



A TRADITION OF
INDEPENDENT
THINKING



Multiple Choice Questions

HOW TO SET THE STANDARD IN MCQ EXAMINATIONS

DR HELEN HYNES, MEDICAL EDUCATION UNIT, UCC

This presentation from the Medical Education Unit in University College Cork, describes how to carry out standard setting for multiple choice examination. In a previous presentation we have looked at what standard setting means and why it is used in medical education.



Standard Setting for MCQ Examinations

The following methods for standard setting MCQs will be described in this presentation:

- Modified Angoff ^{1,2,3,4}
- Ebel ^{2,5}
- Hofstee ^{2,6}
- Cohen ^{7,8}

Please see the Overview of Standard Setting presentation for more details on the theory behind standard setting.

There are a number of different methods of standard setting that have been validated for use in MCQ type examinations.

These include the Modified Angoff ^{1,2,3,4}, Ebel ^{2,5}, Hofstee ^{2,6} and Cohen ^{7,8} methods.

I will describe each of these methods and give references to related publications.

The method that is most used in UCC's Medical Education Unit is the Modified Angoff, so we will describe this in the greatest detail.



Modified Angoff Procedure

The Angoff Method was first described in 1971:

- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington DC: American Council on Education ¹

The method we use is a modified version of the original process.

A group of **expert judges** makes estimates of how a **borderline candidate** would perform on each item in the examination.

- Expert = expert on the subject matter AND on the module content / expected standard of the candidates
- What is a Borderline Candidate?

The Angoff Method was first described in 1971, and since then various modifications have been proposed.

It is widely used in medical education nationally and internationally.

A group of expert judges make estimates on how a borderline candidate would perform on each item in the test.

Ideally a panel of at least 6-8 judges should be involved in the process.



The Experts' Role ^{2,3}

The experts' role is:

- To conceptualise the minimum level of performance required for a pass in the examination.

The experts must:

- be knowledgeable about the candidate population (for example first years or final years);
- understand the standard expected at this level;
- be knowledgeable about the subject matter; and
- understand what has been taught about the subject matter to this cohort of students.

Standard setters should receive training, so that they can provide their judgements in an informed manner. This training should familiarise them with both the standard setting task and the conceptual level required for a pass. It can also provide an opportunity for standard setters to calibrate their expectations using past performance data. ²

The Experts in the standard setting process are 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 (for example first years or final years);
- understand the standard expected at this level;
- be knowledgeable about the subject matter; and
- understand what has been taught about the subject matter to this cohort of students.

Standard setters should receive training, so that they can provide their judgements in an informed manner. This training should familiarise them with both the standard setting task and the conceptual level required for a pass. It should ideally provide an opportunity for standard setters to calibrate their expectations using past performance data.



The Borderline Candidate ^{2,3}

At the beginning of every modified Angoff process, the panel of judges should conceptualize the "Borderline Candidate" / Minimally Competent Candidate.

This candidate demonstrates the knowledge / skills that are just about at the level which differentiates pass or fail.

Definition of Borderline Candidate used for clinical skills:

- Pass doubtful
- Patchy performance of skill.
- Examiner undecided whether to pass or fail candidate.
- Demonstrated some aspects of the skill however omissions and inaccuracies occurred in their performance of the skill.
- Often formulaic in approach and struggled with performing skill.

The next concept that we need to explore is that of the borderline candidate.

At the beginning of every modified Angoff process, the panel of judges should conceptualize the "Borderline Candidate" / Minimally Competent Candidate.

This candidate demonstrates the knowledge / skills that are just about at the level which differentiates pass or fail.

We use the concept of the borderline candidate routinely in our clinical OSCE exams so this is generally familiar to clinical examiners but it may be less familiar to examiners in other fields. We describe the borderline candidate as one where their performance is patchy, they may demonstrate some aspect of the required knowledge or skill but they also demonstrate multiple omissions and errors.

The Work of the Expert Panel in Modified Angoff

Each member of the panel of judges estimates the proportion of borderline examinees who will answer an item correctly.

This is equivalent to estimating the candidate's likelihood of answering an item correctly.⁸

Estimates are averaged over judges and summed over items to create a standard (cut-off score).

Each member of the panel of judges estimates the proportion of borderline examinees who will answer an item correctly.

This is equivalent to estimating the candidate's likelihood of answering an item correctly⁹.

Estimates are averaged over judges and summed over items to create a standard (cut-off score).



How to do a modified Angoff

The Examination coordinator convenes a panel of experts.

Training should be provided to the panel of standard setters.

This focuses on explaining what has been taught to the students and what level the students are at.

The panel should then focus on conceptualising the Borderline Candidate.

The next step is for each member of the panel to read the examination and for each item to answer the question:

"What percentage of borderline candidates would answer this question correctly?"

Each member of the panel should fill out a spreadsheet such as the one shown in the next slide.

The Examination coordinator convenes a panel of experts. These experts might include anyone who teaches on the module, or who teaches on that subject in other modules, tutors, post grads.

Training should be provided to the panel of standard setters.

The training focuses on explaining what has been taught to the students and what level the students are at.

The panel should then focus on conceptualising the Borderline Candidate.

The next step is for each member of the panel to read the examination and for each item to answer the question:

"What percentage of borderline candidates would answer this question correctly?"

Each member of the panel should fill out a spreadsheet such as the one shown in the next slide.



Modified Angoff Spreadsheet

| Question Number | What percentage of borderline candidates do you think would answer this question correctly? | | | |
|-----------------|---|--|--|--|
| | Your Initials | | | |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| ... | | | | |
| ... | | | | |
| Last Question | | | | |

Each examiner fills in their initials as shown and then records the percentage of Borderline candidates that they think would answer each question correctly, based on how difficult they think the question would be for this cohort of students.



Modified Angoff and guessing

If the modified Angoff is being used in a single best answer MCQ, it is important to remember that just by guessing, a completely incompetent candidate would statistically be expected to answer each question correctly.

So for example if the question has 5 possible answers, then 20% of candidates with no prior knowledge would be expected to answer each question correctly.

So when answering the question "What percentage of borderline candidates do you think would answer this question correctly?" we need to bear this in mind.

For a question with 5 possible answers I would ask the examiners to give a percentage between 20% -100% to allow for that.

If the modified Angoff is being used in a single best answer MCQ, it is important to remember that just by guessing, a proportion of completely incompetent candidates would statistically be expected to answer each question correctly.

So for example if the question has 5 possible answers, then 20% of candidates with no prior knowledge would be expected to answer each question correctly.

So when answering the question "What percentage of borderline candidates do you think would answer this question correctly?" we need to bear this in mind.

For a question with 5 possible answers I would ask the examiners to give a percentage between 20% -100% to allow for random chance.



Worked example from a previous MCQ

| Question-number | A | B | C | D | E | F | G | | |
|-----------------|----|----|----|----|----|----|----|--|--|
| 1 | 50 | 30 | 30 | 30 | 50 | 40 | 40 | | |
| 2 | 40 | 25 | 30 | 40 | 30 | 50 | 40 | | |
| 3 | 50 | 60 | 60 | 40 | 40 | 40 | 70 | | |
| 4 | 60 | 75 | 40 | 50 | 40 | 70 | 70 | | |
| 5 | 35 | 30 | 20 | 20 | 50 | 40 | 40 | | |
| 6 | 40 | 40 | 20 | 20 | 30 | 40 | 40 | | |
| 7 | 75 | 40 | 30 | 40 | 40 | 50 | 70 | | |
| 8 | 40 | 50 | 70 | 40 | 40 | 60 | 40 | | |
| 9 | 70 | 75 | 70 | 30 | 50 | 50 | 60 | | |
| 10 | 70 | 35 | 50 | 30 | 40 | 40 | 40 | | |
| 11 | 40 | 40 | 40 | 40 | 70 | 50 | 40 | | |
| 12 | 50 | 70 | 40 | 40 | 40 | 50 | 70 | | |
| 13 | 55 | 50 | 40 | 40 | 40 | 40 | 70 | | |
| 14 | 50 | 30 | 70 | 40 | 30 | 40 | 50 | | |
| 15 | 50 | 40 | 70 | 30 | 50 | 50 | 40 | | |
| 16 | 75 | 70 | 40 | 40 | 40 | 40 | 40 | | |
| 17 | 40 | 40 | 40 | 40 | 30 | 40 | 50 | | |
| 18 | 30 | 40 | 40 | 30 | 50 | 40 | 70 | | |
| 19 | 40 | 50 | 40 | 40 | 40 | 40 | 70 | | |
| 20 | 50 | 50 | 50 | 40 | 50 | 50 | 70 | | |

On this slide we can see a worked example from a previous MCQ used in the School of Medicine on a clinical paper.

We can see that 7 examiners have filled in their percentages for each question. Usually we record the examiners initials but as this is actual data from School of Medicine examiners, I have replaced the initials with A, B, C, D and so on to protect anonymity.

Each column represents a separate examiner and each row represents a separate exam question.

For this exam 7 examiners participated in the standard setting. We usually ask all clinical module coordinators and clinical tutors who are involved in the year to participate. This gives us a good mix of subject expertise and also knowledge of what the students have been taught and the standard expected from them.



Worked example from a previous MCQ

| Question number | What percentage of borderline candidates do you think would answer this question correctly? | | | | | | |
|-----------------|---|----|----|----|----|----|----|
| | A | B | C | D | E | F | G |
| 1 | 30 | 30 | 30 | 30 | 30 | 40 | 40 |
| 2 | 40 | 25 | 30 | 40 | 30 | 50 | 40 |
| 3 | 30 | 60 | 50 | 40 | 40 | 40 | 70 |
| 4 | 40 | 75 | 40 | 50 | 40 | 70 | 70 |
| 5 | 35 | 30 | 20 | 20 | 40 | 40 | 40 |
| 6 | 40 | 40 | 20 | 20 | 30 | 40 | 40 |
| 7 | 75 | 40 | 30 | 40 | 40 | 50 | 70 |
| 8 | 40 | 30 | 70 | 40 | 40 | 40 | 40 |
| 9 | 70 | 75 | 70 | 30 | 50 | 50 | 60 |
| 10 | 70 | 35 | 50 | 30 | 40 | 40 | 40 |
| 11 | 40 | 60 | 40 | 40 | 70 | 50 | 40 |
| 12 | 50 | 30 | 40 | 40 | 40 | 50 | 70 |
| 13 | 35 | 50 | 40 | 40 | 40 | 40 | 70 |
| 14 | 50 | 30 | 70 | 40 | 30 | 40 | 50 |
| 15 | 30 | 40 | 70 | 30 | 30 | 50 | 40 |
| 16 | 75 | 70 | 40 | 40 | 40 | 40 | 40 |
| 17 | 40 | 40 | 40 | 40 | 30 | 40 | 30 |
| 18 | 30 | 40 | 40 | 30 | 50 | 40 | 70 |
| 19 | 40 | 30 | 40 | 40 | 40 | 40 | 70 |
| 20 | 50 | 30 | 70 | 40 | 50 | 50 | 70 |

The next step is for the panel to compare their scores.

Looking at question 1 here, we can see that there is broad agreement between examiners, with estimates ranging from 30% to 50%.

However look at question 14 – here we see a big discrepancy with estimates ranging from 30% to 80%. Now the examination coordinator or usually module coordinator should step in.

Perhaps, for example, this might be a difficult concept, but the students may have had explicit teaching on this subject. Some examiners may be aware of this and others may not.

So at this stage, the questions are reviewed, discrepancies are discussed, and examiners can choose to review their original estimates. If a broad consensus cannot be reaching on any particular question then the module coordinator should consider removing that question from the paper entirely.



Worked example from a previous MCQ

| What percentage of borderline candidates do you think would answer this question correctly? | | | | | | | | |
|---|----|----|----|----|----|----|----|------|
| Question Number | A | B | C | D | E | F | G | Mean |
| 1 | 30 | 50 | 30 | 30 | 50 | 60 | 60 | 38.6 |
| 2 | 60 | 25 | 30 | 60 | 30 | 50 | 60 | 36.4 |
| 3 | 50 | 60 | 60 | 60 | 60 | 60 | 70 | 57.1 |
| 4 | 60 | 75 | 60 | 50 | 60 | 70 | 70 | 57.9 |
| 5 | 35 | 30 | 25 | 20 | 60 | 60 | 60 | 35.0 |
| 6 | 60 | 60 | 25 | 20 | 30 | 60 | 60 | 31.9 |
| 7 | 75 | 60 | 30 | 60 | 60 | 50 | 70 | 49.3 |
| 8 | 60 | 50 | 75 | 60 | 60 | 60 | 60 | 57.1 |
| 9 | 70 | 75 | 75 | 30 | 50 | 50 | 60 | 57.9 |
| 10 | 70 | 35 | 50 | 30 | 60 | 60 | 60 | 43.6 |
| 11 | 60 | 60 | 60 | 60 | 70 | 50 | 60 | 60.0 |
| 12 | 50 | 70 | 60 | 60 | 60 | 50 | 70 | 61.9 |
| 13 | 55 | 50 | 60 | 60 | 60 | 60 | 70 | 61.9 |
| 14 | 50 | 70 | 75 | 60 | 70 | 60 | 50 | 64.3 |
| 15 | 50 | 60 | 75 | 30 | 50 | 50 | 60 | 50.0 |
| 16 | 75 | 70 | 60 | 60 | 60 | 60 | 60 | 66.4 |
| 17 | 60 | 60 | 60 | 60 | 30 | 60 | 50 | 54.3 |
| 18 | 30 | 60 | 60 | 30 | 50 | 60 | 70 | 45.7 |
| 19 | 60 | 50 | 60 | 60 | 60 | 60 | 70 | 51.4 |
| 20 | 50 | 50 | 90 | 40 | 50 | 50 | 70 | 57.1 |
| Overall Mean | | | | | | | | 52.3 |

Mean for each question

Mean overall for all 20 items = New pass mark

The next step is to calculate the mean percentage per question and then the overall mean. The overall mean becomes the new pass mark – in this case 52.3%.



Editing the students' mark to reflect UCC Marks and Standards

As UCC uses 50% as the Pass Mark for examinations in the medical degree programmes, the students' actual marks are amended taking into account the new pass mark (cut score).

This is the formula used:

Amended mark = (actual mark X old pass mark)/ new pass mark.

Old pass mark = 50%

For example, if the student's actual score is 60/100 and the new pass mark (cut-score) is 55%, then the student's amended mark is $(60 \times 50)/55 = 54.5\%$

As UCC uses 50% as the Pass Mark for examinations in the medical degree programmes, the students' actual marks are amended taking into account the new pass mark (cut score).

This is the formula used: Amended mark = (actual mark X old pass mark)/ new pass mark.

For example, if a student's actual score is 60/100 and the new pass mark / cut-off score is 55%, the student's amended mark is $(60 \times 50)/55 = 54.5\%$



Modified Angoff - Summary

Angoff's method is relatively easy to use, there is a sizeable body of research to support it, and it is frequently applied in licensing and certifying settings.

This process can be time consuming when first used.

However it is much easier to use when the panel have done it once or twice in the past.

This method produces absolute standards, so it is well suited to tests that seek to establish competence.

Angoff's method is relatively easy to use, there is a sizeable body of research to support it, and it is frequently applied in licensing and certifying settings.

This process can be time consuming when first used.

However it is much easier to use when the panel have done it once or twice in the past.

This method produces absolute standards, so it is well suited to tests that seek to establish competence.



Ebel 2,5

A team of expert judges reviews the test. They rate each item on 2 dimensions –

- Difficulty : easy / medium / hard
- Importance: essential / important / acceptable (nice to know)

| | Easy | Medium | Hard |
|------------|------|--------|------|
| Essential | | | |
| Important | | | |
| Acceptable | | | |

Another method that can be used for standard setting is the Ebel method. This has been in use since 1986. In the Ebel method, again, we have a team of judges who review each item in the test. They rate each item on 2 dimensions – difficulty and importance.

Each member of a panel of standard-setters completes a 3x3 grid, allocating every question to one of the nine boxes in the grid.



Ebel

A 17-year old male presents to his GP with blood mixed in with his stools. He is very worried because his grandfather died of rectal carcinoma aged 68. On further questioning, the GP learns that he has had intermittent diarrhoea for 6 months, with crampy abdominal pain. These symptoms occasionally wake him from sleep. He also has frequent mouth ulcers and some rectal irritation. He has not been on any foreign travel.

Which of the following is the most likely diagnosis?

- A. Coeliac disease
- B. Crohn's Disease
- C. Irritable Bowel Syndrome
- D. Rectal tumour
- E. Ulcerative Colitis

| Q1 | Easy | Medium | Hard |
|------------|------|--------|------|
| Essential | | | |
| Important | | x | |
| Acceptable | | | |

So looking at this sample MCQ question, an examiner might decide to rate this question as Important and of medium difficulty.



Ebel

Examiners may have differing opinions about how to categorize any given question.

Then the question should be discussed by the panel, including any relevant information about how the topic was covered in teaching.

A consensus is then reached by the panel for each question.

| Q1 | Easy | Medium | Hard |
|------------|------|--------|------|
| Essential | | ? | |
| Important | | ? | |
| Acceptable | | | |

Examiners may have differing opinions about how to categorize any given question.

Then the question should be discussed by the panel, including any relevant information about how the topic was covered in teaching.

A consensus is then reached by the panel for each question.



Ebel

Examiners agree on the definition of the minimally competent / borderline candidate.

Examiners estimate the percentage of questions in each of the 9 categories that a minimally competent candidate would answer correctly.

| | Easy | Medium | Hard |
|------------|------|--------|------|
| Essential | 90 | 80 | 70 |
| Important | 80 | 60 | 50 |
| Acceptable | 70 | 40 | 30 |

Next the experts agree on the definition of a minimally competent examinee. Then another grid is filled out, this time estimating the percentage of questions in each category that a borderline / minimally competent candidate would answer correctly.



Ebel

| | % of questions in this category that a borderline candidate would answer correctly | No of questions in the test in this category | % in this category that borderline candidate would answer correctly x no of questions in category | |
|---------------------|--|--|---|--|
| Essential / Easy | 90 | 7 | 630 | |
| Important / Easy | 80 | 8 | 640 | |
| Acceptable / Easy | 70 | 5 | 350 | |
| Essential / Medium | 80 | 4 | 320 | |
| Important / Medium | 60 | 8 | 480 | |
| Acceptable / Medium | 40 | 4 | 160 | |
| Essential / Hard | 70 | 7 | 490 | |
| Important / Hard | 50 | 7 | 350 | |
| Acceptable / Hard | 30 | 6 | 180 | |
| | | # of questions =60 | Total =3600; Pass Mark =3600/60 = 60 | |

So in this table I have transferred the 9 boxes on the last grid into the first 2 columns we see here.

Next we go back to the test and count how many items were judged to be in each of the 9 categories.

So in this example 7 questions were judged by the expert panel to be Essential and Easy, 8 were judged to be Important and Easy and so on.

The percentage in each category that the panel believed would answer questions in that category correctly is multiplied by the number of questions that category contains.

The passing score is set by averaging the category scores. So in this case the average category score is the total score for all the 9 categories divided by the total number of questions, which is 60. So $3600/60 = 60$ which now becomes the pass mark of the test.



Hofstee Method ^{3,6}

A compromise between relative and absolute standards.

The examiners estimate 4 values:

- The minimum acceptable failure rate
- The maximum acceptable failure rate
- The minimum pass mark (cutscore), even if all examinees failed
- The maximum passmark (cutscore), even if all examinees passed

Their responses serve as the focus for discussion, with all being free to change their estimates.

These minimum and maximum failure rates and percent correct scores are averaged across panelists and projected onto the actual score distribution to derive a passing score.

The Hofstee Method is another way of standard setting. It is described as a compromise method, using a combination of relative and absolute standards.

The examiners estimate 4 values:

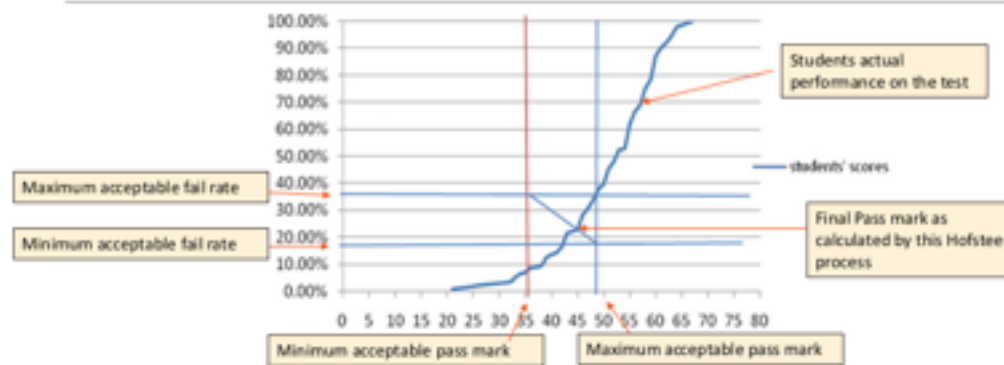
- The minimum acceptable failure rate
- The maximum acceptable failure rate
- The minimum pass mark (cutscore), even if all examinees failed
- The maximum passmark (cutscore), even if all examinees passed

Their responses serve as the focus for discussion, with all being free to change their estimates.

These minimum and maximum failure rates and percent correct scores are averaged across panelists and projected onto the actual score distribution to derive a passing score as we see on the next slide.



Hofstee Worked Example



Kamal et al ¹⁰, Education in Medicine Journal, 2018; 10(2): 15–25; DOI: [10.21315/eimj2018.10.2.3](https://doi.org/10.21315/eimj2018.10.2.3)

This is a worked example taken from Kamal et al ¹⁰.

The data refers to a Final Med MCQ paper with 80 questions on the paper.

As we see on the Y axis, the examiners set the minimum acceptable fail rate at 17% and the maximum acceptable fail rate at 36%.

Looking at the x axis we can see that they set the minimum acceptable pass mark at 36/80 and the maximum acceptable pass mark at 48/80.

Finally we see the curve of the students' actual performance on the test.

Look at the 2 horizontal lines made up by the minimum and maximum fail rates.

Now look at the vertical lines made up by the minimum and maximum acceptable pass marks. These 4 lines intersect in a rectangle as shown on the graph.

A diagonal is drawn across the rectangle. The point where that diagonal line intersects with the curve of the students' actual performance becomes the pass mark

for the exam – so in this case the pass mark is set at 45/80.



Hofstee

Easy to implement

Judges are comfortable with the questions they are asked.

Under some circumstances, however, the pass mark may not be within the bounds they define and when this happens, the standard becomes the maximum or minimum acceptable pass rate.

Consequently, this method would not be ideal for ongoing application in a high stakes test of competence, but is well suited to occasional use and lower stakes settings.

The advantages of the Hofstee method are that it is easy to implement, and that the questions asked of the examiners are less abstract than in some of the other methods.

However, it can happen the the pass mark defined by the process is not within the bounds of the actual scores on the exam and when this happens the standard becomes the maximum or minimum acceptable pass mark identified by the examiners.

For this reason the Hofstee method is less suited to high stakes exams.



Cohen Method ^{7,8}

This is a relative standard – the pass mark is based on the performance of the highest scoring students in the class.

The pass mark is set at 60% of highest achiever's score (or 60% of the mean of the top 3 highest achievers' scores, or 60% of the 90th / 95th centile.)

Need to have at least 100 students to use this method with confidence

Fast, easy to calculate

Useful as a "reality check" if other methods are used.

Acceptable to use with lower stakes exams.

The Cohen Method is a simple and fast way of standard setting.

Various modifications have been suggested in the literature.

The basic method is to set the pass mark at 60% of the highest achievers' score, or 60% of the mean of the top 3 highest achievers' scores, or at 60% of the 90th or 95th centile.

You need to have at least 100 students in the cohort to be able to use this with any degree of statistical confidence.

I use it as part of a post exam evaluation to reality check the pass mark that I have arrived at by doing a modified Angoff.



Summary

Whichever standard setting method is used, we must follow these guidelines ³:

The method must:

- Produce standards consistent with the purpose of the test
- Rely on informed expert judgement, taking into account careful analysis and judgement of acceptable performance
- Take into account test difficulty and student criteria.
- Demonstrate due diligence
- Be easy to explain and implement
- Be supported by a body of research

Whichever standard setting method is used, we must follow these guidelines ³:

The method must:

- Produce standards consistent with the purpose of the test
- Rely on informed expert judgement, taking into account careful analysis and judgement of acceptable performance
- Take into account test difficulty and student criteria.
- Demonstrate due diligence
- Be easy to explain and implement
- Be supported by a body of research

References

1. Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington DC: American Council on Education
2. Ben-David MF. AMEE Guide No. 38: Standard setting in student assessment. *Medical Teacher* 2000; 22(2):120-30
3. Nordini JJ. Setting standards on educational tests. *Medical Education* 2009; 37(5):464-9
4. Bandaranayake RC. *Med Teach*. 2008; 30(9-10):836-45. [Setting and maintaining standards in multiple choice examinations: AMEE Guide No37.](#), doi: 10.1080/01421590802402247.PMID: 19117221
5. Ebel RL. *Essential of educational measurement*. Englewood Cliffs (NJ): Prentice Hall; 1972. p. 622.
6. Hofstee WKB. The case for compromise in educational selection and grading. In: Anderson SB, Helminck JS, editors. *On Educational Testing*. San Francisco: Jossey-Bass; 1983. pp. 109-27.
7. Cohen Schotanus J and Van der Weuten CP. A standard setting method with the best performing students as point of reference: practical and affordable. *Medical Teacher* 2010; 32(2):154-60
8. Celia A. Taylor (2011) Development of a modified Cohen method of standard setting, *Medical Teacher*, 33:12, e678-e682, DOI: 10.3109/0142159X.2011.611192
9. Sizmar S. Look Back in Angoff: A Cautionary Tale. *British Educational Research Journal*, Vol. 23, No. 1 (Feb., 1997), pp. 3-13
10. Kamal et al, *Education in Medicine Journal*. 2018; 10(2): 15-25; DOI: [10.21315/eimj2018.10.2.3](#)