Standard setting of objective structured practical examination by modified Angoff method: A pilot study

M. GANESH KAMATH, VINOD PALLATH, K. RAMNARAYAN, ASHA KAMATH, SHARMILA TORKE, JAMES GONSALVES

ABSTRACT

Background. The undergraduate curriculum at our institution is divided system-wise into four blocks, each block ending with theory and objective structured practical examination (OSPE). The OSPE in Physiology consists of 12 stations, and a conventional minimum score to qualify is 50%. We aimed to incorporate standard setting using the modified Angoff method in OSPE to differentiate the competent from the non-competent student and to explore the possibility of introducing standard setting in Physiology OSPE at our institution.

Methods. Experts rated the OSPE using the modified Angoff method to obtain the standard set cut-off in two of the four blocks. We assessed the OSPE marks of 110 first year medical students. Chi-square test was used to compare the number of students who scored less than standard set cut-off and conventional cut-off; correlation coefficient was used to assess the relation between OSPE and theory marks in both blocks. Feedback was obtained from the experts.

Results. The standard set was 62% and 67% for blocks II and III, respectively. The use of standard set cut-off resulted in 16.3% ($n=18$) and 22.7% ($n=25$) students being declared unsuccessful in blocks II and III, respectively. Comparison between the number, who scored less than standard set and conventional cut-off was statistically significant ($p=0.001$). The correlation coefficient was 0.65 ($p=0.003$) and 0.52 ($p<0.001$) in blocks II and III, respectively. The experts welcomed the idea of standard setting.

Conclusion. Standard setting helped in differentiating the competent from the non-competent student, indicating that standard setting enhances the quality of OSPE as an assessment tool.

Natl Med J India 2016;29:160–2

INTRODUCTION

Standard setting helps to differentiate students with various scores and categorize them by determining an appropriate cut-off score instead of the conventional cut-off of 50%.[1] Various available standard setting methods can be incorporated into the assessment.[2,3] Standard setting can be broadly classified into two methods—test-centred (Angoff, Ebel, Jaeger and Nedelsky methods) and examinee-centred (borderline-group, contrasting groups and Hofstee's methods). Test-centred methods are those where the test items are reviewed by experts to pass a judgement on the test items as 'just an adequate' level of performance. The experts identify an actual group (not a hypothetical) and make judgements on the examinees.[4]

In the Angoff method, the experts make judgements on how hypothetical borderline candidates will perform on each item/question and obtain a cut-off score. The modified Angoff method has additional steps wherein the experts are provided with data on actual performance to understand the difficulty of test items and relook at the scores provided. The Ebel method requires the experts to categorize the test items based on the difficulty level and then decide on the proportion of items in each of the categories the hypothetical borderline students can answer appropriately. The number of items and the number answerable by the hypothetical group in each category is multiplied to get a standard score. The Nedelsky method, which is specifically designed for multiple-choice questions, requires the experts to decide how many distractors to have in a test item a minimally competent student will be able to identify as inappropriate/wrong. In the Jaeger method, the experts rate each item, based on whether an examinee can answer correctly or incorrectly. This method is followed by another group of experts. After which, each of the experts can relook their assessment based on the others' assessments. This method emphasizes more on passing examinees than hypothetical borderline students.[4]

In the borderline group method, the experts identify an actual borderline group instead of a hypothetical group. The scores given by the experts are used to obtain a 'median score', which is then used as the passing score. This contrasts with the group approach in that it requires the experts to divide the examinees into those who are non-competent and competent. This assessment is based on the examinee characteristics and then scored. A graph is plotted for the score distribution for non-competent and competent students. The point of intersection between these two groups is chosen as the passing score. In the Hofstee method, the experts review the test constituents, who in turn provide four values: (i) the minimum failure rate; (ii) the maximum failure rate; (iii) the minimum passing score; and (iv) the maximum passing score. For each of the values, a graph is plotted to obtain a median value.[4] The modified Angoff method has earlier been used to assess the clinical competency of students in an undergraduate curriculum.[5]

The curriculum at our institution for the first year MBBS students is divided system-wise into four blocks (Block I: basic concepts, nerve-muscle, blood; Block II: cardiovascular, respiratory, gastrointestinal system; Block III: endocrine, renal, reproductive system; and Block IV: central nervous system and special senses) for three subjects (anatomy, physiology and biochemistry). Each block has a duration of 10 weeks. At the end of each block, a theory examination and objective structured practical examination (OSPE) are conducted. The OSPE consists of 12 stations and the pass percentage is 50%.

Melaka Manipal Medical College (MMMC), Manipal University, Manipal 576104, Karnataka, India
M. GANESH KAMATH, SHARMILA TORKE, JAMES GONSALVES Department of Physiology
VINOD PALLATH Department of Microbiology
K. RAMNARAYAN Department of Pathology

Kasturba Medical College, Manipal University, Manipal, Karnataka, India
ASHA KAMATH Department of Community Medicine

Correspondence to M. GANESH KAMATH;
ganesh.kamath@manipal.edu, *kamath18@gmail.com*

© *The National Medical Journal of India 2016*

We felt that students at our institution perform better in OSPE than theory during the block examinations. Ensuring the validity of the assessment is essential for maintaining or improving the standard and the products of the course. Although not a fool-proof method, standard setting helps in enhancing the quality of assessment. Through the introduction of standard setting, we intended to achieve appropriate interpretation of the OSPE scores as reliable indicators of minimal acceptable competency required in Physiology, making the OSPE examination a reliable assessment tool to differentiate the competent from the non-competent student.[6,7] The project was also undertaken to explore the possibility of introducing standard setting in the Physiology OSPE at our institution. The overall aim of this study was to show the usefulness of standard setting in enhancing the validity of OSPE.

METHODS

The study was conducted during the OSPE assessment of 110 first year MBBS students of the March 2009 batch of our institution. Faculty members involved in this study were explained their role as expert panelists for standard setting by the modified Angoff method. The expert panelists decided the characteristics of a borderline student. The faculty members involved in block examinations prepared the questions for 12 OSPE stations and the expert panelists rated each question in the OSPE station based on the difficulty level for a borderline student. We defined the competency of a student as the capacity to use the skill and/or knowledge to perform and/or answer the OSPE question(s) appropriately.[8] The achievement of competency in performing or interpreting each OSPE question is better represented by the cut-off score than the conventional 50% cut-off score. Two examples of OSPE stations with the checklist are given below:

Example 1

A strip of an ECG recording (lead II) from a normal person was placed at the OSPE station. The OSPE questions were: (i) calculate the heart rate from the given ECG; (ii) calculate PR interval from the given strip of ECG; (iii) mention the normal range of the PR interval; (iv) if the PR interval is longer than normal, what does it infer?

The checklist/answer key was: (i) heart rate calculated by the formula 60/one cardiac cycle time (paper speed at 25 mm/seconds). The cardiac cycle time calculated as: 1 mm box on the time scale (X-axis) equals to 0.04 second, so 0.04 multiplied by the number of small boxes between the RR intervals; 20 small boxes between the RR interval=20×0.04=0.8, so 60/0.8=75 beats/minute); (ii) four small boxes×0.04=0.16 seconds; (iii) 0.12 to 0.20 seconds; (iv) a PR interval greater than 0.20 second indicates an AV/ conduction block.

Example 2

A peak flow meter was placed to assess the peak flow rate. The OSPE questions were: (i) identify the given instrument; (ii) perform the test on yourself to obtain the reading(s) and note down the value(s) obtained; (iii) mention the normal range obtained in healthy adults; (iv) name one clinical condition where the above test reading is decreased.

The checklist/answer key: (i) peak flow meter; (ii) three readings should be obtained, and the maximum among the three should be considered as the value with litre/minute as the unit; (iii) 400–600 litre/minute; (iv) bronchial asthma.

The experts discussed the ratings provided by each of them. If ratings required adjustment, the experts were asked to give second ratings.

Based on the ratings and consensus among the experts, a standard set cut-off score was obtained. A minimum of six to a maximum of eight faculty members from the Department of Physiology were involved as experts.[9] The checklist for each OSPE station was provided to the experts. The checklists were developed by faculty members of the department and were validated (content and construct validity). The OSPE checklists had been used in previous examinations and feedback was taken from examiners including the external examiners during university examinations. The checklists were used to ensure objectivity while correcting answers for the OSPE stations. The results obtained after the OSPE were analysed (using a checklist) for pass or fail using a conventional cut-off (50%) and the standard set cut-off. The same method was followed for two block (II, III) examinations.

The chi-square test was used to compare the number of students who scored less than standard set pass and conventional cut-off in blocks II and III. The Mann–Whitney test was used to compare the marks obtained in OSPE and theory among those who scored ≥50% or <50%, using the OSPE standard set cut-off in blocks II and III. The data were expressed as median (interquartile range) due to skewness of data.

The correlation coefficient was determined to assess the relation between OSPE and theory marks, obtained in blocks II and III. A value of $p<0.05$ was considered statistically significant.

Qualitative feedback regarding the standard setting process was obtained from the experts through a semi-structured interview. Statistical analysis was done using SPSS 15.0 (South Asia, Bengaluru). The study was approved by our institutional research committee and informed consent was obtained from all experts/ faculty who participated in the study.

RESULTS

The standard set cut-off obtained by the experts using the modified Angoff method was 62% for block II and 67% for block III. All the students passed with the conventional cut-off of 50%. The use of standard set cut-off resulted in 16.3% (18 of 110) and 22.7% (25 of 110) not achieving the standard set cut-off in blocks II and III, respectively. The mean (SD) of the OSPE marks obtained by students were 73 (12) for block II and 80 (11) for block III.

In block II, the number of those who scored ≥50% and ≥62% were 3 (2.7%) and 18 (16.3%), respectively ($p<0.001$). Similarly in block III, 25 (22.7%) and 3 (2.7%) scored ≥67% and ≥50%, respectively. This difference was significant ($p<0.001$).

A comparison between the OSPE and theory marks in block II in those with ≥50% and >50% (i.e. <62% standard set cut-off) showed no statistically significant difference between those with ≥50% OSPE marks compared with the theory marks ($p=0.70$), whereas those with >50% had a statistically significant difference ($p=0.001$), i.e. the OSPE marks scored by the students were higher than the theory marks. Similar results were obtained in block III, where the OSPE marks were higher compared to the theory marks especially in those with >50% ($p=0.001$) (Table I).

The correlation coefficient determined to understand the concurrent nature of the theory and OSPE marks and the correlation coefficients between theory and OSPE marks of the students who did not achieve the standard set criteria for blocks II and III were 0.65 ($p=0.003$) and 0.52 ($p<0.001$), respectively.

The qualitative feedback obtained from the experts showed that they were receptive to the idea of standard setting, but had their reservations on how borderline candidates were identified.

TABLE I. Number of those who scored less than standard set pass and conventional cut-off and marks obtained in objective structured practical examination (OSPE) and theory

Block	Standard set cut-off (%)	Students (*n*) scoring less than standard set cut-off	Conventional cut-off in OSPE		Median (IQR) marks		p value
			Cut-off (%)	Number (%) of students	OSPE	Theory	
II	≤62	18	≤50	03 (16.6)	20.0 (35.0–49.0)	19.5 (14.0–25.0)	0.70
			≥50	15 (83.3)	56.2 (52.5–60.0)	47.0 (39.0–52.0)	0.001*
III	≤67	25	≤50	03 (12.0)	48.0 (44.0–50.0)	41.0 (23.0–45.0)	0.20
			≥50	22 (88.0)	64.0 (61.5–66.0)	54.5 (48.5–60.5)	0.001*

* p<0.05 considered statistically significant IQR interquartile range

They also felt that there was much subjectivity. The feedback given by experts was:

'…I am unclear on how to define a borderline candidate since one student may be efficient in one system (example: cardiovascular system) but may not be efficient in another system (example: endocrinology).'

'…I feel that each question needs to be rated on the basis of the learning objectives, and the difficulty and discrimination index we do not need to base our scores on a borderline candidate as it brings in some subjectivity.'

'…Good to standard set the OSPE, as students are able to score marks comparatively more easily to the theory examination.'

Among the eight experts, six (75%) agreed that the use of standard setting would enhance OSPE as an assessment tool to test the skills of our first year undergraduate medical students.

DISCUSSION

OSPE is useful in ensuring objectivity and uniformity in the assessment of clinical competencies and skills of students.[10,11] OSPE was introduced in our institution a decade ago, particularly in the Department of Physiology. Students provided a positive feedback for OSPE as compared to the traditional practical examination in our institution and in other medical colleges.[10,11] The modified Angoff method is considered a reliable tool for standard setting for OSCE.[12] OSPE is a variant of OSCE, so we used the modified Angoff method to standard set the OSPE in Physiology.[13]

We also compared the OSPE and theory marks obtained in two blocks of our curriculum. Another study being conducted at the same time introduced structuring the theory question paper using SOLO (structure of observed learning outcomes) taxonomy.[14] This standardized the theory paper ensuring a comparative difficulty level of different examinations within the department. This study also aimed to constructively align the teaching, learning and assessment through this process. This justified the use of correlating OSPE with theory examination as a measure of concurrent validity. We noticed that students who scored lower than the standard set cut-off in OSPE had lower marks in theory compared to the OSPE. The results also showed that students scored higher in OSPE in every block. This suggests that standard setting improves the validity of OSPE as an assessment tool.

The study also led to improved awareness among faculty members about the need to check the quality of OSPE. Faculty members were receptive to the idea of standard setting. However, they were not convinced about the process of defining a borderline student, since it was subjective. Most faculty members also felt that instead of changing the OSPE cut-off, it would be appropriate to look into the difficulty level of each OSPE station.

Limitations

First, though each expert came to a consensus on defining a borderline candidate; there was probably a bias due to a lack of clarity among the experts during each block and also between the blocks. This limitation was noticed when feedback was obtained from each expert. Second, standard setting was checked only for two blocks and not for two other blocks. Third, the number of experts varied between the blocks as only those who had more than two years of experience in teaching our students were selected as experts.

Conclusion

Standard setting helped in differentiating the competent from the non-competent student, indicating that standard setting enhances the quality of OSPE, as an assessment tool.

ACKNOWLEDGEMENTS

We would like to thank the faculty at the Department of Physiology, MMMC, and especially those who consented to take part in this study. We would also like to thank the FAIMER faculty and fellows (GSMC-FAIMER Regional Institute, based at Seth G.S. Medical College, Mumbai, India) for their critical analysis and suggestions.

Conflict of interest: None declared.

REFERENCES

1 Cizek GJ. *Setting performance standards: Concepts, methods, and perspectives.* New Jersey:Lawrence Erlbaum; 2001.
2 Cizek GJ, Bunch MB. Standard setting: A guide to establishing and evaluating performance standards for tests. Thousand Oaks, CA:Sage; 2007.
3 Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten CP. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med* 2000;**75:**267–71.
4 Ben-David MF. AMEE Guide No. 18: Standard setting in student assessment. *Med Teach* 2000;**22:**120–30.
5 George S, Haque MS, Oyebode F. Standard setting: Comparison of two methods. *BMC Med Educ* 2006;**6:**46.
6 McKinley DW, Norcini JJ. How to set standards on performance-based examinations: AMEE Guide No. 85. *Med Teach* 2013;**36:**97–110.
7 Searle J. Defining competency—the role of standard setting. *Med Educ* 2000;**34:**363–6.
8 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;**287:**226–35.
9 Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;**37:**464–9.
10 Abraham RR, Raghavendra R, Surekha K, Asha K. A trial of the objective structured practical examination in physiology at Melaka Manipal Medical College, India. *Adv Physiol Educ* 2009;**33:**21–3.
11 Wani PD, Dalvi VS. Objective structured practical examination vs. traditional clinical examination in human physiology: Student's perception. *Int J Med Sci Public Health* 2013;**2:**543–7.
12 Jalili M, Hejri SM, Norcini JJ. Comparison of two methods of standard setting: The performance of the three-level Angoff method. *Med Educ* 2011;**45:**1199–208.
13 Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Med Teach* 2013;**35:**e1437–46.
14 Biggs JB, Collis KF. *Evaluating the quality of learning—the SOLO Taxonomy* (1st ed). New York:Academic Press; 1982.