

STUDY OF STANDARD SETTING IN CONSTRUCTED RESPONSE TYPE WRITTEN EXAMINATION

Dhaval Parmar¹, Chinmay Shah², Rasmika Parmar³

¹ Department of Forensic Medicine, Government Medical College, Bhavnagar, Gujarat, India

² Department of Physiology, Government Medical College, Bhavnagar, Gujarat, India

³ Department of Microbiology, Government Medical College, Bhavnagar, Gujarat, India

Correspondence to: Chinmay Shah (cjshah79@yahoo.co.in)

DOI: 10.5455/ijmsph.2014.170620142

Received Date: 18.05.2013

Accepted Date: 17.06.2014

ABSTRACT

Background: Standard setting is the process of determining what the minimum requirements are to be deemed knowledgeable, or competent to practice. In it, the items prepared for the student assessment is standardized before and/or after the student assessment process. The main objective of standard setting is to determine the cut-off score, or pass mark, of an item or a test.

Aims & Objective: Objective of the present study was to look at standard setting of summative assessment.

Materials and Methods: Present retrospective study of previous 5 years theory papers used Modified Angoff & Hofstee method.

Results: It is observed that Cut-off score of all 5 years by Modified Angoff method was set below 50% while same was set above 50% when derived using Hofstee method. Thus, the Hofstee method consistently produced more stringent passing scores than the Modified Angoff method. Study showed statistically significant difference between averages of cut-off scores obtained by the two methods.

Conclusion: Study provided us with an insight regarding standard setting i.e. standard setting must be done in concert with the test development process and not be treated as a last or separate step independent of the process. It also puts emphasis on the need of variability in cut-off score in different summative examination depending on difficulty level of the paper.

Key Words: Standard Setting; Constructed Response; Written Examination; Modified Angoff Method; Hofstee Method

Introduction

Globalization, mobility of doctors and the rising number of medical institutions make it imperative to have comparable standards in medical teaching learning and assessment.^[1] The goal of licensure and certification testing in medical school is to determine a standard or set a boundary between those who perform well enough & those who do not to hold the credential.^[2] To be fully competent, scoring 100% of marks is unrealistic in most examinations as that will lead to few licensures & cause serious harm to public service.^[3] In reality, examinees show a variety of performances ranging from non-competent to reasonably competent.^[4] Our aim must surely be to strive to eliminate false positives (passing the incompetent) and false negatives (failing the competent)^[5] from our assessment system.

When competency is measured via examinations, it is crucial to ensure that content of the test and its difficulty level is appropriate for the decisions to be made based on the test results.^[6] It is logical that no two tests or examinations have the same difficulty. Small, minor variances in the level of difficulty are acceptable between papers, but we have all come across 'horribly difficult' or 'ridiculously easy' paper^[7] for different summative examinations of same subject. Hence, Medical schools

should not employ fixed pass marks – that is, pass marks which are the same every year –since these do not reflect variations in the difficulty of examinations over time and therefore could result in individuals of identical ability passing in one year and failing in another.^[8] So, any standard setting process used must be able to absorb these variances and adapt the pass mark accordingly.^[9] It must yield cut-off score that permit dichotomous categorization, is sensitive to the examinee's performance, statistically sound and practicable, credible, easy to implement and interpret for the layman.^[10]

Standard setting is the process of determining what the minimum requirements are to be deemed knowledgeable or competent to practice.^[7] In it, the items prepared for the student assessment are standardized before and/or after the student assessment process. The main objective of standard setting is to determine the cut-off score, or pass mark, of an item or a test.^[11] There are different standard setting methods practiced around the world.^[12]

- Fixed or Holistic Standard Setting
- Norm-Referenced Standard Setting
- Criterion-Referenced Standard Setting

In the fixed or holistic standard setting method, the cut-

off score is determined a priori by the university and even the regulatory bodies.^[12,13] The most common or preferred pass mark in health professions' education is 50 percent of the total marks and is set before the assessment.

The norm-referenced standard setting method is used to identify certain number of best examinee from the group.^[12] It is applied after the test and standardized for the particular cohort of students that take the test. There are practices of using percentile, Z-score, T-score, or Stanine scores.^[14] The main limitation of this method is that it is based on the performance of the cohort of student, which will be different from cohort to cohort and results are not comparable across cohorts based on their performance.^[11]

To compensate the fallacies of the fixed and norm referenced standard setting methods, criterion-reference standard setting methods have been formulated.^[12-14] Criterion referenced assessments assess students against specified standards of achievement^[15] and differentiate the competent from the non-competent.^[16] The criterion standard tells about the degree of competence of a particular examinee independent of the performance of other members in the group.^[12,17] Criterion is synonymous with standard or cut-off.^[18]

Currently, in our Institutional settings, summative assessment in UG Forensic Medicine is carried out through written test, which comprises of constructed responses. Same is scored in holistic way for holistic fixed percentage pass mark which is conventionally set at 50% of total marks as per the guideline of Medical Council of India, and is set before the assessment which is the least defensible. Kane proposed that evidence should be supplied in order to defend the performance standards. Thus, our assessments require the evidence of validity to be meaningful.

Present retrospective study was conducted to analyse standard setting of previous 5 years summative assessment theory papers using Modified Angoff & Hofstee methods. Study was also focussed on comparison of both the methods & evaluating their intra-rater reliability. Study would necessitate introspection of our assessment system as well as provide an introduction to methods of setting standards.

Materials and Methods

Present study was undertaken at Forensic Medicine

Dept. of Govt. Medical College Bhavnagar after obtaining permission from the Head of Institute. Panellists selected from Forensic Medicine Dept. of Govt. Medical College Bhavnagar fulfilled all the requirements^[20] as below which can provide strength to our study.

- Experts in the related field of examination
- Familiar with the examination methods
- Good problem solvers
- Familiar with level of candidates
- Interested in education (teachers).

Same panellists were used for both methods of standard setting. All panellists were explained in detail regarding the aims of study and procedure to be followed for standard setting using Modified Angoff & Hofstee method. In our study, while adopting Modified Angoff & Hofstee method of standard setting, we defined 'the borderline examinee' as one who has an exactly 50:50 probability of passing or failing the test. Boursicot and Roberts^[21] referred to the idea of a borderline or minimally competent candidate as a 'nebulous concept' and research has shown that often judges find it difficult to accurately define and understand hypothetical borderline student.^[19,22] We also avoided a group discussion on estimated cut-off scores among judges and no 'reality check' (giving information on actual performance data) was provided.^[23] This is known as the 'purist' approach – not presenting judges with actual performance data.

Modified Angoff Standard-Setting Procedure

Group of expert Judges defined & discussed the characteristics of the border line candidate who is minimally qualified to pass the examination. Panellists then reviewed previous 5 years summative assessment CR (Constructed Response) type-written test - theory papers and estimated the probability to answer each item correctly by this marginal candidate. For each year, estimated passing score is derived for each panellist by summing these item probabilities and final passing score (cut-off score) is calculated by taking the average passing score across panellists.

Hofstee Standard-Setting Procedure

Panellists were asked to review previous 5 years summative assessment CR type - written test - theory papers & give their global impressions of what the minimum and maximum failure rates should be as well as what the minimum and maximum percent correct scores should be for the examination, ensuring that each

judge fully understands each question & its implications. These rates and percent correct scores are projected onto the cumulative score graph to derive a passing score. Average of passing score across panellists is considered final passing score (cut-off score).

Cut-off score derived through both methods are than compared year wise as well as cumulatively. Paired T test was applied to find out statistical significance between these two methods. Intra rater reliability in Modified Angoff method was determined using Intra class correlation coefficient (ICC).

Results

Table 1 shows limited agreement between cut-off score ascertained by same judge using Modified Angoff method and the Hofstee method. It is observed that Cut-off score of all 5 years by Modified Angoff method was set below 50% while the same was set above 50% when derived using Hofstee method (Table 1). Thus, Hofstee method consistently produced more stringent passing scores than the Modified Angoff method.

Table 2 shows statistically significant difference between averages of cut-off scores obtained by both methods. In the year 2011, cut-off score was 32% by Modified Angoff method while it was 51% by Hofstee method - with a significant difference of 19%.

Table-1: Year wise cut-off score by different judges

Year	2008		2009		2010		2011		2012	
Judge	1*	2**	1*	2**	1*	2**	1*	2**	1*	2**
A	50	48	52	50	47	47	43	45	51	47
B	52	42	54	46	55	41	44	41	48	42
C	62	28	65	35	55	29	62	22	60	24
D	50	32	52	41	52	41	51	29	57	38
E	62	42	60	45	61	46	57	37	60	41
F	50	26	48	33	48	25	50	15	48	26
p = 0.0106 p = 0.0172 p = 0.011 p = 0.0346 p = 0.0138										

* 1: Hofstee standard-setting procedure; ** 2: Modified Angoff standard-setting procedure

Table-2: Modified Angoff and Hofstee cut-off score & ICC for Year 2008 to 2012

Year	Hofstee	ICC	M. Angoff	ICC
2008	54	0.34	36	0.85
2009	55		41	0.84
2010	53		38	0.80
2011	51		32	0.78
2012	54		38	0.78
p = 0.00006				

Discussion

There are no true, objectives or "golden" performance standards for any assessment and the performance standards can only be set in a more or less trustworthy

way. Medical Council of India has suggested 50% cut-off score for all summative assessments in Graduate Medical Education Regulations 1997. Present study analyse previous 5 years summative assessment theory papers of UG Forensic Medicine which is CR type-written test. It is an attempt to find out if cut-off score of 50% is appropriate or there is a need to establish credible, defensible, acceptable standard score.

There is no perfect method to determine cut-off score on a test.^[3] And none is agreed upon as the best method.^[24,25] Several standard setting procedures have been developed to face the challenge of making it defensible and controllable.^[5] Significant advancement in standard setting technology has allowed reasonable confidence in the resultant standard.^[26] Nevertheless, the cut-off score is not consistent among the methods. Cut-off score calculated by both the methods with comparison are show in Table 1.

This finding is similar to that reported in previous studies.^[19] Verhoven et al^[27] compared the pass/fail rates on an Individual Statement Questions Examination used in undergraduate medical assessment derived from the modified Angoff method and the norm - reference method (mean minus 1 SD) and found them to be significantly different - failure rates of 56.5% and 10.1% respectively. Similar difference in the cut-off score of both methods was observed in present study which was statistically significant. The higher cut-off score derived in Hofstee method may be because of higher expectations by experts from novices particularly when judgement is holistic rather than item judgment.

Significant difference of cut-off score (Table 2) can have grave effect on students' carrier as well as performance of institute. Hence the best suitable method of standard setting should be employed.

Glass^[18] viewed all standard-setting methods that involved judges making arbitrary decisions as fundamentally flawed. So, it may be argued that the difference resulted due to subjectivity of judges & all well-known rater errors but in present study same panel of judges were recruited to carry out both methods of the study. Also, more than 0.78 intra-rater reliabilities in Modified Angoff method points towards fair, reasonable & careful deliberation of judgements. Norcini stated that although all standards are judgemental, the credibility of each standard varies depending on who sets the standards and the methods they use.^[28,29]

Modified Angoff method of standard setting is more objective and has good inter-rater reliability (0.78-0.85) compare to ICC of Hofstee method (0.34), Although the proportion of candidates that can pass or fail will vary considerably with the method of standard setting used, the Modified Angoff method has self-evident face validity as it replaces grossly arbitrary methods with a reasoned, standardised method that is open to inquiry.^[30]

To assess the validity of standard setting in present study, we asked judges to fill up standard setting feedback questionnaire which suggests that purpose of test, characteristic of borderline examinee & rating task were very clear for them even though they found rating of CR paper for standard setting a difficult task. In opinion of panellist final passing score were appropriate for the examinee & they were confident about appropriateness of cut-off score.

Study provided us with an insight regarding standard setting i.e. standard setting must be done in concert with the test development process and not be treated as a last or separate step independent of the process^[6] and it puts emphasis on the need of variability in cut-off score in different summative examination depending on difficulty level of paper.

Downing et al^[19] argued that all standards are ultimately policy decisions and that 'there is no gold standard for a passing score.' The key is the process of setting the standard. The 4 Key principles that underpin the process of standard setting are that it is systematic, reproducible, absolute & unbiased. Study also showed many educational benefits for future.^[20]

Faculty Development: Standard setting procedures can be employed as a form of faculty development. Faculty experience first-hand information of candidates' performance on the task and are able to compare this with their own expectations relating to the competence.

Quality control of test materials: The process of exposing faculty to test materials, scoring policy and profiles of scored performance constitute a scrutinised quality control procedure. Panellists in the process of reviewing test materials identify inappropriate items, which are either ambiguous, or irrelevant.

As variability in cut-off score is a policy decision, standard setting can be made an integral part of planning for test development. Vigilant faculty can set appropriate

difficulty level of test to lead to appropriate, meaningful & psychometrically sound dichotomous categorization of students at cut-off score of 50%.

Conclusion

This study demonstrates that rigorous and systematic standard-setting methods should be used to derive minimum passing scores for undergraduate students. Appropriate and transparent standard setting is a critical element of good assessment and educational practice. We need to have systematic approach towards implementing explicable, defensible and stable standard setting methods to benefit all stakeholders in health sciences education.

The more standard setting procedures are applied to a variety of tests, the more the practice of high quality testing will be enhanced and the higher will be the confidence in the testing of professional competencies. Much work is still needed to investigate the statistical characteristics of the standard setting method in order to establish its limits and strengths, and its use in undergraduate medical examinations.

References

1. Lilley PM, Harden RM. Standards and medical education. *Med Teach* 2003;25:349-51.
2. An Information Brief from ACT. Available from: URL: <http://www.act.org/workforce/cewd.html>
3. Kane M. Validating the performance standards associated with passing score. *Rev Educ Res* 1994;64:425-61.
4. Barman A. Standard Setting in Student Assessment: Is a Defensible Method Yet to Come? *Ann Acad Med Singapore* 2008;37:957-63.
5. Cusimano MD. Standard setting in medical education. *Acad Med* 1996;71:S112-20.
6. Isaac IB. Standard Setting: What Is It? Why Is It Important? *R & D Connections*; 2008:7.
7. Schoeman, Scarpa. Setting standards in health sciences education – a wake-up call. *African Journal of Health Professions Education* 2011;3:2.
8. Assessment in undergraduate medical education: Supplementary advice to Tomorrow's Doctors. General Medical Council 2011. p. 1-23.
9. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach* 2008;30:836-45.
10. Berk RA. A consumer's guide to setting performance standard on criterion-referenced tests. *Rev Educ Res* 1986;56:137-72.
11. Bhandary S. Standard Setting in Health Professions Education. *Kathmandu Univ Med J (KUMJ)* 2011;9:3-4.
12. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;37:464-9.
13. Kaufman DM, Mann KV, Muijtjens AMM, van der Vlueten CPM. A Comparison of Standard-setting Procedures for and OSCE in Undergraduate Medical Education. *Academic Medicine* 2000;75:267 – 71 .
14. George S, Haque MS, Oyebo F. Standard setting: comparison of two methods. *BMC Med Educ* 2006;6:46.
15. Dunn L, Parry S, Morgan C. Seeking quality in criterion reference assessment. Presented at: Learning Communities and assessment

- Cultures Conference organized by the EARLI Special Interest Group on Assessment and Evaluation. University of Northumbria, 28-30 August 2002. Available from: URL: <http://www.leeds.ac.uk/educol/documents/00002257.htm>.
16. Harden RM. Ten questions to ask when planning a course or curriculum. *Med Educ* 1986;20:356-65.
 17. Glaser R. Instructional technology and the measurement of learning outcomes: Some questions. *Am Psychol* 1963;18:519-21.
 18. Glass GV. Standard and criteria. *J EducMeas* 1978;15:237-61.
 19. Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med* 2006;18:50-7.
 20. Ben-David MF. AMEE Guide No. 18: Standard setting in student assessment. AMEE Guide. *Medical Teacher* 2000;22: 120-30.
 21. 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.
 22. Wayne DB, Fudala MJ, Butter J, Siddall VJ, Feinglass J, Wade LD, et al. Comparison of two standard setting methods for advanced cardiac life support training. *Academic Medicine* 2005;80:S63-6.
 23. Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Medical Education* 2003;37:132-9.
 24. Linn RL. Performance standards: Utility for different uses of assessments. *Educ Pol Anal Arch* 2003;11:1-20.
 25. Boursicot KA, Roberts TE, Pell G. Standard setting for clinical competence at graduation From medical school: a comparison of passing scores across five medical schools. *Adv Health Sci Educ Theory Pract* 2006;11:173-83.
 26. Cizek GJ. Reconsidering standards and criteria. *J EducMeas* 1993;30: 93-106.
 27. 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;36:860-7.
 28. Norcini JJ, Shea JA. The credibility and comparability of standards. *Applied Measurement in Education* 1997;10:39-59.
 29. Norcini JJ, Guille RA. Combining tests and setting standards. *International Handbook of Research in Medical Education* 2002. p. 811-34.
 30. George S, Haque MS, Oyebo F. Standard setting: Comparison of two methods. *BMC Med Educ* 2006;6:46.

Cite this article as: Parmar D, Shah C, Parmar R. Study of standard setting in constructed response type written examination. *Int J Med Sci Public Health* 2014;3:1046-1050.

Source of Support: Nil

Conflict of interest: None declared

IJMSPH