

AKT Standard Setting and Maintenance FAQs

1. How is the standard set?

The standard required to pass the AKT is set using the internationally recognised modified-Angoff methodology. This is the most widely used approach to standard-setting for high stakes multiple-choice tests. A wide range of subject matter experts take part in the judging including GP Registrars who have passed AKT, newly qualified GPs, GP Trainers, MRCGP examiners, Training Programme Directors, Deanery representatives, the BMA, as well as Lay Observers.

After detailed instructions and examples of scoring, each participant individually judges all the questions in an AKT and scores the likelihood of the “just passing candidate” getting each question correct.

The scores are collated and a mean score calculated with one standard error of measurement added.

2. How is the standard maintained from one exam to the next?

To derive an equivalent pass mark for subsequent AKTs, the standard is maintained using ‘item response theory’ (IRT). IRT is acknowledged as the best method for statistical test equating/standards maintenance. IRT permits the use of ‘concurrent’ equating, which allows all historic diets’ data to be utilised to equate the pass mark.

IRT is advantageous in that it establishes a definitive difficulty value per item that is not dependent on the performance of the cohort. In IRT, item difficulty is based on not only how many candidates get it right, but **how able those candidates were** (whether they answered many questions right, and whether those questions were the hardest ones). Similarly, a candidate’s ability is based not only on how many items a candidate gets right, but **how hard those items were**.

Another key feature of IRT is that item difficulty and candidate ability are placed on the same scale, which is consistent from diet to diet as a result of us equating each new diet onto the pre-existing scale. This amounts to using previously used items’ known difficulty to ‘calibrate’ or ‘anchor’ new items onto the IRT scale. Having done this, we can then work out IRT ability values for each candidate in the diet.

There is a defined point on the IRT scale that reflects the passing standard (determined by the Angoff method above). We use regression analysis to find the mark achieved by learners whose ability value is at this standard; this is the pass mark for the diet based on an IRT approach.

This approach is compiled by an independent expert psychometric team. IRT is an internationally validated method used across a range of high stakes national and professional exams.

3. How often do you repeat your standard-setting?

The Angoff procedure is repeated *at least* once every three years. It is undertaken

more frequently if test circumstances change. For example, when a calculator was made available to all candidates for the first time.

The next standard setting meeting is scheduled for November 2025.

To date, the Angoff procedure has taken place in:

- November 2024
- November 2021
- October 2018
- October 2016
- October 2015
- October 2014
- January 2013
- October 2010
- October 2009
- October 2007

4. Why does the pass mark vary?

The exam team aim to construct all AKTs to be of as similar a standard to each other as is possible to predict - for example, by having the same proportions of questions on any given topic and by looking at the difficulty of questions.

However, no two exams composed of multiple questions testing the application of medical knowledge can be of an identical level of difficulty. The pass mark varies slightly from exam to exam because it needs to reflect the subtle differences in how easy or difficult the different questions are.

There are no 'arbitrary' decisions about pass marks. To maintain the standard and therefore be fair to all candidates and patients, the pass mark is mathematically and proportionately raised for easier examinations and lowered for more demanding ones, using the technique of IRT equating.

We take independent expert psychometric advice on maintaining the standard between exams using internationally accepted and published methods, consistent with other postgraduate medical (and non-medical) exams. These methods are approved by the GMC and on several occasions over the years have been reviewed and re-approved by external international assessment experts.

5. Is the pass mark intentionally different between the January, April, July and October exams?

There is no intentional alteration to exam difficulty at different times of the year. It is simply a consequence of selecting an exam of multiple different items that creates some slight differences in the level of challenge set.

The required passing score must reflect the measured 'level of difficulty'.

Example 1:

- Candidate 1 and Candidate 2 - of the same good capability
- Take two different exams - Exam A and Exam B
- At different times in the year - May and October

We want both to be able to pass - even if Candidate 2 takes exam B in October (an exam that is measured to be slightly harder than exam A as sat by Candidate 1 in the previous May).

Example 2:

- For example, if the pass mark was always set at a fixed 60% each sitting
- But the questions in paper A were slightly easier than questions in paper B
- That would make it easier for candidates sitting paper A to pass the exam than candidates sitting paper B.

So, we must reliably adjust the pass mark in line with the difficulty of the questions (in this case, have a slightly higher pass mark for paper A than for paper B).

6. Do you pass a fixed proportion of the candidates?

No, this is not a 'norm-referenced' exam. That term means that the pass mark is not determined with reference to the mean score, nor with reference to a pre-determined proportion of candidates passing the examination.

Instead, this is a 'criterion-referenced' exam to which we apply a 'criterion-referenced' standard. The pass mark is set to reflect the level of competence required for independent practice as a GP.

In theory, we could see a 100% pass rate if all the candidates demonstrated that they had met or exceeded that level of competence and 0% if none of the candidates demonstrated the required level of competence.

7. What does it mean if a result states that a candidate has been unsuccessful by a small percentage?

There is no negative marking in the AKT.

It is important to remember that the AKT is a criterion-referenced examination, as outlined in FAQ 6. Therefore, any score below the designated pass mark (however small) is deemed to have not met the standard required to confer a Certificate of Completion of Training (CCT). This applies even when a candidates' total score is a single mark below the cut-score.

The pass mark for an individual AKT is recommended only after a thorough quality assurance review of all the questions. Candidate scores are machine marked and there can be no alteration of a pass mark for individual candidates. This applies even when it is recognised by the Educational Supervisor that a candidate is doing well in all other aspects of their workplace-based performance.

The final pass mark for an individual AKT is set by a formal standard setting process which is subjected to external scrutiny in the form of an Exam Board. Conferment of a CCT entitles the doctor to practice independently with no further supervision, and the decision to award a CCT must be based on sound evidence, including achieving a passing score in the AKT.

8. How confident are you that the AKT is an accurate assessment?

The accuracy of an examination is tested by estimating its reliability. This is expressed as a coefficient, on a scale from 0 to 1. A value of 0.8 is widely accepted as the minimum desirable for single best answer examinations such as the AKT and a value of 0.9 is recognised as being at the forefront of international best practice.

The AKT consistently shows values very close to 0.9 as estimated by Cronbach's alpha Coefficient. The mean and median alpha for all AKTs combined (AKT 1-54) is over 0.9.

The inherent error of measurement is factored into the passing standard, as is typical for medical assessments. This error becomes smaller as the test reliability increases.

January 2025

Published by the MRCGP AKT team

We welcome feedback on this document via exams@rcgp.org.uk