SHORT COMMUNICATION

Comparing the Borderline Group Method with the Borderline Regression Method in Standard Setting and Measuring OSCE Quality Assurance Metrics in a Malaysian Family Medicine Examination.

Chan Sook Ching[1], Tan Chow Wei[2].

[1]*Community Based Department, Faculty of Medicine, Universiti Kuala Lumpur Royal College of Medicine Perak, 3 Jalan Greentown, 30450 Ipoh, Perak Darul Ridzuan, Malaysia.*
[2]*Academy of Family Physicians of Malaysia, Unit 1.5, Level 1, Enterprise 3B, Technology Park Malaysia, Jalan Innovasi 1, Lebuhraya Puchong-Sungai Besi, 57000 Bukit Jalil, Kuala Lumpur, Malaysia.*

Corresponding Author
Prof. Dr. Chan Sook Ching
Community-Based Department, Faculty of Medicine
Royal College of Medicine Perak, Universiti Kuala Lumpur, Ipoh, Malaysia.
Email: scchan@unikl.edu.my

Abstract

Objective: To determine if there were any significant difference in cut scores and any advantage in using the borderline group method (BGM) versus the borderline regression method (BRM) in standard setting the Objective Structured Clinical Examination (OSCE) in a high stakes Family Medicine examination.
Methods: All the 2018 and 2019 Part 2 Conjoint MAFP/FRACGP examination OSCE marks, together with the cut scores obtained from BGM, were compiled and entered into SPSS version 23. BRM cut scores were obtained by plotting linear regression graphs using the candidates' total marks against their global assessment grade in each OSCE station. The cut scores from both methods were compared to see if there were any significant difference (p value<0.05). OSCE quality assurance metrics (Cronbach's alpha, R^2 coefficient, intergrade discrimination, number of failures, and between-group variation percentage, for each OSCE station) were calculated and the results analysed.
Results: No significant difference in cut scores were found using both methods. Using BRM, additional OSCE quality assurance metrics (R^2 coefficient, intergrade discrimination) were obtained from the linear regression graphs as compared to BGM.
Conclusions: BRM required more work to compute but had the advantage of producing more OSCE quality assurance metrics. Cut scores could be obtained even when no candidates were assessed as borderline in any particular station, making it suitable for small scale OSCEs with fewer candidates. As a result, BRM is now used in standard setting for Part 2 conjoint examinations.

Keywords: *borderline group method, borderline regression method,* Conjoint MAFP/FRACGP, *family medicine, standard setting.*

Introduction

The Academy of Family Physicians of Malaysia (AFPM) conjoint with the Royal Australian College of General Practitioners (RACGP) have been conducting annually a high stakes postgraduate Family Medicine examination since 1996. Successful graduates can register as Family Medicine specialists in the Malaysian National Specialist Register, (established in 2007), after undergoing a credentialing period. The graduates are also eligible to become Members of AFPM (MAFP) and Fellows of RACGP (FRACGP) and thereby to practise in Australia [1].

The conjoint examination consists of a written component (Part 1) and a clinical component (Part 2). After much discussion, deliberation, and training workshops on the standard setting methods to implement, AFPM introduced the modified Angoff method in the 2013 Part 1 examination, and borderline group method (BGM) in the 2015 Part 2 examination [2-6].

In BGM, examiners assessed the candidates using a key features checklist in the Objective Structured Clinical Examination (OSCE). Examiners also gave their global impression on the candidate's performance. The rating form had six options for global assessment, ranging from serious deficiency, competence not demonstrated, competence unclear, minimally competent, clearly competent and highly competent. In BGM all the scores of candidates who were marked as minimally competent (borderline pass) in the global assessment section were compiled and the average mark taken as the cut score or pass mark for the station. The sum of all the borderline marks for all the stations became the passing mark for the OSCE.

In BGM, the cut scores were easy to compute, but were dependent on examiners marking the candidate as borderline (minimally competent). If none of the candidates in a particular station were marked as borderline, BGM could not be used but another method (the modified Angoff method) would be required to get the cut score or pass mark. In the modified Angoff method, a group of raters / judges would have to go through the OSCE question from the station with all the information provided (case scenario, checklist and answers). Based on the information from the OSCE question, the raters would independently decide on the score a borderline candidate would obtain. The average score from the sum of all these scores, would be used as the cut score/pass mark for that station. This was a concern for small scale OSCEs such as the conjoint examination, as the numbers sitting for the annual examination were small usually less than one hundred [7].

In BRM the examiner also rates the candidates' performance just like BGM. However, instead of taking the average of the borderline scores as the cut score or passing mark, all the global scores and the candidates' actual marks are plotted into a graph (x-y axis) for each station. Using linear regression, the formula to calculate the cut score/passing mark is obtained.

As such, it was decided to analyse the examination results of 2018 and 2019 and standard set scores obtained by BGM and compare them to scores calculated using the borderline regression method (BRM) to see if there were any significant differences and whether BRM had advantages over BGM in measuring the quality of OSCE metrics [8-11].

Materials and methods

The specific objectives of the study were: i) to obtain and compare the cut scores using BGM versus BRM in the 2018 and 2019 Part 2 Conjoint MAFP/FRACGP examinations to see if there was any significant difference between the cut scores obtained by the two methods, ii) to measure the quality of OSCE by comparing quality assurance metrics derived from BGM and BRM as recommended in An International Association for Medical Education (AMEE) guide no 49 [11].

Approval to conduct this study was obtained from the Board of Senior Examiners (BOSE), AFPM, which was overall in charge of conducting the conjoint examinations.

All the 2018 and 2019 Part 2 examination marks obtained by candidates in each OSCE station and their global assessment grades were compiled and entered into IBM Statistical Package for the Social Sciences (SPSS) version 23. As the standard setting was conducted using BGM, the cut scores obtained from BGM were available.

To calculate the cut scores using BRM, linear regression graphs were plotted using the candidates' total marks against their global assessment grade in each station. The station cut score was calculated from the formula generated [8]. The BGM and BRM cut scores were then compared to see if there were any significant difference (p value<0.05).
Following the AMEE guide no 49, a number of OSCE quality assurance metrics were calculated including Cronbach's alpha, R^2 coefficient, intergrade discrimination, number of failures and between group variation percentage, for each OSCE station [11]. The results were analysed to see if they were in the acceptable range.

Results

There was no significant difference between the cut scores obtained for each station using both methods for both the 2018 and 2019 examinations (see Table 1). The overall passing mark / cut score obtained from BGM versus BRM were very close i.e., 553.22(69.15%) compared to 555.8 (69.47%) for 2018 and 549.9 (68.74%) compared to 550.1 (68.76%) for 2019 (refer Table I), with a difference of only 0.32% (2.58/800) and 0.02% (0.2/800) respectively for the 2018 and 2019 examinations.
Analysis of OSCE quality metrics are detailed in Tables II and III). Both R^2 coefficient and intergrade discrimination were obtained from the linear regression graphs plotted to get BRM scores.
Cronbach's alpha measuring internal consistency were acceptable: 0.768 (2018), and 0.910 (2019); R^2 coefficient (squared linear correlation between the global rating score and checklist score)

showed good correlation (>0.5). Intergrade discrimination (measuring average increase in scores of the checklist for each grade increase on the global rating) showed adequate discrimination (≥ 5) in 11 out of 15 stations in 2018 and in all 15 stations in 2019 examinations. Between group variation percentage were acceptable (<30%) in all 15 stations in 2018 and in 13 stations in 2019. Two stations in 2019 had group variation percentage of 30.8% and 33.6%. (unacceptable range was >40%). In 2018, six stations and in 2019, four stations had >50% failures respectively.

Discussion and Conclusion

There was no significant difference between the cut scores obtained by BGM and BRM. The overall passing marks obtained by the two methods differed only by 0.32% and 0.02% in 2018 and 2019, respectively. Other studies have shown similar results [7,9,12].
 Although BRM required more work to compute, as linear regression graphs had to be plotted and BRM scores calculated from formula derived from the graphs, it had the advantage of producing more quality assurance metrics in measuring the quality of the OSCE i.e. R^2 coefficient and intergrade discrimination (see Tables 2 & 3).
BRM also had the additional advantage of being able to obtain cut scores / passing mark even if there were no candidates assessed as borderline in any particular station and therefore suitable for small scale OSCEs with fewer candidates. This was also reported in another study [7].
Using BRM, there is no necessity to use another method like the modified Angoff method should there be no borderline candidates in a particular station [13]. The modified Angoff method being "test-centred" and not "examinee-centred" has been utilised as a "back-up" by the conjoint examination. If there are no borderline candidates in any station from BGM, the modified Angoff method is then used to obtain the cut score / passing mark. The medical college where the first

author is attached to, also uses the modified Angoff method to standard set the OSCE questions, prior to their MBBS Final Professional OSCE examination. This serves as a "back-up" in case there is no borderline candidate in any OSCE station.

In addition, because of the small numbers of candidates in small scale OSCEs, some stations may have very few borderline scores. Using BGM, the passing mark will be based on these few scores, whereas in BRM all candidates scores are taken into consideration when plotting the linear regression graphs. The accuracy of the cut scores / passing mark obtained in stations with very few borderline marks is another area of concern.

BRM clearly shows a number of advantages over BGM in standard setting small scale OSCE. Thus, BRM is recommended for any institutions running small scale OSCEs. Based on the findings of this study, the BOSE in AFPM made a decision to switch to using BRM in standard setting its Part 2 Conjoint examinations.

Table 1. Comparing the cut scores obtained by Borderline Group Method (BGM) & Borderline Regression Method (BRM) for 2018 & 2019 Conjoint examinations.

No	Station	2018		2019	
		BGM Cut Score/%	BRM Cut score /%	BGM Cut score /%	BRM Cut score /%
1	CESC1	35.52 (71.0)	35.7 (71.4)	35.7 (71.3)	34.25 (68.5)
2	CESC2	36.56 (73.1)	35.7 (71.4)	35.1 (70.3)	34.61(69.22)
3	PP	33.89 (67.8)	34.7 (69.4)	35.9 (71.7)	35.05 (70.1)
4	SPCSC1	33.69 (67.4)	33.63 (67.26)	31.9 (63.8)	33.7 (67.4)
5	SPCSC2	37.08 (74.2)	36.61 (73.22)	37.1 (74.2)	37.16 (74.32)
6	SPCTC	34.06 (68.1)	33.76 (67.52)	34.4 (68.8)	34.27 (68.54)
7	SPCMI1	34.10 (68.2)	33.91 (67.82)	33.2 (66.4)	33.51 (67.02)
8	SPCMI2	34.23 (68.5)	34.44 (68.88)	35.4 (70.7)	35.59 (71/18)
9	SPCMI3	33.60 (67.2)	34.06 (68.12)	33.9 (67.8)	33.66 (67.32)
10	OEPLD	32.81 (65.6)	32.84 (65.68)	33.0 (66.1)	33.93 (67.86)
11	OEPM	33.91 (67.8)	34.88 (69.76)	33.1 (66.2)	33.39 (66.78)
12	OEPC	36.45 (72.9)	36.64 (73.38)	32.0 (64.0)	32.39 (64.78)
13	OESTC	34.55 (69.1)	34.48 (68.96)	33.4 (66.7)	33.12 (66.24)
14	OELTC	33.80 (67.6)	33.32 (67.32)	33.0 (65.9)	33.37 (66.74)
15	SPCLC	68.97 (69.0)	70.79 (70.79)	72.9 (72.9)	72.12 (72.12)
	Total	553.22 (69.15)	555.8 (69.47)	549.9 (68.74)	550.1 (68.76)

Note: CE – clinical examination, SC- short consultation, PP- practical procedure, SPC – simulated patient consultation, TC- telephone consultation, MI- management interview, OE- Oral examination, PLD- peer level discourse, PM – practice management, PC- preventive care, LTC- long term care, STC- short term care, LC – long consultation

Table 2. Analysis of OSCE quality metrics for 2018 Conjoint examinations

No	Station	Cronbach's alpha if item deleted	R^2 coefficient	Intergrade discrimination	No of failures N=70	Between group variation %
1	CESC1	0.771	0.773	5.7	40	25.6
2	CESC2	0.768	0.859	6.27	22	15.03
3	PP	0.763	0.752	5.89	36	22.77
4	SPCSC1	0.761	0.722	4.88	40	20.17
5	SPCSC2	0.772	0.763	4.81	21	14.3
6	SPCTC	0.760	0.782	5.86	38	26.27
7	SPCMI1	0.760	0.841	7.32	42	27.1
8	SPCMI2	0.760	0.873	6.08	50	18.63
9	SPCMI3	0.732	0.908	7.39	23	28.63
10	PLD	0.736	0.737	5.35	31	18.5
11	OEPM	0.748	0.564	4.19	34	29.6
12	OEPC	0.743	0.634	3.89	19	18.1
13	OESTC	0.747	0.658	5.53	25	19.33
14	OELTC	0.746	0.750	6.12	33	22.39
15	SPCLC*	0.762	0.846	12.62	36	18.8

*Note: * long consultation = 22 minutes compared to other stations =11 minutes*

Table 3. Analysis of OSCE quality metrics for 2019 Conjoint examinations

No	Station	Cronbach's alpha if item deleted	R² coefficient	Intergrade discrimination	No of failure N=75	Between group variation %
1	CESC1	0.904	0.863	6.2	35	21.8
2	CESC2	0.901	0.888	6.38	27	20.9
3	PP	0.903	0.843	7.17	28	20.2
4	SPCSC1	0.910	0.83	6.59	41	25.4
5	SPCSC2	0.902	0.778	6.26	20	26.1
6	SPCTC	0.905	0.827	5.67	52	28.1
7	SPCMI1	0.904	0.721	5.55	33	12.5
8	SPCMI2	0.905	0.864	5.57	40	22.5
9	SPCMI3	0.903	0.88	7.71	27	19.4
10	PLD	0.904	0.829	6.1	44	11.1
11	OEPM	0.903	0.826	6.54	24	20.4
12	OEPC	0.904	0.748	6.45	35	33.6
13	OESTC	0.901	0.802	6.29	29	20
14	OELTC	0.901	0.832	6.15	37	17.7
15	SPCLC*	0.908	0.786	11.27*	24	30.8

*Note: * long consultation = 22 minutes compared to other stations =11 minutes*

References

[1]. Handbook for Candidates, Mentors & Examiners. Conjoint MAFP/icFRCAGP Examination Year 2023. Academy of Family Physicians of Malaysia

[2]. Norcini JJ. Setting standards on educational tests. Med. Educ 2003;37:464-9.

[3]. Barman A. Standard Setting in Student Assessment: Is a Defensible Method Yet to Come? Ann Acad Med Singapore 2008;37:957-63.

[4]. 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(1).50-57.

[5]. Friedman M. AMEE Guide No. 18. Standard setting in student assessment. Medical Teacher 2000;22:120-30.

[6]. Chan SC, Sheikh MA, Lee TW. Implementing Standard setting into the Conjoint MAFP/FRACGP Part 1 Examination - Process & Issues. Malaysian Family Physician journal 2016 Vol 11(2&3):2-8.

[7]. 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. Advances in Health Sciences Education (2006) 11:115-122. DOI 10.1007/s10459-005-7853-1.

[8]. Standard setting simplified – Borderline Regression at https://www.maximity.co.uk/blog/standard-setting-setting-simplified-borderline-regression

[9]. Reid KJ, Dodds A. Comparing the borderline group and borderline regression approaches to setting Objective Structured Clinical Examination cut scores. Journal of contemporary Medical Education 2014;2(1): 8-12

[10]. Mortaz Hejri S, Jalili M, Muijtjens AMM, Van Der Vieuten CPM. Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. J. Res Med Sci 2013;18:887-91.

[11]. Pell G et al. How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. Medical Teacher, 2010;32:10, 802-811. DOI;10.3109/014215X.2010.507716.

[12]. SY Park , SH Lee , MJ Kim, KH Ji, JH Ryu . Comparing the cut score for the borderline group method and borderline regression method with norm-referenced standard setting in an objective structured clinical examination in medical school in Korea. Journal of Educational Evaluation for Health Professions 18:25. doi: 10.3352/jeehp.2021.18.25. Epub 2021 Sep 27. September 2021.

[13]. Anneke Kramer, Arno Muijtjens, Koos Jansen et al. Comparison of a rational and an empirical standard setting procedure for an OSCE. Medical Education. 2003;37:132-139.