Education in Medicine Journal

ORIGINAL ARTICLE

Title: Comparing the Borderline Regression Standard Setting Method to a Traditional Method in an Objective Structured Clinical Examination

Authors: Nur Ain Mahat, Norhafizah Ab Manan, Rafidah Bahari, Amelah Mohammed Abdul Qader, Rohimah Md Yusoff, Maria Zalina Abdul Rahim

Submitted Date: 16-01-2024

Accepted Date: 28-03-2025

Please cite this article as: Mahat NA, Ab Manan N, Bahari R, Mohammed Abdul Qader A, Md Yusoff R, Abdul Rahim MZ. Comparing the borderline regression standard setting method to a traditional method in an objective structured clinical examination. Education in Medicine Journal. 2025; early view.

Comparing the Borderline Regression Standard Setting Method to a Traditional Method in an Objective Structured Clinical Examination

Nur Ain Mahat[1], Norhafizah Ab Manan[2], Rafidah Bahari[3], Amelah Mohammed Abdul Qader[2], Rohimah Md Yusoff[2], Maria Zalina Abdul Rahim[2]

[1]Professional Nursing Studies, International Islamic University Malaysia, Pahang, MALAYSIA
[2]Cyberjaya University College of Medical Sciences, Selangor, MALAYSIA
[3]Newcastle University Medicine Malaysia, Johor, MALAYSIA

ABSTRACT

The Objective Structured Clinical Examination (OSCE) is regarded as an effective assessment method for evaluating the clinical skills and competencies of healthcare professionals. In recent years, there has been a growing interest in employing more data-driven approaches to standard setting in the OSCE; one such method is the Borderline Regression Method (BRM). This study aims to compare the BRM with the traditional method to determine the passing score of the OSCE. This study, conducted at the University of Cyberjaya, analyzed data from 131 medical students who took the OSCE in the Final Professional Examination. Two methods were used to set the passing score for each OSCE station: the traditional method, which used a 50% passing score, and the BRM method, which used simple linear regression. In the BRM method, checklist scores were regressed onto global ratings, and the borderline global rating of 2 was used to determine the passing score. A margin of error was applied to assess the accuracy of the passing scores. For the traditional method, this was calculated as 1.96 × standard error. Of 131 candidates, the pass mark for BRM was 1.93 points higher than the traditional method (mean = 11.93 vs. mean = 10.00) for all 20 stations. The margin of error was smaller for the passing mark using BRM compared to the traditional method, with a difference of 0.09 (0.39 vs.0.48), $t = 1.96$, $p < 0.05$. The BRM resulted in a higher pass mark with a smaller margin of error, while the traditional method had the opposite effect with a broader margin of error. The study suggests considering the adoption of BRM in OSCE in medical school, but the optimal method for high-stakes clinical examinations is still inconclusive.

Keywords: *Standard setting, OSCE, Borderline regression method, Traditional Method, Comparison*

CORRESPONDING AUTHOR

Nur Ain Mahat, Professional Nursing Studies, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota , 25200 Kuantan, Pahang, Malaysia

Email: nurainmahat@iium.edu.my

INTRODUCTION

The Objective Structured Clinical Examination (OSCE) is widely regarded in the field of medical education as an effective assessment method for evaluating the clinical skills and competencies of

healthcare professionals. One critical aspect of implementing OSCE is the process of standard setting, which involves determining the passing score and performance levels that define competencies for candidates (1). The process of setting the standard in clinical examination systems is crucial in ensuring fair and reliable assessment of candidates' performance (2).

Standard settings, such as the modified Angoff or the Ebel methods, have been commonly used. These methods involve a panel of experts who carefully review and rate the difficulty of examination tasks and then collectively determine the passing score based on their professional judgment (3). This approach provides valuable insights and perspectives from experienced professionals who possess a deep understanding of the knowledge, skills, and competencies required for clinical practice. The Angoff method is reliable and has the added advantage of not requiring the expert group to be present during the conduct of the examination (4).

Nevertheless, traditional norm-reference methods in determining the cut score for passing OSCE do not take into account the level of difficulty of individual stations and have now been discarded in favor of standard setting methods (5). Choosing the right standard setting method is extremely important and may have a potentially detrimental effect on the judgment of examinees' performance (6). The right method reflects an examinee's ability well; too high a cut score may disadvantage good-enough students, while a cut score that is too lenient may allow weak students who are able to follow instructions but lack the clinical intuition in tackling real cases to pass the examination (7). The best standard setting method for an OSCE must be transparent, credible, feasible, suit the time and resources available, and, last but not least, represent the examinee's true performance (6).

In certain instances, challenges may emerge, such as resource scarcity, limited statistical expertise, and difficulty in recruiting judges, leading to unintentional bias due to a small number of examinees in the borderline range. Therefore, the modified Borderline-Group Method (BGM) is employed for its simplicity, rendering it suitable for small-scale OSCEs. This method identifies the mean scores of examinees who are exclusively rated as borderline and computes them for each station. The final OSCE cut score is determined by averaging the station mean scores. The examiner's judgment is prioritized in the final decision, which is derived from observations of actual performances (8).

Furthermore, in recent years, there has been a growing interest in employing more data-driven approaches to the standard setting in OSCE. One such method that has gained popularity is the Borderline Regression Method (BRM). A standard setting method known as the BRM is similar to the modified BGM but incorporates a linear regression approach, allowing the cut score to be set using the scores from all examinees (9).

The BRM is often used in clinical assessments such as the OSCE and has been widely accepted as a reliable method for standard setting in such assessments (7). In this method, the examiner assesses students' performance in each OSCE station using a checklist, which includes a global rating scale. The marks from all examiners from each station are collated and regressed on the ascribed global rating scale to produce a linear equation. Substitution of the global rating score representing borderline performance into the equation will then predict the cutoff passing mark for the station (1).

Between the BGM and BRM, the former is time efficient, where examiners evaluate students' performance on a global rating scale such as good, borderline, and fail, and is straightforward to implement (8). The cut scores are the median scores of examinees with borderline ratings. There is

no notable contrast in the cut score between the BGM and BRM. If the OSCE's cut scores serve as the minimal assessment of acceptable medical practice performance, examinee-centered standard setting methods are more fitting for this purpose. Additionally, as demonstrated in this study, both BGM and BRM approaches were easily implementable in practice (5). With sufficient agreement on the borderline group criteria, the use of BRM for standard setting could serve as a more defensible method to establish the OSCE stations' cut scores (1). However, it has some limitations in that achieving consensus on the borderline group is difficult, and when the number of borderline examinees is small, the cut scores may have low validity (5).

This statistical method utilizes the relationships between candidates' performance on examination items and their overall scores to determine the passing scores (3-6, 10). By analyzing the performance characteristics and difficulty of specific items, the BRM establishes a cutoff point that distinguishes competent and non-competent candidates.

This study chose the BRM because it is the method that best suited the present situation; achieving consensus on the borderline group was challenging and the study focused on using actual performance data to establish the cutoff scores through regression analysis. Since the BRM relies on a familiar statistical method and cutoff scores are easily calculated (8), it is reproducible and transparent. Its credibility, reliability, and validity in OSCE has also been well reported (4-5, 8, 10, 11-14). In its application, the method is highly feasible in different settings (5, 8) and consumes less time and resources (13). The method also provides an accurate representation of examinees' performance (6), demonstrated by the high correlation between station checklists and passing scores (14). It can be concluded that the cut score obtained from the BRM is defensible for use in OSCE (15).

This study aims to compare the traditional method with the standard setting BRM within the context of OSCE. By examining the strengths and limitations of each approach, a better understanding of the potential benefits of using more data-driven methodologies can be gained, which can contribute to the validity, reliability, and fairness of the standard setting process in OSCE.

METHODOLOGY

Examination Procedure

This study took place at the University of Cyberjaya (UoC) in 2022. Data were collected from 131 final-year Bachelor of Medicine and Bachelor of Surgery (MBBS) students who participated in the OSCE as part of their Final Professional Examination. The assessment comprised 20 stations, each of which was assessed by a single clinician examiner. Among these stations, 16 were designated as short OSCEs, encompassing key components of clinical practice such as history-taking, physical examination, data interpretation, formulating the diagnosis, performing clinical skills, patient management, and patient counseling. Additionally, the assessment featured four stations designated as long OSCEs, aimed at providing a more in-depth evaluation of those clinical skills.

Each station was meticulously constructed based on a predetermined blueprint to ensure comprehensive coverage of pertinent clinical competencies. Stations that adhered to this blueprint were thoroughly vetted to ensure that the assessment was aligned with the desired learning

outcomes and reflected the students' experience. Prior to implementation, all stations underwent a thorough review to maintain consistency and validity across the assessment process.

To uphold the standardization of assessment procedures, examiners and simulated patients received extensive training prior to the examination day. All examiners in this study underwent rigorous training to standardize their understanding of the scoring criteria. Examiners were guided through the process, with detailed explanations provided during department and central vetting sessions. The central vetting involved station leaders, including a team of experts, who evaluated all examination materials, such as station content, scenarios, checklists, and scoring rubrics, before the examination was administered. Explicit instructions were also provided on the marking scheme. Additionally, all examiners were gathered and briefed again in a 1-hour session by the station leaders prior to the examination. Meanwhile, the training sessions for the simulated patients were conducted by the station leaders, including role-playing exercises to simulate the examination and clarify expectations. Each of the simulated patients was required to sign an oath of secrecy to ensure privacy and confidentiality.

The assessment was conducted in a clinical skill training room, which provided a controlled environment conducive to evaluating clinical skills. This setting facilitated the simulation of authentic clinical scenarios while maintaining the necessary level of standardization and objectivity.

Specific time allocations were established for each type of OSCE station to ensure efficiency and fairness. Short OSCEs were allotted 6 minutes per station, allowing candidates to demonstrate their proficiency concisely. In contrast, long OSCEs were allotted 20 minutes per station, providing candidates with sufficient time to engage in more detailed and comprehensive clinical assessments.

Each student received a score and rating. The score of the student was obtained by counting the ticks on the checklist by the examiner. Each station had a different total mark, ranging from 15 to 51. For the analysis, the score for each station was standardized to 20 marks. Meanwhile, the rating was derived from the global rating provided by the examiner, categorized as follows: 1: Clear fail, 2: Borderline, 3: Clear pass, and 4: Good pass.

The Final Professional Examination is comprised of 100%, with 70% derived from 16 short OSCE stations, each carrying 20 marks, totaling 320 marks. The remaining 30% comes from 4 long OSCE stations, each also carrying 20 marks, totaling 80 marks.

The OSCE results for students who did not meet passing scores were thoroughly examined in the Examination Board Meeting to verify that their scores accurately reflected unsatisfactory performance.

Statistical Analysis

For descriptive statistics, the mean, standard deviation, and minimum and maximum values were calculated to analyze the score, while the frequency and percentage were utilized to tabulate the total number of students who failed in each station.

Two methods were used to determine the passing score for each OSCE station: traditional and BRM. The traditional method used 50% as the passing score, whereas the BRM used simple linear regression, where the scores from the checklist were regressed onto the global ratings, providing a linear equation. Then, the global rating that represents borderline performance, rated as 2, was substituted into the equation to determine the passing score for each station (5).

A margin of error around the cut scores for each station was used to determine the accuracy of the passing mark determined by each method. For the traditional method of a 50% passing mark, the margin of error was calculated using the formula 1.96 × standard error (16). Meanwhile, for cut scores derived from the BRM, a linear regression was performed for each station. The margin of error for the regression lines was calculated as follows (9, 16):

$$S_{\hat{Y}_{XO}} = S_{Y|X}\sqrt{\frac{1}{n} + \frac{(X_O + \bar{X})^2}{(n-1)S_X^2}}$$

Where,

$S_{\hat{Y}XO}$ = standard error of the regression line

$S_{Y|X}$ = standard error of the estimate

n = number of students

X_O = cut score

$\hat{X}$ = mean of scores

S_X^2 = variance of scores

After generating the standard error of the regression line, the margin of error was calculated by multiplying the standard error of the regression line by the t value at p = 0.05 and d.f. = n – 2.

Then, the margin of error of each station was added to each method. Finally, the difference in the combined margin of error was calculated. The method with a lower margin of error was considered more accurate (16).

Ethical Consideration

This study did not undergo ethical approval since it is an evaluation study. However, the Dean and examination committee were aware of this evaluation and have approved it for publication. All identifiers were removed before the data analysis.

RESULTS

A total of 131 candidates participated in the OSCE. Table 1 presents the descriptive statistics for all OSCE stations. The scores for each station were standardized to 20, and the average mean score of all 20 stations was 14.3, with the highest mean at 17.9 and the lowest mean at 9.4.

Figures 1 to 5 depict the BRM, indicating how the linear regression analysis was conducted, where the examiners' checklist scores were regressed on the examiners' global rating scores to determine the passing mark at each station, representing Stations 1 to 5.

Table 2 lists the descriptive statistics, including the standardized total scores, station pass marks, failure rates, and margin of error for the BRM and traditional methods. Overall, the pass mark for the BRM is 1.93 points higher than the traditional method (mean = 11.93 vs. mean = 10.00) for all 20 stations. Meanwhile, the failure rate for the BRM is, on average, 9.7% higher than the traditional method (20.4% vs.10.7%). The average failure rate was calculated by taking the mean of failure percentage across all 20 stations. Furthermore, the margin of error is smaller for the passing mark using the BRM compared to the traditional method, with a difference of 0.09 (0.39 vs.0.48), $t = 1.96$, $p < 0.05$. Based on the t distribution table, the t value is 1.96 at $p = 0.05$ and d.f. = 129.

DISCUSSION

The BRM set the pass mark at 1.93 points higher than the traditional method (mean = 11.93 vs. 10.00). This difference suggests that the BRM may set a higher standard for passing, reflecting its potentially more stringent criteria based on regression analysis of borderline performance. This is because BRM aims to ensure only those who are truly competent pass, while the traditional method might pass candidates who are borderline or underperforming. This finding is supported by previous studies, which also observed similar results (5, 11).

The failure rate associated with the BRM was, on average, 9.7% higher than that of the traditional method (20.4% vs. 10.7%). It indicates that a greater proportion of examinees failed under the BRM, which is coherent with the finding that the BRM resulted in a higher passing score.

The higher failure rate highlights that the BRM is more cautious in passing determinations, potentially due to its reliance on statistical modeling of borderline performance. The same results were noted in prior studies, where the overall cut scores of the BRM were higher than those of the norm-referenced method, resulting in a higher number of failed examinees (5, 11). Nevertheless, this contradicts a prior study, which revealed that the BRM showed the highest pass rates due to a lower passing mark (17), while in another study, the traditional method resulted in the highest failure rate (11).

Moreover, the margin of error for the passing mark was smaller for the BRM compared to the traditional method, with a difference of 0.09 (0.39 vs. 0.48). The reduced margin of error for the BRM suggests greater precision in determining the passing mark, which may contribute to its higher accuracy in reflecting true competence levels. This is consistent with a prior study where the root mean square error in measuring reliability was small, indicating that BRM is a reliable method of standard setting for OSCE (1). The statistically significant p-value indicates that the differences observed between the BRM and the traditional methods are statistically significant, supporting the

robustness of the BRM in setting passing scores with greater precision, as supported by Park et al. (5).

Providing a variety of indices is crucial in measuring the validity of the OSCE in using borderline regression (11). In this study, the variations of the cutoff passing mark in each station strengthen the validity of BRM in analyzing the level of competencies of the students, the level of station construction, and the rating of the expert examiners. Hejri et al. also reported that BRM is much more convenient and less resource-consuming compared to other procedures like the Angoff method (1). A study revealed that when comparing the BRM, BGM, and Angoff methods, pass marks did not significantly differ between the BGM and BRM. However, the Angoff method consistently established significantly lower pass marks compared to the BRM (12).

This study found that the BRM, compared to the traditional method, provided a better cutoff passing mark for each station than just the arbitrary 50%. Kilminster and Roberts reported similar results when they tried the method during their Year 3 OSCE examinations 20 years ago (18), which were replicated by Pell and Roberts and many researchers since (19). Although the borderline methods generally are the preferred method for standard setting in many institutions, the BRM is able to inform examiners of the level of discrimination between weaker and stronger students (20).

Nevertheless, not one standard setting method is perfect. Dwyer (3) argued that determining a pass mark only after the examination has been conducted, as in the case of BRM, may not be suitable for all OSCE. Furthermore, there is evidence that this standard setting method may not be as accurate in smaller examinations (< 50 examinees) (15). In our situation, setting the marks after the examination was not an issue, and the cohort was large enough for BRM to be accurate.

The strength of this study is the use of a well-accepted standard setting method, the BRM. However, we acknowledge that applying one method of standard setting may not be adequate with high stake examinations. The limitation of this study may be improved with the utilization of more than one standard setting method, as suggested by Dwivedi et al. (11). Moreover, the analysis was conducted in only one cohort, making it insufficient for generalization.

CONCLUSION

This research was conducted to compare the standard setting BRM with the traditional method in OSCEs. The findings indicate that employing the BRM led to a higher pass mark, with a smaller margin of error, convenience, and required fewer resources. In contrast, the traditional method resulted in a broader margin of error. While the traditional method can be simple, it does not consider the complexity of the OSCE or the varying difficulties of the stations. BRM offers a more data-driven and objective approach to determining the cutoff, which may improve the fairness and accuracy of the assessment without the need for additional resources to reassess performance.

Nonetheless, this study is limited by its sample size of 131 medical students and the use of a single cohort from one institution, which may restrict the generalizability of the findings to broader populations or different educational settings. Future research should include larger, multi-institutional cohorts to enhance the validity and applicability of the findings.

Based on the study outcomes, we suggest considering the adoption of BRM in OSCEs for high-stake clinical examinations, such as the Final Professional Examination, due to its objective, data-driven

approach and its ability to reduce the margin of error compared to traditional methods. Since high-stake assessments require both fairness and accuracy, BRM provides a more refined pass mark determination by incorporating real-time student performance data rather than relying solely on fixed thresholds. Further validation through larger-scale and multi-institutional studies is recommended to strengthen its applicability in clinical assessments.

A few recommendations can be considered, such as introducing further standardizations in examiner training, with a specific focus on the criteria for borderline students. This proactive measure ensures a more consistent and objective evaluation process. However, determining the optimal method for high-stakes clinical examinations remains inconclusive.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the Faculty of Medicine, University of Cyberjaya, for the permission to utilize data from the Final Professional Exam in the course of this study. This research did not receive any funding from external sources. The authors declare no conflict of interest.

REFERENCES

1. Hejri SM, Jalili M, Muijtjens AM, Van-Der-Vleuten CP. Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. J Res Med Sci. 2013;18(10):887-891. PMID: 24497861; PMCID: PMC3897074

2. Dewan P, Khalil S, Gupta P. Objective structured clinical examination for teaching and assessment: evidence-based critique. Clin Epidemiology Glob Health. In press 2023. https://doi.org/10.1016/j.cegh.2023.101477

3. Dwyer T, Wright S, Kulasegaram KM, Theodoropoulos J, Chahal J, Wasserstein D, et al. How to set the bar in competency-based medical education: standard setting after an objective structured clinical examination (OSCE). BMC Med Educ. 2016;16(1):1-7. https://doi.org/10.1186/s12909-015-0506-z

4. Milan FB, Grochowalski JH. A resource efficient and reliable standard setting method for OSCEs: borderline regression method using standardized patients as sole raters in clinical case encounters with medical students. Med Teach. 2022;44(8):878–885. https://doi.org/10.1080/0142159X.2022.2041586

5. Park SY, Lee S-H, Kim M-J, Ji K-H, Ryu JH. 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. J Educ Evaluation Health Prof. 2021;18(25):1-10. https://doi.org/10.3352/jeehp.2021.18.25

6. Malau-Aduli BS, Teague PA, D'Souza K, Heal C, Turner R, Garne DL, et al. A collaborative comparison of objective structured clinical examination (OSCE) standard setting methods at

Australian medical schools. Med Teach. 2017;39(12):1261–1267. https://doi.org/10.1080/0142159X.2017.1372565

7. Hays R. Standard setting. Clin Teach. 2015;12(4):226 – 230.
8. Smee S, Coetzee K, Bartman I, Roy M, Monteiro S. OSCE standard setting: three borderline group methods. Med Sci Educ. 2022;32(6):1439–1445. https://doi.org/10.1007/s40670-022-01667-x
9. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. Belmont, CA: Duxbury Press; 1988.
10. Moreno-López R, Hope D. Can borderline regression method be used to standard set OSCEs in small cohorts? Eur J Dent Educ. 2022;26(4):686–691. https://doi.org/10.1111/eje.12747
11. Dwivedi NR, Vijayashankar NP, Hansda M, Dubey AK, Nwachukwu F, Curran V, et al. Comparing standard setting methods for objective structured clinical examinations in a Caribbean medical school. J Med Educ Curric Dev. 2020;7:1-10. https://doi.org/10.1177/2382120520981992
12. Boursicot KAM, Roberts TE, Pell G. Using borderline methods to compare passing standards for OSCEs at graduation across three medical schools. Med Educ. 2007;41(11):1024–1031. https://doi.org/10.1111/j.1365-2923.2007.02857.x(2007)
13. Kaufman DM, Mann KV, Muijtjens AMM, Van Der Vleuten CPM. A Comparison of standard-setting procedures for an OSCE in undergraduate medical education. Acad Med. 2000;75(3):267–271. http://journals.lww.com/academicmedicine
14. Kramer A, Muijtjens A, Jansen K, Düsman H, Tan L, Van Der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Med Educ. 2003;37(2):132–139. https://doi.org/10.1046/j.1365-2923.2003.01429.x
15. Homer M, Fuller R, Hallam J, Pell G. Setting defensible standards in small cohort OSCEs: understanding better when borderline regression can 'work.' Med Teach. 2020;42(3):306–315. https://doi.org/10.1080/0142159X.2019.1681388
16. 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. 2006;11:115-122. https://doi.org/10.1007/s10459-005-7853-1
17. Kamal D, Sallam M, Gouda E, Fouad S. Is There a "Best" Method for Standard Setting in OSCE Exams? Comparison between Four Methods (A Cross-Sectional Descriptive Study). J Med Educ. 2020;19(1): e106600. http://doi.org/10.5812/jme.106600
18. Kilminster S, Roberts T. Setting standards for OSCEs: Trial of borderline approach. Adv Health Sci Educ Theory Pract. 2004;9:20 –109.

19. Pell G, Roberts TE. Setting standards for student assessment. Int J Res Method Educ. 2006;29(1):91–103.

20. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. Med Teach. 2010;32(10):802-811. http://doi.org/10.3109/0142159X.2010.507716