

ARTICLE INFO

Submitted: 04-04-2018

Accepted: 08-05-2018

Online: 29-06-2018

Evaluation of Two Applied Methods for Standard Setting in Undergraduate Medical Programme at the Faculty of Medicine, Suez Canal University

Doaa Kamal¹, Shima ElAraby¹, Mohamed Hany Kamel²,
Somaya Hosny³

¹Medical Education Department, Faculty of Medicine, Suez Canal University, Ismailia, EGYPT

²Family Medicine Department, Faculty of Medicine, Helwan University Cairo, Helwan, EGYPT

³Faculty of Medicine (Former Dean), Suez Canal University, Ismailia, EGYPT

To cite this article: Kamal D, ElAraby S, Kamel MH, Hosny S. Evaluation of two applied methods for standard setting in undergraduate medical programme at the Faculty of Medicine, Suez Canal University. *Education in Medicine Journal*. 2018;10(2):15–25. <https://doi.org/10.21315/eimj2018.10.2.3>

To link to this article: <https://doi.org/10.21315/eimj2018.10.2.3>

ABSTRACT

Standard setting process is known to overcome the problem of the arbitrary pass score which is prone to underrate or overrate students. At Faculty of Medicine, Suez Canal University (FOM/SCU), arbitrary 60% cutoff score is used in written exams. The objective of the study is to evaluate the efficacy of applying two standard setting methods in undergraduate medical programme of FOM/SCU. Two standard setting methods (Angoff and Hofstee) were applied on the integrated multiple choice question (MCQ) exam of the third year FOM/SCU undergraduate students. Eleven judges from different specialties participated in this study. The cutoff score was determined by each method. Focus group discussion was conducted to get the judges' perception regarding standard setting process, importance, judges' selection and challenges. The determined cutoff scores using Angoff and Hofstee methods were 54.98% and 44% respectively. Thematic analysis of focus group discussion revealed judges' acceptance for standard setting application on all exams and they agreed on involving non-specialists in the judges' committee. There is a difference between the fixed cutoff score currently used at FOM/SCU and those determined using Angoff and Hofstee methods. Multidisciplinary judges are satisfied with this difference and they echoed the use of standard setting methods in improving assessment process.

Keywords: *Angoff method, Hofstee method, Cutoff score*

CORRESPONDING AUTHOR

Shima ElAraby, McS, PhD, Medical Education Department, Faculty of Medicine, Suez Canal University, 41522, Round Road, Ismailia, Egypt | E-mail: shimaa80me@yahoo.com

INTRODUCTION

Standard setting is defined as the proper following of a prescribed, rational system of rules or procedures resulting

in the assignment of a number to differentiate between two or more states or degrees of performance (1). Another suggested definition of standard setting is the systematic way of collecting value

judgements, reaching consensus, and transforming that consensus into a single score on a test (2).

Standard setting overcomes the problem associated with the traditional pass score in which there is a possibility of either failing examinees who have achieved the sufficient level of knowledge and skills if the standard is very high or passing examinees who did not achieve the requirements if the standard is too low (3).

In 1977, Angoff developed a systematic procedure for deciding on the raw scores for passing a test. In his procedure, a group of judges was asked to make estimates about how borderline students would perform on items in the examination i.e. the proportion of borderline students who will answer each item correctly. Judges estimates were then averaged and summed over items to determine a cut-off score (4). The steps applied in Angoff method include: a) Selecting at least five raters (judges) of different genders, ages, races, and educational backgrounds; b) Taking the test by the raters and then giving their feedback on the test items; c) Reaching to a consensus regarding the definition of the minimally qualified examinee or the borderline student; d) Rating the test items individually to estimate the percentage of borderline examinees that would be able to answer the item correctly; e) Reviewing the ratings and getting the average percentage correct for every test item; f) Establishing the cut score of the whole exam by adding the average percentage correct numbers and dividing by the total number of the test items (3).

In 1983, Willem Hofstee proposed a compromise method in which panelists were asked the following questions: What are the maximum and minimum acceptable cut-off scores?; What are maximum and minimum acceptable failure rates? The answers to those questions were then graphed in a (Hofstee) plot. The main aim of the Hofstee method is to consider the true scores of the students while determining the standards and cutoff score values and failure rates (5).

The credibility of the cutoff score resulted from any standard setting method depends on the consistency of the used method with the purpose of the test and the judgement of the experts who have data about examinees performance. Furthermore, the standard setting exercise should neither be so brief to indicate a lack of credible consideration nor too long to cause a burden on the standard setter. It should also be transparent, and easy to explain and be implemented (2).

Several factors were found to affect judges' decisions; these factors are divided into: factors within the judges, in the social milieu in which the process of standard setting is carried out; and factors that arise by the interaction between the previous factors. For example, a weak judge will be affected by the opinions of his peers when discussions are made between their respective standards. In addition, many judges who have not clearly understood the standard setting process may face difficulty in reaching probabilities and decisions. They also tend to either forget or misunderstand the concept of the borderline examinee (6).

FOM/SCU adopts innovative educational strategies since its establishment in 1978. The educational strategies include problem-based learning (PBL), community-based education (CBE), horizontal & vertical integration, and evidence-based medicine (EBM). In this regard, comprehensive student assessment is used to align with the school strategies. The used assessment methods cover the three learning domains and have several formats for each domain (7).

The written assessment tests, in FOM/SCU, are in the form of modified essay questions (MEQ), short answer questions and MCQ. In all types, the school uses the traditional approaches in defining the cut-off point (arbitrary cutoff score of 60%) which do not provide strong and valid evidence for pass/fail decisions. Thus this study aims at evaluating two standard setting methods that were applied on one of the MCQ exams in order to investigate the feasibility of their

application on all the written exams of the school which will be reflected on improving the validity of the assessment scores and improving the item construction process.

METHODS

This is a mixed quantitative and qualitative study, conducted in the period from 2016 to 2017. In this study, the Angoff and Hofstee methods were applied on the final integrated MCQ exam of the third year medical students in the academic Year of 2014/2015 at Faculty of Medicine, Suez Canal University, Ismailia, Egypt.

Although there are many types of standard setting methods, we used Angoff and Hofstee methods in this study. The reason for selecting those two methods is that the Angoff method is one of the most widely used criterion-referenced methods, while Hofstee method is a compromise type which addresses both the criterion and norm-referenced expectations.

Participants and Materials

A sample of 11 judges from the FOM/SCU faculty members volunteered to participate in the study. The volunteered judges are representing all basic sciences departments (pathology, anatomy, physiology, pharmacology, parasitology, histology, genetics, biochemistry, community medicine, family medicine) who participated in the teaching and learning activities of the medical students and thus were familiar with the curriculum.

The MCQ exam included 75 items covering 10 disciplines, the number of items of each discipline is determined according to the assessment blueprint (11 pathology, 13 anatomy, 9 pharmacology, 9 physiology, 9 microbiology, 9 biochemistry, 3 community medicine, 4 histology, 3 genetics, and 5 parasitology).

The scores of all (139) third-year medical students in the final MCQs exam were

used, after taking permission of the school administration. No exclusion criteria were included.

Procedures

A two-day training workshop was conducted. It started with an orientation session in which the standard setting importance, classifications, and methods were presented by the principle investigator followed by identification and reaching to a consensus about the definition of the borderline student. After the orientation session the standard setting process started by applying the Angoff method in the following sequence:

Judges received Angoff rating forms containing the MCQ test items under investigation.

1. The judges were requested to review each test item and give it an Angoff score (the percentage of borderline students who would be able to correctly answer the item).
2. After rating the first five items of the test, a discussion with the judges was held, to ensure their understanding of the process of rating.
3. The process of rating and discussion was repeated for the remaining items under timed conditions.
4. After completion of the rating process, each judge ratings were placed into spreadsheet form and projected onto a screen visible to all judges. For each item, the average was calculated.
5. The cutoff score of the test was estimated as shown in Table 1.

In the next day and after estimating the cutoff score of the test using Angoff method, the judges were requested to apply the Hofstee method procedures (independently from the cutoff score estimated using the Angoff method) as follows:

1. After reviewing all test items, judges were requested to answer the following four questions:
 - a. What is the maximum passing score of the test? (The highest acceptable percent score to allow the borderline examinee to pass);
 - b. What is the minimum passing score of the test? (The lowest acceptable percent score to allow the borderline examinee to pass);
 - c. What is the maximum failure rate of the students? (The highest acceptable percentage of students to fail);
 - d. What is the minimum failure rate of the students? (The lowest acceptable percentage of students to fail).
2. The answers of each judge on the previous four questions were collected.
3. Then the average of each judge answer was then calculated for all the judges (Table 2).
4. The average scores and percentages for the four questions given by the judges and also the true scores of the 139 students were plotted on a cumulative frequency graph to determine the cutoff score for the MCQ exam.
5. The maximum passing mark, minimum passing mark, maximum failure rate and the minimum failure rate determined by the judges were then connected by a rectangle, and then a diagonal line was drawn from the highest fail rate to the lowest fail rate. A line was then drawn

Table 1: Descriptive statistics associated with the test items' Angoff scores

| Item (n = 75) | Rater | | | | | | | | | | | Mean |
|---------------|-------|------|------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| Pathology | 52.3 | 59.1 | 69.5 | 61.8 | 71.8 | 56.4 | 55.9 | 59.1 | 55.9 | 58.2 | 49.1 | 64.3 |
| Anatomy | 59.2 | 56.9 | 55.4 | 61.5 | 65.4 | 53.1 | 54.2 | 61.9 | 55.0 | 57.7 | 55.0 | 56.2 |
| Pharmacology | 57.2 | 52.8 | 55.6 | 54.4 | 55.6 | 45.6 | 43.9 | 53.3 | 66.1 | 50.6 | 49.4 | 51.9 |
| Physiology | 60.0 | 59.4 | 54.4 | 61.1 | 56.7 | 52.2 | 53.9 | 58.9 | 53.3 | 56.1 | 53.9 | 55.8 |
| Microbiology | 61.1 | 58.9 | 51.1 | 63.3 | 53.3 | 56.7 | 58.3 | 53.3 | 59.4 | 62.2 | 55.0 | 55 |
| Biochemistry | 58.9 | 58.3 | 60.0 | 65.6 | 50.0 | 43.3 | 48.9 | 56.1 | 52.2 | 60.6 | 53.3 | 59.2 |
| Community | 58.3 | 60.0 | 68.3 | 76.7 | 56.7 | 56.7 | 55.0 | 65.0 | 53.3 | 60.0 | 48.3 | 64.2 |
| Histology | 66.3 | 48.8 | 56.3 | 57.5 | 52.5 | 45.0 | 50.0 | 65.0 | 63.8 | 60.0 | 46.3 | 57 |
| Genetics | 65.0 | 48.3 | 51.7 | 56.7 | 46.7 | 60.0 | 46.7 | 56.7 | 53.3 | 58.3 | 50.0 | 50 |
| Parasitology | 63.0 | 51.0 | 48.0 | 68.0 | 54.0 | 52.0 | 47.0 | 62.0 | 52.0 | 55.0 | 54.0 | 49.5 |
| Pass score | 54.98 | | | | | | | | | | | |

Table 2: Descriptive statistics associated with the test items' Hofstee scores

| | Rater | | | | | | | | | | | Mean |
|-----------------------|-------|----|----|----|----|----|----|----|----|----|----|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| Minimum passing score | 45 | 45 | 45 | 60 | 40 | 60 | 50 | 40 | 50 | 65 | 40 | 49.1 |
| Maximum passing score | 35 | 35 | 35 | 40 | 30 | 40 | 40 | 35 | 30 | 40 | 30 | 35.5 |
| Minimum fail rate | 30 | 55 | 10 | 40 | 30 | 30 | 50 | 40 | 30 | 50 | 50 | 37.7 |
| Maximum fail rate | 0 | 40 | 0 | 20 | 15 | 15 | 30 | 20 | 0 | 30 | 20 | 17.3 |

vertically to the score axis and the cutoff score is determined as shown in Figure 1.

At the end of the workshop, focus group discussion was conducted with the judges to explore their perception regarding standard setting process, importance, judges' selection, and challenges.

ANALYSIS

Data analysis was performed using the Statistical Package for the Social Sciences (SPSS Version 20, IBM (SPSS, Inc., Chicago, IL, USA). The pass rates were calculated based on the pass scores set by the two applied methods. Thematic analysis was conducted for the focus group discussion to determine the perception of judges regarding standard setting methods. Thematic analysis involved reading and rereading the data transcripts many times, identifying initial themes and subthemes, constructing a hierarchy of themes and subthemes, and summarising and synthesising the data. The analysis was checked and confirmed by the other researchers involved in the study without knowing the identity of the participants to ensure the reliability of data. Validity checks included the experiences of the interviewers, the probing of the questions, and seeking for more clarification from the interviewees.

RESULTS

The Determined Passing Score Using Angoff and Hofstee Methods

For the test under investigation in this study, the pass rate determined using the Angoff method was 54.98% (Table 1) while it was 44% using the Hofstee method. Regardless of standard setting method used, we found that the resulting cutoff scores were 5%–15% lower than the 60% cut score used currently in our school (Figure 2).

The total number of students who failed using the current 60% pass mark was 29/139 (20.9%), however, it was 18/139 (12.9%) using Angoff method and 12/139 (8.6%) using Hofstee method.

Judges' Perception of the Application of Standard Setting Methods Used in this Study

The key findings related to judges' perception of standard setting methods application were shown in Table 3.

The collected data was categorised into four themes: a) The importance of standard setting in student assessment; b) The selection of appropriate judges; c) The benefits of standard setting; and d) Difficulties and challenges of standard setting methods. Subthemes were extracted from the main themes (Table 1).

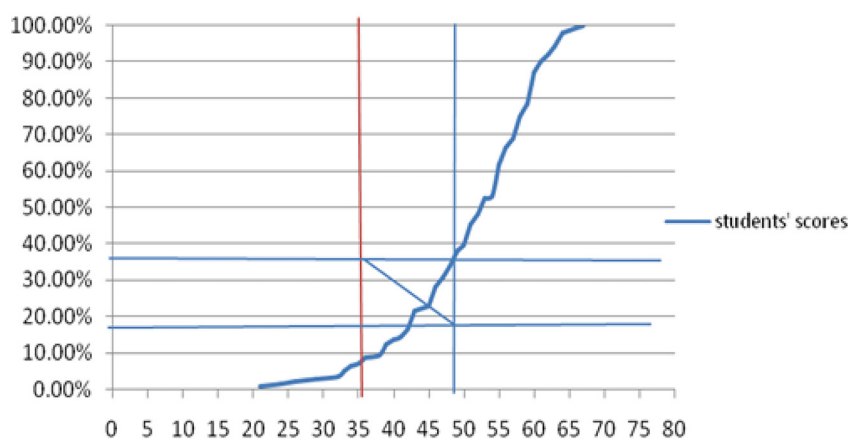


Figure 1: Hofstee calculation for (2014–2015) year three final MCQ exam.

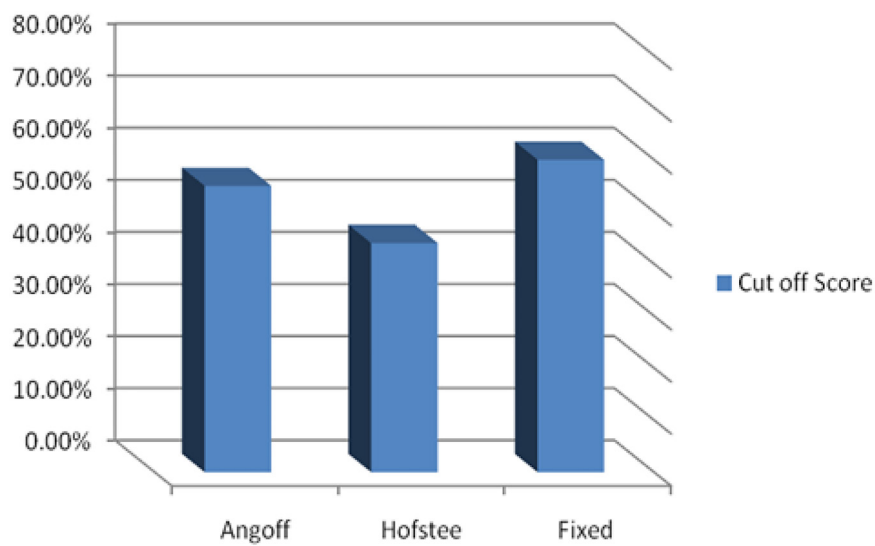


Figure 2: Cutoff scores determined by the three methods.

Table 3: Key findings of the focus group discussion related to judges’ perception of standard setting methods application

| Themes | Subthemes |
|---|---|
| The importance of standard setting in assessment | Difference between traditional cutoff score and that reached after using Angoff and Hofstee methods. Fairness insurance. The gap between students' scores and actual performance level will be reduced. |
| The selection of appropriate judges | Familiarity with students. Subject matter experts vs. non subject matter experts. |
| The benefits of standard setting | Determining item difficulty. Effect on test reliability. |
| Difficulties and challenges of standard setting methods | Definition of borderline students. Time consuming. Confidentiality issues. |

The importance of standard setting in student assessment

Nearly all participants described standard setting as an essential step to ensure the validity of the assessment procedures. They found that the modification of the cutoff score would lead to a significant change in the number of students passing/failing the exam under investigation. For example, one judge said:

Sometimes, I feel shocked when I know the exam scores of one of my students, as I expected him to have a high score, but actually, his score in the exam was low, so I think that standard setting methods will resolve this issue as we revise all the exam questions and decide their pass mark ...

Most of the participants preferred the Hofstee method, they argued that Hofstee method is less subjective than the Angoff method as it takes the students' raw scores in consideration and also more feasible, one judge said that:

Hofstee method is better than Angoff as it takes less time in its procedures; also it takes into consideration the students' scores which will ensure more fairness to the students.

The selection of appropriate judges

The majority of the participants in the focus groups described an important and essential point in judges' selection that the judges should be those who taught, evaluated, spent a lot of time with the students (for example as mentors or facilitators) in order to be familiar with them and consequently be able to properly specify the borderline student's performance. As one participant said:

...but it's important for the senior staff who will participate as judges in the standard setting process to have more contact with the students, they need to teach, mentor the students and become familiar with them.

As the exam is an integrated one, so most of the participants preferred the representation of the judges from different disciplines in the judges' committee to give their input in other disciplines questions as this will lead to a balance in deciding the overall pass mark of a test, this is captured well in the conclusions of one participant:

I think that multiple specialties in one committee are better as it enabled me to put myself in the borderline student's place in specialties other than mine...

The benefits of standard setting

The used methods of standard setting in our workshop allowed the participants to elicit

the difficulty level of each question; as a result, the majority of them identified a need to decrease the variance of the item difficulty distribution in further exams for more reliable test scores. The previous concept is captured in the words of the participant:

I think that standard setting is the final stage in the assessment process as it included judging the item difficulty, which will adjust test difficulty of the subsequent exams.

Difficulties and challenges of standard setting methods

A major source of difficulty for the judges in determining a pass mark was the definition of borderline student, especially if the judges were not familiar with the students, accordingly, the majority of the participants stressed on the importance of proper selection and training of the judges to be able to overcome such difficulty. These sentiments are captured in the comments of one participant who explained:

I personally think that I will not be able to determine who will be the borderline student, unless I teach them more than one time.

Applying standard setting methods in all formative and summative exams seems to be time-consuming as agreed by most of the participants; they recommended the application of such methods in high stakes exams. One participant clarified:

I think that we can perform Angoff and Hofstee methods in the final exams only as it will be difficult and time consuming to use them in determining the pass mark for all periodic exams in all years.

Participants also expressed concerns about the security of the exam, especially when applying standard setting methods before the application of a test. This view was echoed by most of the participants:

Confidentiality of the test could be affected especially when different judges committees participated in the standard setting process...

DISCUSSION

In this study, we applied two standard setting methods on one of the integrated MCQ exams at the Faculty of Medicine, Suez Canal University to determine the pass mark and explore the difference between it and the traditional arbitrary 60% cutoff score.

The results of our study showed that the cutoff score was 54.98% as determined by the Angoff method and 44% as determined by the Hofstee method. This means that the cutoff score of the test under investigation should have been different than the fixed 60% cutoff score used by FOM/SCU.

Different previous studies also found a difference between the arbitrary cutoff score and the cutoff score determined using one or more of the standard setting methods, they also found a difference between the cutoff scores determined using the two methods. Among these studies is the one performed by Hussein et al. which was conducted at the College of Medicine, University of Sharjah, United Arab Emirates. In their study, they used Angoff and Hofstee methods to compare the cutoff scores of examination results of a total of eight cohorts of undergraduate medical students from Year 1, Year 2, or Year 3 in the two academic years 2007/2008 and 2008/2009. The results of their study revealed that the Angoff cutoff scores for the eight cohorts ranged between 57% and 66%, while according to Hofstee method they ranged between 47% and 55%, both groups of cutoff scores were different from the previous fixed 60% cutoff scores used in their institution (8).

The results of our study are also consistent with those of Yousef, et al. whose study was conducted at the Faculty of Dentistry

at King Abdulaziz University. They used Angoff and Hofstee methods to determine the cutoff scores for summative examinations for three academic years (2012, 2013, and 2014). They also compared the new cutoff scores with the existed fixed passing score of 60%. The results of their study showed that both methods resulted in a cutoff score which is different from the fixed previous 60% cutoff score (64.5% using Angoff method and 63.5% using Hofstee method) (9).

Our findings also agreed with those of the study conducted by Elfaki and his colleague in which two standard setting methods were applied on the raw scores of 106 final medical students on a multiple-choice exam in internal medicine at the faculty of Medicine of King Khalid University. They used the modified Angoff method and the norm-reference method. The cutoff scores derived from the two methods were different from the arbitrary 60% cutoff score (10).

The results of the current study also revealed a real need to use standard setting methods to set the cutoff scores in assessment; this was clear in the focus group analysis which revealed different themes. Almost all participants described that the changed cutoff score resulted in a significant modification of the number of students passing the exam; this made standard setting an essential step to ensure the soundness of the assessment measures.

Some of the participants in the focus group discussion in our study preferred the use of Hofstee method than the Angoff method to increase the credibility of the cutoff score as they found it more objective than the Angoff method and also because Hofstee method takes into consideration the actual scores of the examinees. This is congruent with the finding of a study conducted by Mogapi in Botswana. In his study, the Angoff grading model was applied following the introduction of criterion-referenced testing. By comparing the cutoff scores determined using the Angoff method with those determined using holistic grading

models; he found that Angoff method produced markedly different cutoff scores which he considered as a weakness in the Angoff method. He suggested the use of other methods in combination with the Angoff method. He attributed this weakness to the difficulty that judges may face in conceptualising the borderline student in the Angoff process (11).

The focus group discussion analysis showed that judges need extensive training if the school will apply standard methods setting process on all its exams. This finding is consistent with the study conducted by Manias and McNamara in which they explore the views of experts in medicine and nursing in setting the standards for an occupational English language screening test for immigrant health professionals. The results of their study revealed the importance of the orientation regarding the process of standard setting (12).

Also among the important points that emerged from the focus group discussion with our judges were the challenges and difficulties of standard setting application. One of the challenges reported by the participants is that the time spent in the process of reaching consensus and determining the cutoff scores of each exam is considered one of the main difficulties and limitations that may hinder the future application of standard setting strategies. These conclusions are matching to some extent those of Senthong, et al. who compared a modified Angoff individual method to the conventional group method. In their study, six judges were divided into two groups matched by teaching experience. Their results revealed that judges preferred to perform their judgements individually to avoid the time spent in the consensus reaching process (13).

Another challenge in the process of determining a pass mark as expressed by the focus group participants in our study was the definition of a borderline student as well as the security issues of the exam, especially when applying standard setting

methods before the application of a test. This is somewhat similar to those of the study that was conducted by Pitoniak et al. at the University of Massachusetts, School of Education. In their study, two standard setting methods were implemented: the Direct Consensus method, and the Item Cluster method. Two separate standard setting panels had participated in the study and each panel completed each method, the judges were asked about the limitations of each method and their answers involved several factors, one of them was the definition of a borderline candidate (14).

Focus group discussion also revealed the recommendation of a multi-disciplinary panel to judge the exam as done in our study. Some participants defended that by assuming the non-content expert will not be biased in determining the level of the borderline group in answering the questions which will decrease the subjectivity of a critical step in standard setting process. The previous finding is similar to the finding of Clauser and his colleagues who conducted an operational standard setting exercise for Step 1 of the USMLE, to examine the impact of rating unfamiliar items on Angoff passing scores. Standard setting judges in their study involved practicing physicians as well as non-physician PhDs (e.g., Professors of Microbiology). Their results concluded that when the non-expert judges were asked to determine the cutoff score of an unfamiliar item, they provide lower passing score than for a familiar item (15).

It was reported that the determined cutoff score by the judges can vary depending on the difficulty of the test (16). This is confirmed in our results as the majority of the judges identified a need to determine the difficulty level of each question to adjust the test difficulty for more reliable test scores. Another study that found matching results is a one that compared the bookmark and Angoff standard setting methods. It concluded that there is a positive correlation between the estimation of the probability of the right answers of the items by the judges and the items' real difficulty (17). Also, a

very old study estimated the relation of item difficulty and inter-item correlation with the test variance and reliability, revealed that in order to maximise the reliability of a test, the variance of the difficulty level distribution should be decreased (18).

Applying the Angoff or Hofstee method, is encouraged by the faculty members as appeared from the focus group discussion as the cut off score determined by either method yielded a different passing score currently used by the FOM.

Some limitations of our study should be acknowledged. Most notably, that our findings cannot be generalised to other assessment methods used in our school such as the modified essay questions. Also, the number of judges who volunteered to share was limited.

CONCLUSION

The current study concluded that whether using the Angoff or Hofstee method, or both, the determined cutoff score is different than the fixed 60% passing score currently used by the FOM/SCU. The study also concluded that the concept of introducing standard setting methods in assessment at FOM/SCU is encouraged by the faculty members as it will improve the validity of assessment scores and may help the school to adjust the test difficulty for further exams.

ACKNOWLEDGEMENTS

The authors would like to thank the participants for their willingness to participate in this study. They would also like to thank all members of the Medical Education Department at FOM/SCU for their continuous help and support. Ethical approval has been granted from the Ethical Committee at the Faculty of Medicine, Suez Canal University.

REFERENCES

1. Cizek G. Setting performance standards: foundations, methods, and innovations. 2nd ed. London: Routledge; 2012.
2. Norcini JJ. Setting standards on educational tests. *Med Educ.* 2003;37(5):464–9. <https://doi.org/10.1046/j.1365-2923.2003.01495.x>
3. Cizek, GJ, Bunch MB. Standard setting: a guide to establishing and evaluating performance standards on tests. Thousand Oaks: SAGE Publications; 2007. <https://doi.org/10.4135/9781412985918>
4. Ben-David MF. AMEE guide no. 18: standard setting in student assessment. *Med Teach.* 2000;22(2):120–30. <https://doi.org/10.1080/01421590078526>
5. De Champlain AF. Standard setting methods in medical education. In: Swanwick T, editor. *Understanding medical education: evidence, theory and practice.* Chichester, West Sussex: John Wiley & Sons Ltd; 2014. p. 305–16.
6. Mortaz Hejri S, Jalili M. Standard setting in medical education: fundamental concepts and emerging challenges. *Med J Islam Repub Iran.* 2014;28:34.
7. Hosny S, El Wazir Y, El Kalioby M, Farouk O, Ghaly M. Role of Suez Canal University, Faculty of Medicine in Egyptian Medical Education Reform. *Health Professions Education.* 2016;2(1):44–50. <https://doi.org/10.1016/j.hpe.2016.01.007>
8. Hussein A, Abdelkhalek N, Hamdy H. Setting and maintaining standards in multiple choice examinations: guide supplement 37.3 – practical application. *Med Teach.* 2010: 610–12. <https://doi.org/10.3109/01421591003730977>
9. Yousef MK, Alshawwa L, Tekian A, Park YS. Challenging the arbitrary cutoff score of 60%: standard setting evidence from preclinical operative dentistry course. *Med Teach.* 2017;24:1–8. <https://doi.org/10.1080/0142159X.2016.1254752>

10. Elfaki OA, Salih KM. Comparison of two standard setting methods in a medical students MCQs exam in internal medicine. *American Journal of Medicine and Medical Sciences*. 2015;5(4):164–7.
11. Mogapi MO. Exploring the contextual limitations of Angoff grading model: the case of Botswana. *International Journal of Scientific Research in Education*. 2012;5(4):278–86.
12. Manias E, McNamara T. Standard setting in specific-purpose language testing: what can a qualitative study add? *Language Testing*. 2016;33(2):235–49. <https://doi.org/10.1177/0265532215608411>
13. Senthong V, Chindaprasirt J, Sawanyawisuth K, Aekphachaisawat N, Chaowattanapanit S, Limpawattana P, Choonhakarn C, Sookprasert A. Group versus modified individual standard-setting on multiple-choice questions with the Angoff method for fourth-year medical students in the internal medicine clerkship. *Adv Med Educ Pract*. 2013;27(4):195–200.
14. Pitoniak MJ, Hambleton RK, Sireci SG. *Advances in standard setting for professional licensure examinations (Laboratory of Psychometric and Evaluative Research Report No. 423)*. Amherst: University of Massachusetts, School of Education; 2002.
15. Clauser JC, Hambleton RK, Baldwin P. The effect of rating unfamiliar items on Angoff passing scores. *Educational and Psychological Measurement*. 2016;77(6):901–16. <https://doi.org/10.1177/0013164416670983>
16. Tekian A, Norcini J. Overcome the 60% passing score and improve the quality of assessment. *GMS Z Med Ausbild*. 2015;32(4):Doc43.
17. Çetin S, Gelbal SA. Comparison of bookmark and Angoff standard setting methods. *Educational Sciences: Theory and Practice*. 2013;13(4):2169–75.
18. Gulliksen H. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*. 1945;10(2):79–91. <https://doi.org/10.1007/BF02288877>