Research Article

Evaluation of Passing Scores in Semiotics: An Objective Structured Clinical Examination for Medical Students of Mashhad University of Medical Sciences, Mashhad, Iran, 2015

Abbas Makaren,[1] Hamid Mahdavifard,[1,*] and Hasan Gholami[1]

[1]Department of Medical Education, Mashhad University of Medical Sciences, Mashhad, IR Iran

[*]*Corresponding author*: Hamid Mahdavifard, Department of Medical Education, Mashhad University of Medical Sciences, Mashhad, IR Iran. E-mail: mahdavifh1@mums.ac.ir

Abstract

Background: Numerous exams are held at different levels and in different fields of medical sciences to evaluate students' practical knowledge. In pass-fail exams where several examiners score the students, it is important to determine "the minimum passing score" or "the passing score" to determine whether students have passed or failed; this score is sometimes called the "cut-off point" or "standard score." The objective structured clinical examination (OSCE) method is employed for the final assessment of medical students in Semiotics I in Mashhad University of Medical Sciences, Mashhad, Iran. The commonly used standard scoring method for this lesson is the fixed score method, which sometimes results in a discrepancy between educational management and the lecturers. Hence, the current study aims to compare 4 different methods—the Cohen, borderline-group, borderline regression, and Hofstee methods—of determining the passing score in the semiotics course and comparing the results with those of the fixed score method.

Methods: A 6-station OSCE was used to assess Semiotics I in Mashhad University of Medical Sciences in 2015. In the current study, in order to determine a standard scale for scoring the students, two forms, Forms 1 and 2, and a checklist were completed for each student. In Form 1, a 5-option Likert scale scoring system, graded from poor to excellent, was used. Data from Form 1 were analyzed using the borderline regression and borderline-group methods. Form 2 included 4 items and the collected data were analyzed using the Hofstee method. Data collected from both forms were analyzed, after the exams, using SPSS version 16.

Results: The cut-off point established by the Cohen method was very close to that of the common method. In other words, there was no significant difference between the cut-off point determined by the Cohen method (11.73) and that of the common method (12). The other study methods, however, such as borderline regression and borderline-group methods proposed higher cut-off points, which were significantly different from that of the common method: more students failed Semiotics I using these methods. The Hofstee method cannot be used in the OSCE, as the results were insignificant.

Conclusions: Because there was a significant difference in the number of students who passed the exam based on the fixed score and Cohen methods, and on the borderline-group and borderline regression methods, it is recommended that the latter methods not be widely employed. In addition, it is suggested that different methods should be used to define a mean standard passing score because, according to the statistics, an accurate and efficient estimator with minimum variances accuracy should be employed to evaluate population parameters, and the mean estimator would benefit from such advantages.

Keywords: Objective Structured Clinical Examination, Standard Score Determining Methods, Cut-Off Point

1. Background

To assess students' knowledge, a final exam is held for each course (1, 2) that is used as an index to define which students have passed and which have failed. There are different exams used to assess students, some of which are paper-pencil (cognitive) type, and some of which are practical (3-5). In medical education, practical tests are often applied due to the nature of lessons. The objective structured clinical examination (OSCE) is one of the practical exams used in medical education. The OSCE assesses students' potency in certain clinical practices at several stations. There are different methods used to determine the passing score in the OSCE (5-7) and one of the most common is the fixed

score method (5, 8, 9), although there are different methods such as the Cohen, borderline-group, regression borderline, the Hofstee, etc. (8, 10-13). This study aims to compare the common fixed score method with some of these other methods in order to determine the standard score.

Different exams are held by medical universities every year (3, 14). In paper-pencil (cognitive) exams, the fixed score method is usually used to determine the passing score; for example, in medical education, scores of 10 and 12 are the baseline for fundamental sciences and clinical practice courses respectively. Different methods are used in different medical universities to determine the passing score. Different studies have shown that using differ-

ent methods to determine the passing score in a particular course result in different cut-off points (15-17). In other words, different methods produce different cut-off points (10, 14, 15, 18, 19).

To determine a standard score based on a particular exam, it is necessary to determine the cut-off point or the minimum passing score to define whether students have passed or failed (1, 18, 20-22). The standard score is a judgment made by the professionals; the purpose of the exam, students' potency, and their socio-educational status can also affect the professionals' decision (2, 3).

To date, different methods have been used to determine the passing score for different exams (18-20, 23). The fixed score, the Angoff, the Nedelsky, the Abel, distinct groups (interaction), borderline-group, regression borderline, the Hofstee, and the Cohen methods are employed in practical exams for medical students (1, 20). Based on the purpose of the exam and dominant circumstances, a particular method is used to determine the passing score (5, 7, 8, 24).

In Semiotics I, pathophysiological topics assessed by the OSCE are usually scored based on the fixed score method. In other words, a score of 12 is required to pass the exam; however, different methods define different passing scores. The current study aims to evaluate 4 methods used to determine the passing score in the OSCE on pathophysiological topics, and also compares the results with the scores of the common method in order to distinguish the most reliable method of determining the passing score in Semiotics I for medical students at Mashhad University of Medical Sciences in August 2015.

2. Methods

A week prior to the exam, a briefing session was held with the examiner at the clinical skills learning center of Qaem hospital to fully familiarize the teachers with the methods employed in the current study. Accordingly, the teachers were asked to comment on the minimum and maximum passing scores, as well as the minimum and maximum number of the students who passed and failed the exam, based on a 5-option Likert scale scoring model – poor, borderline, acceptable, good, and excellent – in Form 1 (1, 6, 15, 25) and on Form 2, which included standard multiple-choice questions previously distributed among the teachers (1, 15, 26-28). For this purpose, the teachers were briefed on how to complete the forms and asked to complete both forms and the checklist. The teachers categorized the students as having passed or failed according to their practical knowledge in Form 1, which was designed to define the passing scores based on the borderline-group and borderline regression methods. Form 2 included 4

questions to determine the maximum and minimum passing scores or the lowest fail score, and was designed based on the Hofstee method.

Before holding the session, the teachers were informed via e-mail, and the forms were also distributed to them this way.

The exam was held on the predetermined date at the clinical skills learning center in the university. A total of 126 students were examined to sit the exam, of which 125 attended the exam. The students were assigned into groups of 6. Each station was repeated twice; accordingly, 2 students answered the questions. A total of 125 copies of Form 1 were distributed in each station and 1 form was completed for each student; accordingly, each student was classified as either poor, borderline, acceptable, good, or excellent based on the teachers' comments. In total, 750 copies of Form 1 were completed at the 6 stations. To complete Form 2, the teachers were asked to attend the Clinical Skills Learning Center after the exam where Form 2 was distributed and completed by teachers based on the data previously provided at the briefing session.

The maximum score for each station was 100; after completing the scoring process at all stations, the scores were totaled and the students were scored on a 20-point scale. Each student spent 5 minutes at each station.

To observe ethical considerations, the student number was used to identify the students in the checklists and forms, and the researchers were blind to the students' scores.

2.1. Standard Scoring Methods

According to the Cohen method, the students' scores were sorted from low to high; the 95% confidence interval (95%CI) or top 5% of scores was then determined, the mean was measured, and finally 60% of total mean score was taken as the standard score (6, 13, 19, 27, 29).

In the borderline-group method, the students were assessed by the examiner in each station using the checklist and general assessment. After completing Form 1 and identifying the borderline students, based on the general assessment, their scores were extracted from the checklists and the mean score for each station was measured and taken as "the standard score of the station" (6, 7, 19, 30, 31).

In the regression borderline method, the students were assessed by the examiner in each station using the checklist and general assessment (completing Form 1). A regression equation was then used for each station to predict the checklist score of the borderline student, with the checklist score as the dependent variable and the general assessment score (based on a 5-option Likert scale scoring system) as the independent variable (3, 6, 28, 30, 32).

To analyze the data, SPSS version 16 was used; tables and figures were created using Microsoft Excel.

To determine the passing score based on the Hofstee method, a chart was first designed using Excel. The maximum passing scores and the minimum students' scores were shown on the X axis. The minimum rate of failed students and maximum rate of passed students were marked on the Y axis as "a" and "b" respectively. The measures were calculated for each student on the chart and then, a cumulative chart was drawn based on the mean scores of the students. By stretching the 2 measures on the X axis and "a" and "b" on the Y axis, a rectangle was drawn. The cumulative chart was then drawn based on the students' scores and the passing score was determined at the intersection point of the rectangle diameter and the cumulative chart of the students' scores (1, 3, 28).

According to Figure 1, there was no intersection between the rectangle diameter and the cumulative chart of the students' scores. Since, according to the Hofstee method, the intersection point between the rectangle diameter and the cumulative chart of students' score is considered as the cut-off point and there was no cross point in the provided chart, the Hofstee method was not applicable to the current study.
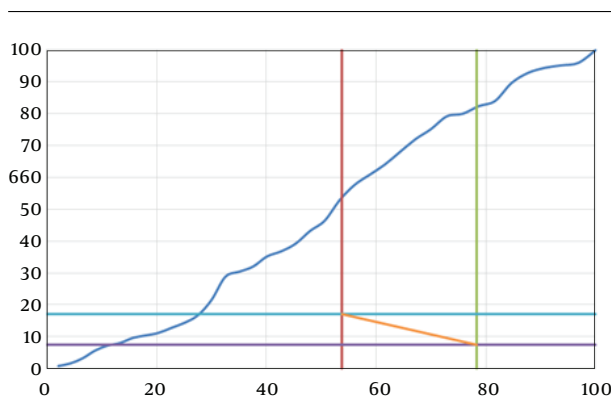

Figure 1. Students' Scores in the Final Exam of Semiotics I, Based on the Hofstee Method

3. Results

The practical exam for Semiotics I was held in August 2015 for medical students admitted to the university in September 2012. In this practical exam, held based on the OSCE, 125 students participated, of which 55 (44%) were male and 70 (56%) were female.

According to the Cohen method, the mean cut-off point for each station was as follows: gastrointestinal, 12;

heart, 12; lungs auscultation, 12; communication skills, 12; basic resuscitation, 10.39; and emergency medicine, 12.

The determined cut-off point for all the items in the exam was 11.73 on the 20-point scale, based on the Cohen method.

The cut-off points of the 6-station OSCE, based on the borderline-group method, are shown in Table 1.

According to Table 1, in order to determine the passing score based on the borderline-group method, the passing score was first measured for each station and the mean passing score was then measured as 16.44.

The exam passing score was the mean score of the stations, which was 16.44 based on the 20-point scale, according to the borderline-group method.

The cut-off points of the 6-station OSCE, based on the regression borderline method, are shown in Table 2.

According to Table 2, in order to determine the passing score based on the regression borderline method, the linear equation of $Y = aX + b$ was used; a and b were measured from the chart, the fixed coefficient of X was obtained from the statistics analysis with SPSS, and Y was the cut-off point of different stations. The mean cut-off point (Y) for the 6 stations of 16.79 was taken as the passing score, based on the regression borderline method.

According to Table 2, the cut-off point of the exam was 16.79 based on the 20-point scale, using the regression borderline method.

According to the Hofstee method, after drawing the X and Y axes, the rectangle, and its diameter, no crossing point was observed between the rectangle diameter and the cumulative chart of the students' scores. Hence, the Hofstee method was not applicable to obtaining the cut-off point in the exam considered in this study.

If the Cohen method is used as a pass-fail scale, no student fails the exam, as the cut-off point was 11.73. If using the borderline-group to determine the passing score of the exam, however, the cut-off point is 16.44 and 44 (35.2%) students fail the exam, which is a high failure rate. According to the borderline regression method, with the cut-off point of 16.79, 49 (39.2%) students fail the exam, which is also a high failure rate.

4. Discussion and Conclusion

According to the results of a study by Jalili and Mortazhejri in Tehran University of Medical Sciences in 2009, comparing 4 scoring methods – the fixed score, Angoff, borderline regression, and Cohen methods – in the pre-internship exam (27), the highest and lowest pass rates were obtained using the borderline regression and fixed score methods. In the current study, however, the Cohen

Table 1. The Distribution of Students' Scores in the Exam Stations, Based on the Borderline-group Method for Semiotics I

Station	Number of Points in Each Station	Minimum Score	Maximum Score	Mean Score	Standard Deviation	Passing Score for the Station
Gastrointestinal	8	15	20	18.25	1.66	18.25
Heart	5	16	20	19	1.73	19
Lungs auscultation	20	17	20	18.95	1.05	18.95
Communicational skills	32	8	20	13.71	4.55	13.71
Basic resuscitation	6	16	18	16.66	1.03	16.66
Emergency medicine	31	4	15	12.09	2.44	12.09

Table 2. The Linear Regression Equation for for Students' Scores in Six Stations for Semiotics I

Station	a	b	Y	X	Cut-off Point
Gastrointestinal	17.333	0.155	17.798	3	17.80
Heart	17.037	0.273	17.856	3	17.86
Lungs auscultation	18.726	0.035	15.831	3	18.83
Communicational skills	11.003	1.398	15.197	3	15.2
Basic resuscitation	14.439	0.449	15.786	3	15.79
Emergency medicine	7.082	2.726	15.26	3	15.26

and borderline regression methods showed the highest and lowest rates of passing scores respectively.

To compare the results of the current study with those of Wood et al. (24), the determined passing scores based on the borderline-group and the borderline regression were lower than those of other methods evaluated in both studies.

Research conducted by Kaufman et al. from 1996 to 1997 compared 5 passing score methods and showed that the standard scores of the borderline and the Angoff methods were similar.

In a study by Kramer et al., the borderline regression, the Angoff individual, and the Angoff methods were compared based on the exact scores; results showed that the standard score of the borderline method was the lowest, compared with those of the other methods studied (15). The results of the current study show that distinguishing the best method is practically impossible, since the difference between the standard scores of different methods was so high. However, the borderline regression method seems to be more applicable as it was more flexible and benefited from good validity and reliability. It was specifically well-designed for the performance-based tests and is preferred to the Angoff method. The results of the current study indicated, however, that both the borderline methods gave higher passing scores and were therefore stricter than the other methods studied.

Reid et al. in 2014 reported a passing rate of > 50% following the use of the borderline regression and the borderline-group methods, which is similar to the results of the current study.

Since there is a significant difference between the passing rates of the fixed score and the Cohen compared to those of the borderline-group and borderline regression methods, an evaluation of other passing score methods is recommended, in which the peer reviewers discuss the test items before holding the exam (test-oriented methods).

In test-oriented methods, each peer reviewer independently evaluates the probability of correct answers to each of the checklist questions for a borderline student, and the average comments of the peer reviewers are then considered as the passing score. Therefore, the peer reviewers can comment on the checklist, which seems logical.

It is also recommended that the mean scores of different passing score methods are measured and then considered as the cut-off point. A combined index is a better scale for decision making and is statistically more defendable; in addition, in order to find the community parameters statistically, an accurate and efficient linear estimator eliciting the minimum variances is required. According to the theoretical statistics, a mean estimator would benefit from the mentioned properties.

The current study was the first to use the Hofstee method in Mashhad University of Medical Sciences and re-

sults of the study indicated the high cut-off point of the Hofstee method, compared with that of the fixed score method. It seems that a revision is necessary to promote the level of knowledge among graduates. However, changing the standard scoring system was not welcomed by the educational management of the university and the authorities preferred to use traditional scoring methods, resulting in the authors of this study encountering difficulties. In addition, the lecturers were unfamiliar with the new scoring methods and needed to attend briefing sessions for this, but showed little interest due to their involvement in clinics.

4.1. Conclusions

Results showed that, considering only the checklist score as the passing score and the fixed score and Cohen methods to determine the cut-off point, the cut-off point was lower and accordingly the number of passed students increased; as a result, all exam participants passed the exam. However, when the checklist score and students' grades based on the Likert scale were used to determine the cut-off point, as well as the borderline regression and borderline-group methods, the passing score significantly increased and about one-third of the students failed the exam, producing a high failure rate that was unacceptable.

Because the minimum passing score for Semiotics I was 12, all the participating students passed the exam, which can be justified in 2 ways: a high level of practical knowledge among the students or the lecturers' choice of an easy exam to assess the students.

Supplementary Material

Supplementary material(s) is available here.

Acknowledgments

References

1. Mortaz Hejri S, Jalili M. Standard setting in medical education: fundamental concepts and emerging challenges. *Med J Islam Repub Iran.* 2014;**28**:34. [PubMed: 25250275].
2. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ.* 1979;**13**(1):41–54. [PubMed: 763183].
3. Cusimano MD. Standard setting in medical education. *Acad Med.* 1996;**71**(10 Suppl):S112–20. [PubMed: 8940953].
4. Wayne D, Cohen E. Standard Setting in Competency Evaluation. ; 2010.
5. Norcini JJ. Setting standards on educational tests. *Med Educ.* 2003;**37**(5):464–9. [PubMed: 12709190].
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. Boursicot KA, Roberts TE, Pell G. Using borderline methods to compare passing standards for OSCEs at graduation across three medical schools. *Med Educ.* 2007;**41**(11):1024–31. doi: 10.1111/j.1365-2923.2007.02857.x. [PubMed: 17973762].
8. Kane MT, Crooks TJ, Cohen AS. Designing and Evaluating Standard-Setting Procedures for Licensure and Certification Tests. *Adv Health Sci Educ Theory Pract.* 1999;**4**(3):195–207. doi: 10.1023/A:1009849528247. [PubMed: 12386478].
9. Schoonheim-Klein M, Muijtjens A, Habets L, Manogue M, van der Vleuten C, van der Velden U. Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting methods. *Eur J Dent Educ.* 2009;**13**(3):162–71. doi: 10.1111/j.1600-0579.2008.00568.x. [PubMed: 19630935].
10. Ben-David MF. AMEE Guide No. 18: Standard setting in student assessment. *Med Teach.* 2009;**22**(2):120–30. doi: 10.1080/01421590078526.
11. 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. [PubMed: 10724316].
12. Plake BS, Hambleton RK, Jaeger RM. A New Standard-Setting Method for Performance Assessments: The Dominant Profile Judgment Method and Some Field-Test Results. *Educ Psychol Measure.* 2016;**57**(3):400–11. doi: 10.1177/0013164497057003002.
13. Ricker KL. Setting cut-scores: A critical review of the Angoff and modified Angoff methods. *Alberta J Educ Res.* 2006;**52**(1):53.
14. Maurer TJ, Alexander RA, Callahan CM, Bailey JJ, Dambrot FH. Methodological and Psychometric Issues in Setting Cutoff Scores Using the Angoff Method. *Personnel Psychol.* 2006;**44**(2):235–62. doi: 10.1111/j.1744-6570.1991.tb00958.x.
15. 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. [PubMed: 12558884].
16. Brandon PR. Conclusions About Frequently Studied Modified Angoff Standard-Setting Topics. *Appl Measure Educ.* 2004;**17**(1):59–88. doi: 10.1207/s15324818ame1701_4.
17. 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**(12):1199–208. doi: 10.1111/j.1365-2923.2011.04073.x. [PubMed: 22122428].
18. Chinn RN, Hertz NR. Alternative Approaches to Standard Setting for Licensing and Certification Examinations. *Appl Measure Educ.* 2002;**15**(1):1–14. doi: 10.1207/s15324818ame1501_01.
19. Troncon LE. Clinical skills assessment: limitations to the introduction of an "OSCE" (Objective Structured Clinical Examination) in a traditional Brazilian medical school. *Sao Paulo Med J.* 2004;**122**(1):12–7. [PubMed: 15160521].
20. 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. [PubMed: 11703640].
21. Davison I, Bullock AD. Evaluation of the Introduction of the Objective Structured Public Health Examination. Birmingham: The University of Birmingham; 2007.
22. Smee SM, Blackmore DE. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Med Educ.* 2001;**35**(11):1009–10. [PubMed: 11703635].
23. Humphrey-Murto S, MacFadyen JC. Standard setting: a comparison of case-author and modified borderline-group methods in a small-scale OSCE. *Acad Med.* 2002;**77**(7):729–32. [PubMed: 12114151].

24. Wood TJ, Humphrey-Murto SM, Norman GR. Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci Educ Theory Pract.* 2006;**11**(2):115–22. doi: 10.1007/s10459-005-7853-1. [PubMed: 16729239].

25. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach.* 2008;**30**(9-10):836–45. doi: 10.1080/01421590802402247. [PubMed: 19117221].

26. Searle J. Defining competency - the role of standard setting. *Med Educ.* 2000;**34**(5):363–6. [PubMed: 10760120].

27. Jalili M, Mortazhejri S. Standard Setting for Objective Structured Clinical Exam Using Four Methods: Pre-fixed score, Angoff, Borderline Regression and Cohen's. *Strid Dev Med Educ.* 2012;**9**(1):77–84.

28. Cizek GJ. Standard-Setting Guidelines. *Educ Measure Issues Pract.* 2005;**15**(1):13–21. doi: 10.1111/j.1745-3992.1996.tb00802.x.

29. Bay L. Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests by Cizek, GJ, & Bunch, MB. Wiley Online Library; 2010.

30. Reid K, Dodds A. Comparing the borderline group and borderline regression approaches to setting Objective Structured Clinical Examination cut scores. *J Contemp Med Educ.* 2014;**2**(1):8. doi: 10.5455/jcme.20130921013523.

31. Barman A. Standard setting in student assessment: is a defensible method yet to come? *Ann Acad Med Singapore.* 2008;**37**(11):957–63. [PubMed: 19082204].

32. Taylor CA. Development of a modified Cohen method of standard setting. *Med Teach.* 2011;**33**(12):e678–82. doi: 10.3109/0142159X.2011.611192. [PubMed: 22225450].