The correlation between the marks given by two or more examiners rating the same performance is called the interrater reliability. The higher the consistency of the examiners' marking, the higher the interrater reliability will be. Interrater reliability is improved by removing subjectivity as much as possible, by using structured objective rating forms and by training examiners to be proficient in observation, recording, questioning and rating. In oral examinations, structured questions improve reliability.

Reliability will improve by increasing the number of measurements made of an individual candidate and by increasing the number of examiners making a judgment on each candidate's performance. One of the reasons the college assessment processes have multiple tests and multiple examiners is to improve reliability.

Reliability is also increased by removing poorly performing items from assessment processes after analysis of candidate performance and before results are ratified.

Relationship between validity and reliability

In tests of complex skills and attributes such as those tested in the college's assessment processes, there needs to be a compromise between reliability and validity. As one pushes towards higher and higher reliability, validity invariably suffers. It is unwise to put too much weight on a highly reliable test if it fails to measure important factors relating to the doctor's professional knowledge, skills and attitudes. There is no point in having a highly valid test which produces unpredictable results, where some examiners rate the candidate's performance very poorly and others rate the same performance very highly.

The blind pursuit of reliability or validity at the expense of the other is poor practice. While everything should be done to achieve the highest levels of both reliability and validity, inevitably a balance between the two is needed. The reliability of the college assessment processes compares well with internationally acceptable levels.

Standard setting

A standard is a conceptual boundary between acceptable and unacceptable performance. For the college examination, the standard of performance required of candidates is demonstration of competence for unsupervised general practice in Australia.

Standard setting is the process by which a standard is translated into a passing score, intended to divide a group of candidates into those who are at or above the standard, and those whose performance is below the standard.

The aim of the standard setting process is to accurately and fairly translate the concept of the minimum competent performance level into an actual figure, or a 'passing score'.

The selection of standard setting methods used by the college for the current assessment processes was made with consideration of both theoretical and practical implications. Problems and issues with traditional methods of standard setting are in many ways addressed by the techniques now used.

The procedures used to set the 'cutting scores' for the applied knowledge test and key feature problems involve a group of expert examiners making judgments about hypothetical borderline candidates' performance in each question. The actual performance of candidates in the same questions can then be considered before determining final scores.

This standard setting process has an additional benefit to candidates. The quality of every question used in the examination is scrutinised by the group of expert examiners for face validity and statistical reliability and quality. Any questions considered to be of less than excellent quality are debated at length and often omitted.

The 'borderline group method' is used to set the cutting score for the objective structured clinical examination. This method uses each examiner at each clinical station throughout Australia. At each of the 14 clinical stations, each candidate is judged on a number of different performance domains, as well as on their overall performance. Standard setting is achieved through statistical analysis of the performance domain scores against the overall performance scores.

Although the overall pass mark may vary from assessment process to assessment process, the standards used in determining those pass marks remain constant. Regardless of the procedure used, setting standards requires judgment. The processes employed by the college in determining cutting scores are less arbitrary than simply choosing a pass mark. They involve both judgments by examiners and analysis of actual candidate performance in the assessment tasks. The result is a process that is both fair and accurate.

Item banking

This term is used when referring to the storage and retrieval of test information. Each assessment segment is analysed in detail and the performance of candidates is reviewed on a state by state, case by case and question by question basis. This enables a global analysis of the assessment process and the contribution individual segments make to the process of discriminating between pass and fail candidates. In this way, weak questions and cases can be strengthened, and unreliable ones discarded.

Thus, the assessment process is subject to continual review and subsequent fine tuning to reduce or eliminate weaknesses.

Candidate performance reporting

The Board of Censors will review the scores achieved by all candidates at the conclusion of the administration of each segment of the examinations. This information is deidentified so that the Board of Censors are unaware of a candidate's identity until after results are ratified. The administration of the assessment segments, the candidates' scores and the cutting score for each segment are reviewed by the Board of Censors to affirm that due process has been accorded in the calculation of the cutting score. Ratification, undertaken by the Board of Censors on behalf of RACGP Council, involves the confirmation and approval of the segment cutting scores, and the overall result for each candidate as determined by the Board of Censors.

The examination results, as ratified by the Board of Censors on behalf of RACGP Council, shall be final and will not be subject to review except:

- at the instigation of the candidate in conformity with the procedure, and subject to the rules relating to the Board of Censors' decisions review policy.
- pursuant to a resolution of the RACGP Council, duly passed, directing the censor in chief to review examination results for any purpose the council deems fit in its sole unfettered discretion
- the censor in chief, having conducted a review of examination results as ratified, may make recommendations to the RACGP Council and any resolution of the council as to such recommendation shall be final.

Completion of the examination cycle

There are three segments of the examination: the applied knowledge test, the key feature problems and the clinical examination. Candidates are required to pass each segment.

The maximum duration of a candidate's examination cycle is three years, during which time each segment of the examinations must be completed successfully. A pass mark in any one segment of the examination may be carried over for a period of three years from the first sitting of the AKT. A failed segment may be repeated until a pass mark is achieved in that segment. If a candidate is unable to successfully complete each of the segments within this three year period, a result of 'fail' will be recorded for that segment of the examination.

The Board of Censors has the right to request remediation of a candidate and evidence of completion of remediation if a candidate fails an assessment segment.

If, at the end of the three year examination cycle, not all three segments have been

completed successfully, credit can longer be carried over. A new examination cycle will need to be commenced and each of the three segments will need to be undertaken and successfully completed regardless of the result achieved in a previous examination cycle.

Results statements

A list of successful candidates for each segment, identified by RACGP number, will be posted on the RACGP website. The RACGP reserves the right to alter the dates of the release of results as required.

Any candidate may be required to undergo further examination should the censor in chief deem that insufficient data is available to determine whether a candidate has reached the necessary standard.

All candidates will receive written notification of their results. The written results will be mailed to candidates within 7 days of the date of publication. It is the responsibility of all candidates to ensure that the Assessment Department has the correct address for the mailing of examination results.

Examination results will not be provided by telephone or faxed.

Feedback to candidates

An important element of any educational process such as the college assessment processes is feedback. The results that are sent to candidates will include national data on mean scores and pass rates, allowing candidates to determine their level of performance in relation to their peers.

Feedback may be requested by the candidate from the censor in each state. This feedback process will commence 14 working days after the publication of results. College censors give feedback on areas in which candidates performed well, as well as the areas of concern. Candidates who fail assessment processes are advised to avail themselves of this service. Feedback may be delivered as a group activity with other unsuccessful candidates.

Feedback is an essential part of the assessment process since it encourages candidates to identify their strengths and areas for improvement in practice and to obtain further training in those areas. A censor may also require a remediation plan be undertaken by a failed candidate before further assessment segments or cycles are undertaken. Thus the examination has an educational value over and above its value as an assessment process.

Please note that candidates who request a procedural error clarification of assessment results or who use the censor in chief's appeal committee will not be eligible for feedback until these processes are complete.

Request for a reconsideration

Candidates may request procedural error clarification if they believe there has been a procedural error during the examination.

The closing date to apply for a reconsideration is 10 working days after the publication of results. If a candidate believes that there has been a procedural error, they should inform the supervisor/invigilator or assessment panel chair on the day of the alleged error and complete an incident report at the same time. This will benefit the candidate, as it will be used as substantiating evidence.

The process for reconsideration does not include the examination paper being re-marked.

The process considers the overall result of the candidate, the circumstances of the actual incident and the possible impact of the incident on the performance of the candidate.

Reconsideration will **not** be conducted on the grounds of an alleged error in content or marking of the examination, nor on the basis of personal expectation of performance.

For example: during a clinical case you feel that the explanation regarding your treatment of a particular problem is correct, but the examiner requests other management.

A request for procedural clarification of results, or use of the censor in chief's appeal committee, based on the presumption that this case has been marked incorrectly will not be considered.

Procedural errors

Procedural errors include:

- incorrect information or instructions being given to a candidate during an assessment process or instructions that are not correctly available to a candidate
- faults in the timing of an assessment process
- disruptions to the running of the clinical cases (OSCE)
- required materials or equipment not being available.

A procedural error has not occurred where a candidate receives the correct information but misinterprets or misunderstands that information, eg. where a candidate is instructed to go to Room 4 but actually goes to Room 5. The candidate was given the correct information but did not follow the instructions correctly.

Incidents that would not be regarded as procedural errors would include:

- candidates being stopped during the written examination and told that they are no longer required to answer a particular question, or that the label number on a picture is incorrect. This is an unavoidable part of the assessment process and time for these incidents has already been provided for
- the objective structured clinical examination not starting at the advertised time or being delayed. This is not a procedural error if the correct time is allowed for once the clinical exam is commenced.

Reconsideration outcomes

In the event of a procedural error the following outcomes are the most likely.

Written (AKT and KFP) examination

- The candidate may be able to resit a future AKT or KFP examination
- The candidate may be allowed to sit a future AKT or KFP examination without payment of the normal enrolment fee
- Other outcome as deemed appropriate by the censor in chief or his/her delegate.

Objective structured clinical examination

A candidate's mark for a particular case may be amended if a procedural error occurs during that case in the clinical segment of the examination. Though alleged procedural errors may be unsettling, it is understood that such events are sometimes unavoidable. Disruptions often occur in real life situations and they must be dealt with in that context. Any amendment to an assigned mark will only apply to the particular case during which the error occurred; no consequential effect on following cases will be considered.

Board of censors decision reviews

The college may in its absolute and unfettered discretion decide whether to upgrade to a pass or to affirm a fail result on any application made by a candidate for reconsideration or the censor in chief's decision reviews committee. The decision in relation to any assessment reconsideration or censor in chief's decision review committee shall be final. In reaching any determination as to an application for special consideration or procedural error clarification or use of the censor in chief's committee of marking, the college shall not be bound by the rules of natural justice.

Policies in relation to special consideration, the Board of Censors' decision reviews are provided on the RACGP website at www.racgp.org.au/assessment/policy.