Research Article

"Is There a "Best" Method for Standard Setting in OSCE Exams? Comparison between Four Methods (A Cross-Sectional Descriptive Study)

Doaa Kamal [1, *], Moataz Sallam [1], Enas Gouda [1] and Sally Fouad [1]

[1]Suez Canal University, Ismailia, Egypt

[*] *Corresponding author*: Lecturer of Medical Education, Suez Canal University, Ismailia, Egypt Email: doaakherbo@gmail.com

Abstract

Background: One of the essential considerations while designing an OSCE exam is "standard-setting", which refers to the score with which a student could be considered to pass or fail an exam. The selection of proper standard-setting method is based on different criteria, including the applicability of the method, the university bylaws, and the purpose of the test.

Objectives: To examine the difference between four different standard-setting methods: the modified Cohen's, borderline regression, Hofstee methods, and the fixed 60% arbitrary method in determining the passing score in ophthalmology OSCE exam.

Methods: Two periodic ophthalmology OSCE were selected to examine the differences in failure rates and pass scores. The four standard setting methods were applied with a sample size that included 38 (year 5 undergraduate) students at the Faculty of Medicine, Suez Canal University.

Results: Modified Cohen's method resulted in the lowest passing score (54% and 58%), while the Hofstee method led to the highest passing score (69.2% and 75%). Comparing the percentage of students who are supposed to pass the exam, we can observe that there is no statistically significant difference among these standard-setting methods were used except for the Hofstee method.

Conclusions: There is no single best method for setting the passing mark of an exam. We could obtain more practical outcomes if we considered more than one method and the average pass mark.

Keywords: Standard Setting, OSCE, Modified Cohen, Borderline Regression, Hofstee

1. Background

Objective Structured Clinical Examination (OSCE) is one of the important assessment methods applied for performance assessment. It can be defined as "an assessment tool that is characterized by being objective and standardized. During this exam, students move through a series of time-bound stations in the circuit for the purposes of assessment of performance in a safe environment. Students are assessed using standardized scoring rubrics by trained examiners" (1).

Known for high-level validity and reliability levels, OSCE tests have been incorporated more extensively in the assessment strategies of different medical schools with different methods, such as long case and short case examinations. Organizing and developing OSCE is not an easy task, which requires several preparations for all those involved (2).

At our institution, in an attempt to improve standardization and objectivity of assessment in undergraduate years, different standardized assessment methods were used to improve outcomes of assessment (e.g., long case exams have been completely replaced by OSCE exams). There are some worrying factors that affect the reproducibility in OSCE exams, including students' performance across stations, inter-rater reliability as well as examination and station length. Recently, more attention has been given to standard-setting procedures (3).

There are multiple methods used for this purpose, which have been divided into three groups: norm-referenced methods, criterion-referenced methods, and combination methods (3).

In norm-referenced methods, the pass/fail scores are determined by the relative scores of students (e.g., Cohen methods). These types are usually considered unacceptable in licensing tests. On the other hand, in criterion-referenced methods, a group of experts examine each test item to determine its difficulty and relevance (e.g., borderline group and contrasting groups methods) (2).

The combination/compromise methods were designed to provide a balance between norm-referenced and criterion-referenced judgment. The main idea in these methods is that this compromise will help avoid the unreasonably high or low scores (e.g., the Hofstee methods) (4).

Studies have shown that various standard-setting methods may lead to different results. The credibility of the passing score obtained from any assessment method will be high if this method produced a standard which is consistent with the purpose of the test and based on the judgment of experts who fit the "criteria of judge selection" (2).

In 2001, Wilkinson et al., conducted a study to examine the validity and reliability of using global ratings of borderline performance to set the pass mark. They concluded that this method yielded a valid and reliable cut-off score (5).

On the other hand, Boulet JR et al., concluded that selecting the proper standard-setting method for OSCE exams should depend on the purpose of the assessment and the availability of the resources (6).

In our study, we compared four methods of standard-setting in order to determine the most effective method for establishing an appropriate passing score for a low-stake OSCE exam, these methods include the Modified Cohen's, the borderline regression, the arbitrary fixed 60% score method and Hofstee method.

1.1. Cohen Method of Standard Setting

Cohen method is one of the norm-referenced standard-setting methods that can set the standards in 'lower stakes' exams. It uses the best performing students' mark as a reference point to define the difficulty of the exam (7).

According to the Cohen method, the students' scores are arranged from the lowest to the highest scores; the 95% confidence interval (95% CI) or top 5% of the scores is highlighted, the mean was determined, and finally, 60% of the total mean score is calculated and considered as the standard/passing score. This can be expressed by the formula:

Pass Mark = R + 0.6 (X – R),

Where R is the mark which could be obtained by random guessing, and X is the mark of the 95th percentile student (7).

The previous formula was modified by Taylor (2011); accordingly, he used the score of the 90th percentile student to determine the passing score as 65% of the total mean score, making no adjustment for random marks. The formula was adjusted to be as follows.

Pass Mark = 0.65Y,

Where Y is the mark of the 90th percentile student.

1.2. Borderline Regression Method

Borderline regression is considered the best method in OSCE exams. Examiners are asked to award a global score in each OSCE station based on their subjective opinions. The global score should be selected form 3 - 5 grade descriptors, such as good pass, pass, borderline, or fail. The borderline score reflects students whom the examiner feels that they have not performed good enough to pass the test nor performed so bad to fail that test or part of it. The global scores are collected and statistically regressed against the station's checklist. The passing score is then calculated using a linear equation by assigning the midpoint of the global rating scale against the borderline group(s) scores (8).

1.3. Hofstee Method

It is one of the compromised methods of standard-setting, which has similar characteristics with both norm-referenced and criterion-referenced methods. It considers the scores of the students as well as the judges' expert group agreement about the maximum passing mark, minimum passing mark, maximum accepted failure rate, and minimum tolerated failure rate of the students (4).

The fixed arbitrary 60% method:

In our institution, as many other medical schools, the passing score as well as the failure and passing rates are previously determined according to the institution's by-laws (60%). Accordingly, the standard of any exam will be fixed.

2. Objectives

This study aims to decide the best and the most feasible method to determine the passing score for the OSCE exam by comparing four different methods-the modified Cohen's, borderline regression, Hofstee method, and the fixed arbitrary methods.Aims:

1. To determine the pass mark in one of two periodic ophthalmology OSCE exams via four different standard-setting methods: Modified Cohen method, borderline regression method, Hofstee method in addition to the arbitrary fixed 60% method.

2. To compare between the resulted pass mark using the previously mentioned methods.

3. To recommend the best method among the four used methods, if possible.

2.1. Study Questions

1. Is there a statistically significant difference in the passing score between the four different standard-setting methods: Modified Cohen method, borderline regression

method, Hofstee method, and the arbitrary fixed 60% method?

2. Is there a statistically significant difference in the number of students passing/failing the exam between the four different standard-setting methods?

3. Is there a good method for setting the passing score for a low-stake OSCE exam?

2.2. Type of the Study

It is a cross-sectional descriptive study.

2.3. Sample Type

Convenient sample.

3. Methods

-Two low-stakes ophthalmology OSCE exams (each included 5 different dynamic stations) were included in the study.

- The students' sample size included 38 (year 5) students at the Faculty of Medicine, Suez Canal University.

- The stations used for the first group examined the following topics: History taking from a patient with cataract, History taking from a patient with primary angle-closure glaucoma, History taking from a patient with exotropia, History taking from a patient with subconjunctival hemorrhage and color discrimination test 'clinical exam'.

- For the second group, the following topics were addressed: Levator function test, Confrontation test, Ocular motility testing, Pupil examination, and Ruler test for assessment of proptosis.

3.1. Standard Setting Methods
3.1.1. Modified Cohen Method

The OSCE scores of 38 students (in two groups) were plotted, and the score of the best performing student in each group was determined as a reference point according to which the pass mark was calculated as the following formula:

Pass Mark = 0.65Y,

Where Y is the mark of the 90th percentile student.

3.1.2. Borderline Regression Method

Examiners used a structured checklist for each OSCE station. They also assigned a global rating/score for each station (comprising a five-point Likert scale where 1 = fail, 2 = borderline, 3 = pass, 4 = good pass, 5 = excellent). Specific examiner training was conducted before the scheduled OSCEs to identify the borderline candidates' criteria and to explain the significance if the global rating. The two scores of each station (the true score and the global score) were regressed, and the passing score was determined. We used SPSS version 22 to analyze the data. was used.

3.1.3. Hofstee Method

We conducted a briefing session with all examiners/judges who volunteered to participate in the study to explain the steps and concepts of the Hofstee method. Then we asked each examiner to determine the maximum and the minimum students' passing scores (out of 100%) as well as the maximum and minimum rates of failed students in the whole OSCE exam (by answering the four Hofstee questions). The judges' means of the previous scores and rates were calculated. The students' scores were plotted, a cumulative chart was drawn, and the passing score was determined as clarified in the following section.

4. Results

1. The passing scores of each group in both OSCE exams were determined using Hofstee method as shown in Figures 1 and 2. According to this method, the maximum and minimum passing scores as well as the maximum and minimum failure rates were determined, then a rectangle joining the 4 points was drawn after that an intersection point between this rectangle and the cumulative chart of the students' scores was obtained. This intersection point is considered as the cut-off point/passing score of each exam: 135 out of 195 (69.2%) for group 1 and 155 out of 207 (75%) for group 2.

2. According to the borderline regression method, (Figure 2) the cut-off scores of each station in each 5-station exam was determined separately before determination of the cut-off score of the whole exam. The cut-off point of the whole exam regarding group 1 students was 113 out of 195 (58%) while for group 2 students was 151 out of 207 (73%).

3. According to the Modified Cohen method, the cut-off score of each station was determined separately according to the equation (Pass Mark = 0.65Y), where Y is the mark of the 90th percentile student. The determined cut-off score for the whole exam was determined as 105.3 out of 195 (54%) for group 1 exam and 120.8 out 207 (58%) for group 2 exam.

Tables 1 and 2 show the number and percentage of students who passed each station in group 1 and group 2 exams respectively using Cohen's and BRM.

Table 3 shows that there is a statistically significant difference the number/percentage of students passing the exam when using Hofstee method especially in group 1.

5. Discussion

In our research, we compared the results of four different standard-setting methods: the arbitrary fixed 60% standard setting method used at our medical school, a norm-referenced standard setting method, a compromise
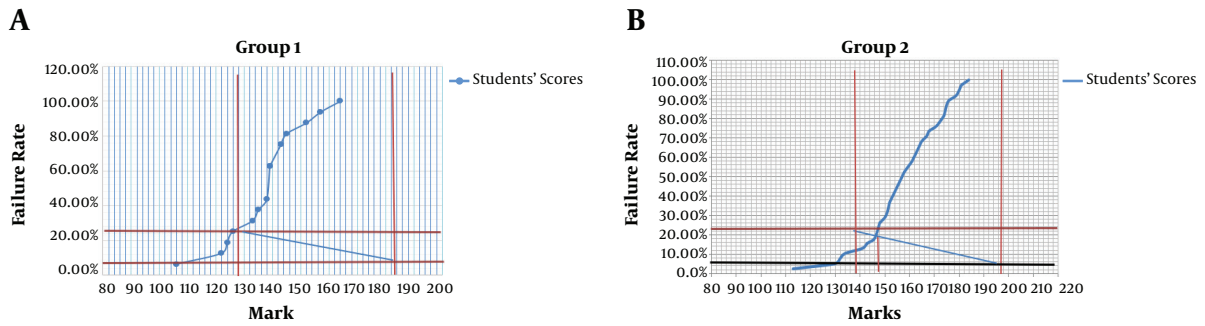
Figure 1. Students' scores in the ophthalmology periodic OSCE exam of year 5 medical students, based on the Hofstee Method

Table 1. Frequency of Passed Students in Group 1 Stations from 1 - 5 by BRM and Cohen Methods (N = 16 Students)

	BRM, No. (%)	Cohen, No. (%)	P Value[a]
Station 1	8 (50)	10 (62.5)	0.500
Station 2	15 (93.7)	13 (61.9)	1
Station 3	19 (90.5)	19 (90.5)	1
Station 4	16 (100)	16 (100)	
Station 5	6 (37.5)	11 (68.7)	0.063
Total score (195)	15 (93.7)	16 (100)	1

[a] McNemar test

Table 2. Frequency of Passed Students in Group 2 Stations from 6 - 10 by BRM and Cohen Methods (N = 21 Students)

	BRM, No. (%)	Cohen, No. (%)	P Value[a]
Station 6	15 (71.4)	17 (81)	0.500
Station 7	15 (57.1)	15 (93.7)	1
Station 8	15 (93.7)	15 (93.7)	1
Station 9	21 (100)	21 (100)	
Station 10	20 (95.2)	21 (100)	1
Total score (207)	21 (100)	21 (100)	

[a] McNemar test

method and the Modified Cohen's method. Using Hofstee method, the passing score was 135 out of 195 (69.2%) for group and 155 out of 207 (75%) for group 2 while it was 113 out of 195 (58%) for group 1 and for group 2 was 151 out of 207 (73%) using BRM. Using the Modified Cohen's method revealed passing score of 105.3 out of 195 (54%) for group 1 and 120.8 out 207 (58%) for group 2 and finally using the arbitrary 60% method revealed passing score of 117 out of 195 for group 1 and 124.2 out of 207 for group 2.

There is no statistically significant difference in percentage of students who passed the exam between the

used standard setting methods except for the Hofstee method, and this could be referred to the unfamiliarity of the judges' committee with this new method.

According to the results of a study conducted by Jalili and Mortazhejri in Tehran University of Medical Sciences in 2009, in which they used four standard setting methods -the fixed score, Angoff, borderline regression, and Cohen methods- in the preinternship exam (9), the passing scores of the total test in the pre-fixed score, Angoff, borderline regression and Cohen's methods were respectively as follows: 60, 49.15, 42.39 and 42.74. Their results showed that both the BRM and the Cohen's method yielded lower standards than that of the pre-fixed score. This is to some extent consistent with our results regarding as both the BRM and the Cohen's methods produced lower scores in comparison of the fixed arbitrary method especially in the first group.

In our study, Modified Cohen's and borderline regression methods showed the highest rates of students passing the exams, while Hofstee method showed the lowest rates. This can be congruent with the results of Makaren et al. in Mashhad, University of Medical Sciences, Iran in 2015 who aimed to evaluate the Passing Scores in Semiotics (OSCE) (10), in which the Cohen's method showed the highest passing rate among all the compared methods, while BRM showed the lowest passing rate and this as was explained by Maharen could be attributed to the subjectivity nature of the BRM.

Wright (7) used Cohen and modified Cohen standard setting methods at the Faculty of Medicine, University of Botswana, to apply the two methods to the exams of eight groups of nearly 50 students from first and second years of the MBBS course and set the passing score. His results showed that the two methods decreased the average test failure rates than the Fixed method and this is also matching the results of our study.
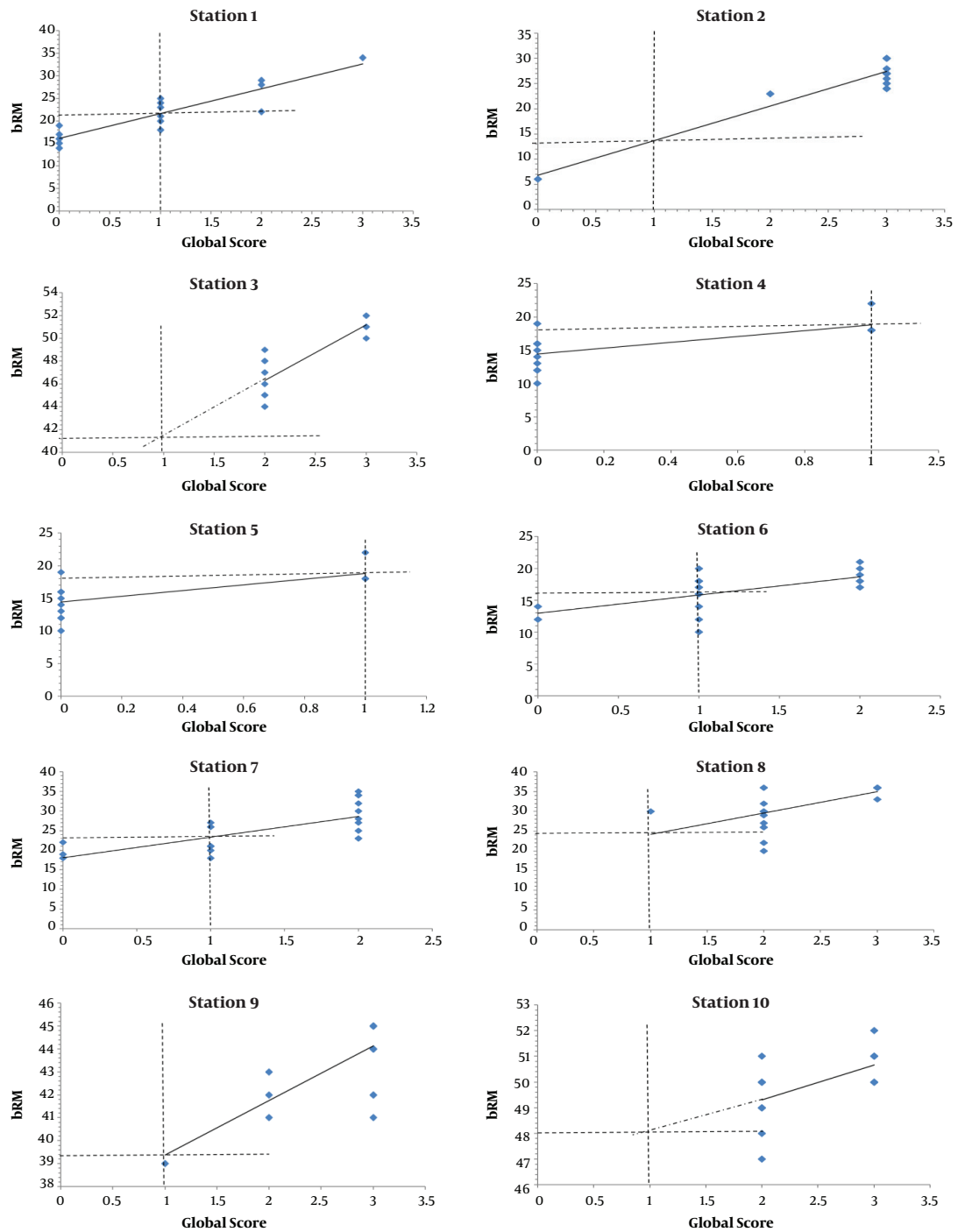
Figure 2. Scatter plots of the checklist score versus the global score for the ten stations in the in the periodic ophthalmology objective structured clinical examination (OSCE) in 2 groups of candidates

5.1. Limitations of the Study

• Unfamiliarity of the faculty at the FOM-SCU regarding standard setting methods and its importance in setting the pass mark of the exam.

• Most of the evaluators as well as the judges included in the study were Senior Professors and staff members who

Table 3. Percent of Students Passed the Exam After Using BRM, Cohen and Hofstee Methods Compared to the Fixed Arbitrary 60% Method

	BRM	Cohen	Fixed	Hofstee	P Value[a]
Stations 1-5 (195) (n = 16 students)	15 (93.7%)	16 (100%)	15 (93.7%)	9 (56.3%)[b]	0.001[c]
Station 6 – 10 (207) (n = 21 students	21 (100%)	21 (100%)	21 (100%)	21 (100%)	

[a]Cochran's Q test
[b]Statistically significant when compared with BRM, statistically significant when compared with Cohen; statistically significant when compared with fixed method.
[c]Statistically significant at $P < 0.05$

have limited contact with the students compared to the more junior staff members, which make the criteria of defining the borderline student & borderline score will be questionable.

• The training process was not evaluated for effectiveness.

5.2. Conclusions

This study concluded that using the fixed 60% arbitrary method to determine the pass mark for all exams irrespective of the differences between students as well as teaching and training for each group will result in marked difference in the failure and pass rates among students. Also, as it was reached from this study that Hofstee method could be the less accurate method for setting the pass mark of an exam especially when there is poor training and practice of the involved judges. Finally, it is concluded that there is no single method could be the best for setting the pass mark of an exam, instead it would be better if the selection of standard setting method was based on the feasibility and applicability of each method, also a better outcome could be reached if more than one method were applied in combination and the average passing score is considered.

Footnotes

Authors' Contribution: All authors developed the original idea and the protocol, abstracted and analyzed data, wrote the manuscript, and is a guarantor, they also contributed to the development of the protocol, abstracted data, and prepared the manuscript.

Conflict of Interests: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Ethical Approval: This research was approved by the FOM-SCU research Ethics Committee at Faculty of Medicine, Suez Canal University.

Funding/Support: No funding support.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

References

1. 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**(9):e1437–46. doi: 10.3109/0142159X.2013.818634. [PubMed: 23968323].

2. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach.* 2013;**35**(9):e1447–63. doi: 10.3109/0142159X.2013.818635. [PubMed: 23968324].

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**(3):267–71. doi: 10.1097/00001888-200003000-00018. [PubMed: 10724316].

4. Bay L. Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests by Cizek, G. J., & Bunch, M. B. *Journal of Educational Measurement.* 2010;**47**(4):482–6. doi: 10.1111/j.1745-3984.2010.00126.x.

5. Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. *Med Educ.* 2001;**35**(11):1043–9. doi: 10.1046/j.1365-2923.2001.01041.x. [PubMed: 11703640].

6. Boulet JR, De Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach.* 2003;**25**(3):245–9. doi: 10.1080/0142159031000100274. [PubMed: 12881044].

7. Wright JD. The Application of Simple, Practical and Affordable Standard Setting Methods in Small Groups of Students. *MedEdPublish.* 2016;**5**(3). doi: 10.15694/mep.2016.000106.

8. Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Objective structured clinical examinations. *Med Educ.* 2003;**37**(2):132–9. doi: 10.1046/j.1365-2923.2003.01429.x. [PubMed: 12558884].

9. Jalili M, Mortazhejri S. Standard Setting for Objective Structured Clinical Exam Using Four Methods: Pre-fixed score, Angoff, Borderline Regression and Cohen's. *Strides in Development of Medical Education.* 2012;**9**(1):77–84.

10. Makaren A, Mahdavifard H, Gholami H. Evaluation of Passing Scores in Semiotics: An Objective Structured Clinical Examination for Medical Students of Mashhad University of Medical Sciences, Mashhad, Iran, 2015. *Strides in Development of Medical Education.* 2017;**14**(1). doi: 10.5812/sdme.59227.