

Comparison of Two Standard Setting Methods in a Medical Students MCQs Exam in Internal Medicine

Omer Abdelgadir Elfaki^{1,*}, Karimeldin M. A. Salih²

¹A/Prof of Internal Medicine and Medical Educationist, Department of Internal Medicine and Department of Medical Education, College of Medicine, King Khalid University

²A/Prof of Pediatrics and Medical Educationist, Department of Pediatrics and Department of Medical Education, College of Medicine, King Khalid University

Abstract Background: There is no single recommended approach to standard setting and many methods exist. These include norm-referenced methods and the criterion-referenced methods. The Angoff method is the most widely used and researched criterion-referenced method of standard setting. Because the outcome of assessments is determined by the standard-setting method used and because different methods of setting standards result in different standards, the choice and process of the method used is of utmost importance. The aim of this study was to compare two standard setting methods, norm-referenced and Angoff methods. **Methods:** Two different methods of standard setting were applied on the raw scores of 106 final medical students on a multiple choice (MCQ) examination in internal medicine. One of these was the modified Angoff method and the other the norm-reference method of standard setting (mean minus 1 SD). The pass rates derived from the two methods were compared. **Results:** The pass rate with the norm-reference method was 88% (93/106) and that by the Angoff method was 39% (41/106). The percentage agreement between Angoff method and norm-reference was 36% (95% CI 36% – 81%). **Conclusions:** The two standard-setting methods used, produced significantly different outcomes. This was demonstrated by the different pass rates.

Keywords Standard, Setting, Angoff, Norm-reference, Assessment

1. Background

A standard is a conceptual boundary on the true-score scale between acceptable and non-acceptable performance [1]. Generally, there are 2 types of standards: absolute or criterion-reference and relative or norm-reference [2-4]. In absolute standards, the performance is measured against a predetermined criterion and therefore is independent of the performance of the group of examinees. On the other hand, in relative standards, the performance of the examinee is compared to others who took the test and hence the pass/fail depends on the performance of that group.

The outcome of any assessment is determined by the standard-setting method used. Standard – setting is defined as "the process of deciding what is good enough" [5]. There is no single method of choice for standard setting and many methods exist [6]. These include norm – reference methods and the criterion reference methods. Each standard-setting method has its advantages and disadvantages, and there is no gold standard. A criterion-referenced standard is generally preferred to a norm-referenced (fixed pass rate) or holistic

model (arbitrary pass mark at, say, 60%) [7]. The Angoff method is the most widely used and researched criterion-referenced method of standard setting and is well supported by evidence [2, 8, 9]. However, it has some disadvantages which include the high time and number of personnel needed [2, 10] and the difficulty inherited in applying the concept of the borderline candidate [2, 11-13]. In this method, a panel of judges examines each MCQ item and estimates the probability that the "minimally competent" or "borderline" candidate would answer the item correctly [8]. Then the scores are discussed in the group and consensus is reached if possible. This stage is not done in the modified approach. Each judge's estimate scores on all items are added up and averaged and the test standard is the average of these means for all the judges. The norm reference methods are easy to use and understand but some examinees will always fail irrespective of their performance and the pass score is not known in advance and can deliberately be influenced by the students [9].

It has been found that different methods of setting standards result in different standards and hence it is argued that the validity of a test is determined as much by the method used to set the standard, as by the test content itself. Downing et al [14] argued that all standards are ultimately policy decisions and that 'there is no gold standard for a passing score. What is important is the process of setting the

* Corresponding author:

elfaroog@gmail.com (Omer Abdelgadir Elfaki)

Published online at <http://journal.sapub.org/ajmms>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

standard. The four important criteria that underpin the process of standard setting are that it is systematic, reproducible, absolute and unbiased.

The objective of this study was to compare two standard setting methods, norm – reference and Angoff methods for MCQs examination.

2. Methods

This study was conducted to answer the research question: What is the agreement of the standards and the pass rates resulting from application of these two standard setting methods on the same MCQs exam?

The scores of 106 final medical students on MCQs paper in their internal medicine course at the College of Medicine of King Khalid University were collected after ethical approval from the research committee. The MCQs paper had 90 one best answer type items with four options. The questions covered topics in general internal medicine and clinical medicine.

Using the two different standard-setting methods-the norm-reference method and the modified Angoff method-two standards were determined. The two methods were compared by the pass rates. In the norm-reference method the standard was determined by calculating the mean of the scores and the Standard Deviation (SD). The standard was set as the adjusted mean minus 1.0 SD.

In the modified Angoff method, a panel of seven judges participated in the standard-setting round. All of them were consultant physicians who participated in teaching the medical students and thus were familiar with the curriculum. A consensus on the definition of a minimally acceptable, that

is borderline candidate, was reached. Based on that definition, each rater judged each MCQ item and the probability that a borderline candidate would answer the item correctly. As the modified Angoff method was used, no group discussion and consensus was established for each item. All ratings were collected and the mean of each rater's total judgment scores on all 90 items was calculated. This mean score represents the score that a minimally competent candidate would obtain according to the rater's judgment.

Statistical analysis was done using SPSS version 20.0. The pass rates were calculated based on the pass scores set for each of the two methods. Comparison of the Angoff and norm reference methods was done by looking at their percentage agreement, which was determined by calculating the percentage of students that gets the same result (pass or fail) by the 2 different methods.

3. Results

The mean of scores and the standard deviation are shown in table 1. The pass rate with the norm-reference method i.e. means minus 1.0 SD was 88% (93/106) and that by the modified Angoff method was 39% (41/106) (table 2). The choice of mean minus 1.0 SD as the pass/fail cut -off score was entirely arbitrary (although this is common practice among educationalists). The two methods of standard – setting, i.e. norm – reference (mean minus 1.0 SD) and modified Angoff method, were compared by looking at the percentage agreement between them (Figure 1). The percentage agreement between the Angoff and the norm – reference method was 36% (95% Confidence Interval = 36% – 81%).

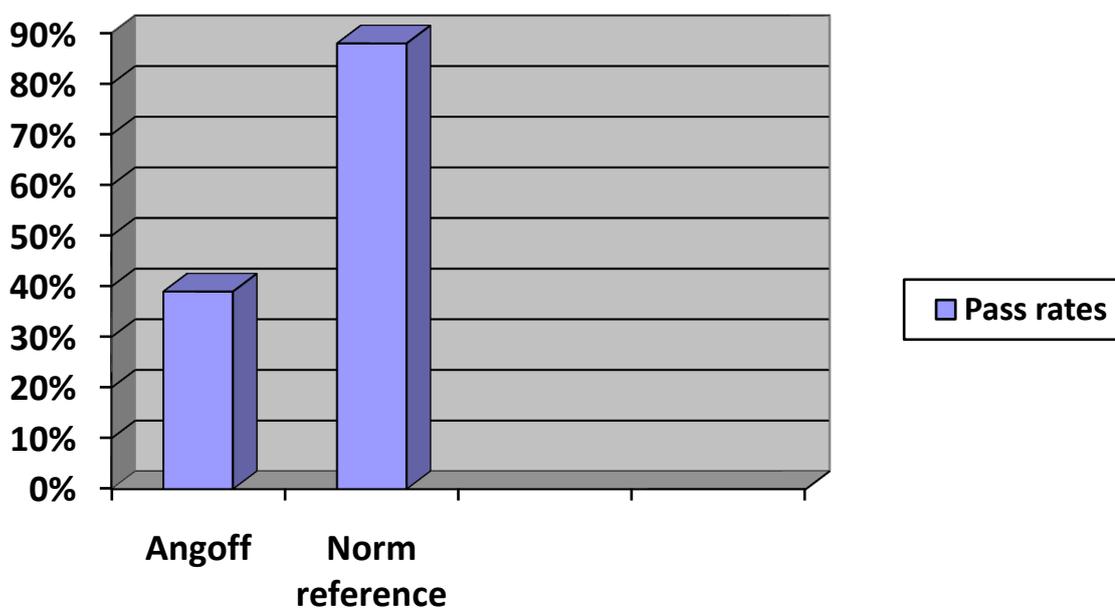


Figure 1. The pass rates of the two methods

Table 1. Descriptive statistics for the two methods of standard setting

Total number of students	106
Total number of test items	90
Mean of the scores	46.5
Standard Deviation	11.96

Table 2. Passing scores and rates for the two methods

Parameter	Angoff method	Norm reference method
Passing score	48	35
Pass rate	39%	88%
Percentage agreement	36%	

4. Discussion

In this study, there was limited agreement between the modified Angoff method and the norm-reference method in determining the outcome of MCQs paper for a batch of medical students. The pass rate was found to be 88% with the norm – reference method, whilst by the Angoff method was 39%. Thus, these two different standard setting methods yielded different standards and the percentage agreement between the two was only 36%. This finding is similar to that reported in previous studies [14-19]. Verhoven et al [20] compared the pass/fail rates derived from the modified Angoff method and the norm – reference method (mean minus 1 SD) and found them to be significantly different with failure rates of 56.5% and 10.1% respectively. Standards were also found to be very different in other studies where different standard setting methods to OSCEs in undergraduate medical examinations had been used [21, 22]. Although it is now fairly well established that different standard setting methods result in different outcomes or passing scores, they can be made credible, defensible and acceptable by ensuring the credibility of judges and using a systematic approach to collect their judgments [14].

The number of judges who participated in the Angoff method of standard setting in this study was relatively small, though it might be acceptable. Some researchers recommended that the number of judges should be between 5 and 10 [23] while others suggested between 5 and 30 judges [24]. Although there is no clear consensus among researchers on the most appropriate number of judges, larger numbers might yield more valid findings.

5. Conclusions

The pass rates generated by the two methods proved to be significantly different. The percentage agreement of the pass rates by the two methods is very low.

REFERENCES

- [1] Kane M. Validating the performance standards associated with passing scores. *Rev Educ Res.* 1994; 64:425–61. doi: 10.2307/1170678.
- [2] Boursicot K, Roberts T. Setting standards in a professional higher education course: Defining the concept of the minimally competent student in performance based assessment at the level of graduation from medical school. *Higher Education Quarterly.* 2006; 60: 74–90. doi: 10.1111/j.1468 2273.2006.00308.
- [3] Searle J. Defining competency-the role of standard setting. *Medical Education.* 2000; 34: 363–366. doi: 10.1046/j.1365-2923.2000.00690.
- [4] Case SM, Swanson DB. *Constructing written test questions for the basic and clinical sciences.* Philadelphia: National Board of Medical Examiners; 1998.
- [5] Cusimano M. Standard-setting in medical education. *Acad Med.* 1996; 71:112–120. doi: 10.1097/00001888-199610000-00062.
- [6] Ben-David, M.F. (2000) AMEE Guide No. 18: Standard setting in student assessment, *Medical Teacher*, 22, pp. 120–130.
- [7] Case, S.M. & Swanson, D.B. (2001) *Constructing Written Test Questions for the Basic and Clinical Sciences*, 3rd edn (Philadelphia, National Board of Medical Examiners).
- [8] Talente G, Haist SA, Wilson JF. A model for setting performance standards for standardised patient examinations. *Evaluation and the Health Professions.* 2003;26:427–446. doi: 10.1177/0163278703258105.
- [9] Norcini JJ. Setting standards on educational tests. *Medical Education.* 2003;37:464–469. doi: 10.1046/j.1365-2923.2003.01495.x.
- [10] Kilminster S, Roberts T. Standard setting for OSCEs: Trial of borderline approach. *Advances in Health Sciences Education.* 2004; 9:201–209. doi: 10.1023/B:AHSE.0000038208.06099.9a.
- [11] Impara JC. Setting standards using Angoff's method: Does the method meet the standard? *Invited address to Division D of the Midwestern Educational Research Association, Chicago.* 1997.
- [12] National Research Council. Setting reasonable and useful performance standards. In: Pelligrino JW, Jones LR, Mitchellw KJ, editor. *Grading the Nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* Washington, DC: National Academy Press; 1999. pp. 164–184.
- [13] Impara JC, Plake BS. Teachers' ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method. *Journal of Educational Measurement.* 1998; 35:69–81. doi: 10.1111/j.1745-3984.1998.tb00528.
- [14] Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine.* 2006;18:50–57. doi: 10.1207/s15328015t1m1801_11.

- [15] Verhoeven BH, Van der Steeg AFW, Scherpbier AJJA, Muijtjens AMM, Verwijnen GM, van der Vleuten CPM. Reliability and credibility of an Angoff standard setting procedure in progress testing using recent graduates as judges. *Medical education*. 1999; 33: 832–837. doi: 10.1046/j.1365-2923.1999.00487.
- [16] Humphry-Murto S, MacFayden JC. Standard setting: A comparison of case author and modified borderline group methods in a small scale OSCE. *Academic Medicine*. 2002;77:729–732.
- [17] Kaufman DM, Mann KV, Muijtjens AMM, van der Vleuten CPM. A comparison of standard setting procedures for an OSCE in undergraduate medical education. *Academic Medicine*. 2000;75:267–271.
- [18] Fehrmann ML, Woehr DJ, Arthur W. The Angoff cutoff score method: The impact of frame-of-reference training. *Educational and Psychological Measurement*. 1991; 51: 857–872.
- [19] Impara JC, Plake BS. Standard setting: An alternative approach. *Journal of Educational Measurement*. 1997; 34: 353–366. doi: 10.1111/j.1745-3984.1997.tb00523.
- [20] Verhoeven BH, Verwijnen GM, Muijtjens AMM, Scherpbier AJJA, van der Vleuten CPM. Panel expertise for an Angoff standard setting procedure in progress testing: item writers compared to recently graduated students. *Medical Education*. 2002; 36: 860–867. doi: 10.1046/j.1365-2923.2002.01301.
- [21] Wayne DB, Fudala MJ, Butter J, Siddall VJ, Feinglass J, Wade LD, McGaghie WC. Comparison of two standard setting methods for advanced cardiac life support training. *Academic Medicine*. 2005;80:S63–S66. doi: 10.1097/00001888-200510001-00018.
- [22] Kaufman DM, Mann KV, Muijtjens AMM, van der Vleuten CPM. A comparison of standard setting procedures for an OSCE in undergraduate medical education. *Academic Medicine*. 2000;75:267–271.
- [23] Norcini JJ, Shea JA. The credibility and comparability of standards. *Applied Measurement in Education*. 1997; 10: 39–59. doi: 10.1207/s15324818ame1001_3.
- [24] Zieky MJ, Livingston SA. *Manual for setting standards on the basic skills assessment tests*. Princeton, NJ: Educational Testing Service; 1977.