International Journal of Medical Science and Health Research

Vol.9, No. 04; 2025

ISSN: 2581-3366

Precision in Clinical Evaluation: Integrating Borderline Regression Into OSCEs at Defence Services Medical Academy (DSMA)

Tun Tun Naing[1], Thaung Tan Soe[1],Tayzar Hein[1],Zaw Phyo1,Pyi Hein Kyaw[1], Zin Min Htike[1]
Darli Aung[2], Khun Tun Hlaing[1] , Khin Aung Htun[1] & Ye Phyo Aung[1]
[1.] Defence Services Medical Academy, Myanmar
[2.]No.(1)Defence Services Obstetrics, Gynaecology& Children Hospital (300-Bedded), Myanmar
Correspondence: Tun Tun Naing, Defence Services Medical Academy, No.94, D-1,
Pyay Road, Mingaladon Township, Yangon, Myanmar, Tel: +95 95053402, Email:

doi: 10.51505/ijmshr.2025.9402 URL: http://dx.doi.org/10.51505/ijmshr.2025.9402

Received: Jun 16, 2025 Accepted: July 02, 2025 Online Published: July 12, 2025

Abstract

Background: The Defence Services Medical Academy (DSMA) uses Objective Structured Clinical Examinations (OSCEs) for assessing clinical competencies, including Obstetric and Gynaecology High-Stakes OSCE. However, the optimal method for assessing clinical performance remains contentious.

Objective: This paper aims to compare the results from two assessment methods—checklist scoring (CL) and global rating score with borderline regression (BRM)—to see if there are significant differences in how students' clinical skills are evaluated.

Methods: This study investigated data from thirty three students who participated in 15 OSCE stations at DSMA. Each student was evaluated utilising both comprehensive ratings and checklist scores. The passing thresholds for each station were determined using the Borderline Regression method. Results from both approaches were compared using paired T-tests to find any statistically significant differences in performance evaluation.

*Results***:** The study found no significant differences between Borderline Regression and checklist methods in performance evaluations, with all stations' p-values being either 1.000 or closely approximated 1.000.

Discussion: The study found that Borderline Regression can be effectively combined with checklist scoring to improve the fairness and accuracy of passing thresholds without disrupting the assessment process.

Conclusion: The study suggests that integrating both assessment methods at DSMA can enhance the reliability and fairness of OSCE evaluations, but further research is needed to understand their long-term impact.

Vol.9, No. 04; 2025

ISSN: 2581-3366

Keywords: Borderline Regression Method, Traditional checklist scoring & Objective Structured Clinical Examinations

1. Introduction

Objective Structured Clinical Examinations (OSCEs) are a crucial part of medical educations clinical skills assessment, because of their ability to replicate real-world clinical settings and evaluate a wide range of competencies. The Defence Services Medical Academy(DSMA) has implemented Outcome-Based Education (OBE) for eight years to ensure that Year 5 undergraduate students achieve specific learning objectives.

Currently, DSMA's Obstetrics and Gynecology High-Stakes OSCE for Year 5 students consist of several stations. The purpose of each station is to assess a different aspect of clinical competency in gynecology and obstetrics. The activities carried out at these stations such as collecting medical histories conducting physical exams and carrying out clinical procedures are crucial for determining whether or not students are prepared to continue their education and practice on their own in the future.

An important problem at DSMA is that assessors' use of OSCE scoring methods varies widely. The checklist scoring method which attempts to offer a systematic assessment of students proficiency in completing predetermined tasks has drawn criticism for being overly strict and neglecting to take into consideration the subtleties of clinical performance. The global rating system on the other hand is more subjective and is based on the examiners overall evaluation of the students performance across a variety of domains such as technical proficiency clinical reasoning and communication abilities. Especially when assessors have varying levels of experience or interpret expected performance standards differently this subjectivity has led to disparities.

Many medical schools have looked into various methods for creating OSCE passing standards in order to get around these challenges. Obviously, borderline regression establishes a more dependable passing threshold by comparing checklist scores with global ratings using regression analysis. This method seeks to create a more equitable and trustworthy evaluation system by determining a borderline passing score using both subjective and objective assessments.

Research has shown that combining global rating systems with checklist scoring results in a more comprehensive evaluation of students performance. Aduli Malau et al. (2012) found that combining the two methods resulted in clinical competency assessments with fair and accuracy(Malau-Aduli et al., 2012). Furthermore, the usefulness of the OSCE as an assessment tool has been questioned due to its limited capacity to precisely gauge the complexity of clinical practice. Hejri et al. (2013) observed that although OSCEs offer standardized and structured evaluations, they might not fully capture the variety of abilities needed in actual clinical settings like flexibility and judgment. Moreover, assessor bias presents another review issue in OSCE ratings(Hejri et al., 2013). Examiners judgments are often tainted by personal bias despite the

Vol.9, No. 04; 2025

ISSN: 2581-3366

high standards of training and evaluation which compromises the assessments objectivity and quality. Research by Setyonugroho et al. (2015) provided a case study to demonstrate the impact of these biases and suggested that using multiple examiners and consistent scoring systems may help lessen the bias(Setyonugroho et al., 2015). One technique to reduce any additional biases that might compromise the assessments objectivity is to use techniques like double-blind assessments in which the assessors remain anonymous to one another while providing ratings.

In spite of these difficulties, OSCEs continue to be among the most popular and trustworthy ways to assess clinical competencies in medical school. OSCEs may become more accurate and equitable as standardized evaluation techniques like Borderline Regression are developed further. In order to support the major implications for the goals of an outcome-based education and to guarantee that measurements are valid and reliable, DSMA recognises the importance of enhancing these skills.

Throughout the previous eight years, DSMA has made outcome-based education (OBE) a priority. OBE seeks to ensure that students meet predetermined learning objectives by emphasizing the application of knowledge in real-world situations. Through the coordination of teaching learning and assessment approache, OBE guarantees that students possess the skills required for their future employment as medical professionals. The success of OSCEs notwithstanding, there are still challenges in accurately assessing clinical competencies.

Therefore, the clinical competencies of Year 5 students at DSMA are assessed through the Obstetric and Gynaecology High-Stakes OSCE. The inconsistent implementation of the checklist and global rating methodologies by assessors has resulted in issues. This study aims to evaluate the reliability of two methods—checklist scoring (CL) and Borderline Regression (BRM)—in assessing student performance during the Obstetric and Gynaecology High-Stakes OSCE at DSMA. The primary objective is to determine if the Borderline Regression method, which adjusts passing levels based on individual station performance, provides a more precise assessment of student competencies compared to the traditional checklist scoring method. This study examines whether the integration of both methodologies could enhance the equity and reliability of the evaluation system for this significant examination.

2. Materials and Methods

All the students who was attending the Year-5 participated in the study. Thirty-three students who were enrolled in 15 OSCE stations at the Obstetric and Gynecology High Stake OSCE for Year 5 students at DSMA provided performance data for this study. To ascertain whether there are any appreciable differences in student performance assessments, the study will compare the conventional checklist scoring method (CL) and global rating with the Borderline Regression method (BRM). Prior to data collection, ethical clearance was obtained from the DSMA internal ethics review board.

Vol.9, No. 04; 2025

ISSN: 2581-3366

2.1 Data Collection

Every student finished fifteen OSCE stations each of which evaluated a distinct clinical skill. For every student information was gathered on both global ratings and checklist scores. The passing thresholds for every station were determined using the Borderline Regression technique. The scores from the two evaluation techniques were then compared using paired T-tests.

2.2 Statistical Analysis

- **Borderline Regression**: Based on student performance data, borderline regression was utilized to determine the cutoff points for passing scores at global rating score of each station. Because of the varying degrees of difficulty of the tasks evaluated the cutoffs differed by station.
- **Paired T-test**: For each station the checklist scores and the Borderline Regression scores were compared using the Paired T-test to see if there were any notable variations between the two scoring systems.

3. Results

3.1 Borderline Regression Analysis Results

The Borderline Regression method was applied to each of the 15 OSCE stations, calculating the cutoff points for passing scores. These thresholds varied across stations, reflecting the differences in the complexity of each task. The following cutoff points were calculated for each station:

Vol.9, No. 04; 2025

ISSN: 2581-3366

Table 1: Cut-off Point at Different OSCE stations after applying Borderline Regression Method

No.	Type of Station	Stastion Number	Task to be Performed	Cut-off Point
1	History	1	Antenatal History taking of a grandmultip woman at booking visit	6.1
2	Examination	1	Antenatal examination	6.3
3	Explanation	1	Communication and explanation of management of a pregnant woman with severe preeclampsia	5.4
4	Procedure	1	Demonstration of placental delivery by controlled cord traction method	5.7
5	History	1	Focused history taking of a pregnant woman presenting with abdominal pain at term	5.7
6	Examination	2	Postnatal examination	5.8
7	Explanation	2	Antenatal health education	5.4
8	Procedure	2	Episiotomy repair	6.3
9	History	3	Foucused history taking of a woman presenting with vaginal discharge	5.3
10	Examination	3	Pelvic examination of a woman presenting with vaginal discharge with mannequin	5.4
11	Explanation	3	Explain and communicate with a patient who has had laparotomy and salpingectomy for ruptured ectopic pregnancy	5.4
12	Procedure	3	Manual Vacuum Aspiration procedure	5.6
13	History	4	Focused history taking of a woman with pelvic mass	6.2
14	Explanation	4	Explanation of proper use of condom	5.5
15	Procedure	4	Taking cervical smear	5.4

These cutoff points reflect the relative difficulty of each OSCE station. For example, stations assessing more complex skills, such as physical examinations and procedures, had higher cutoff points, while stations focused on history-taking and explanation had slightly lower cutoff points.

Vol.9, No. 04; 2025

ISSN: 2581-3366

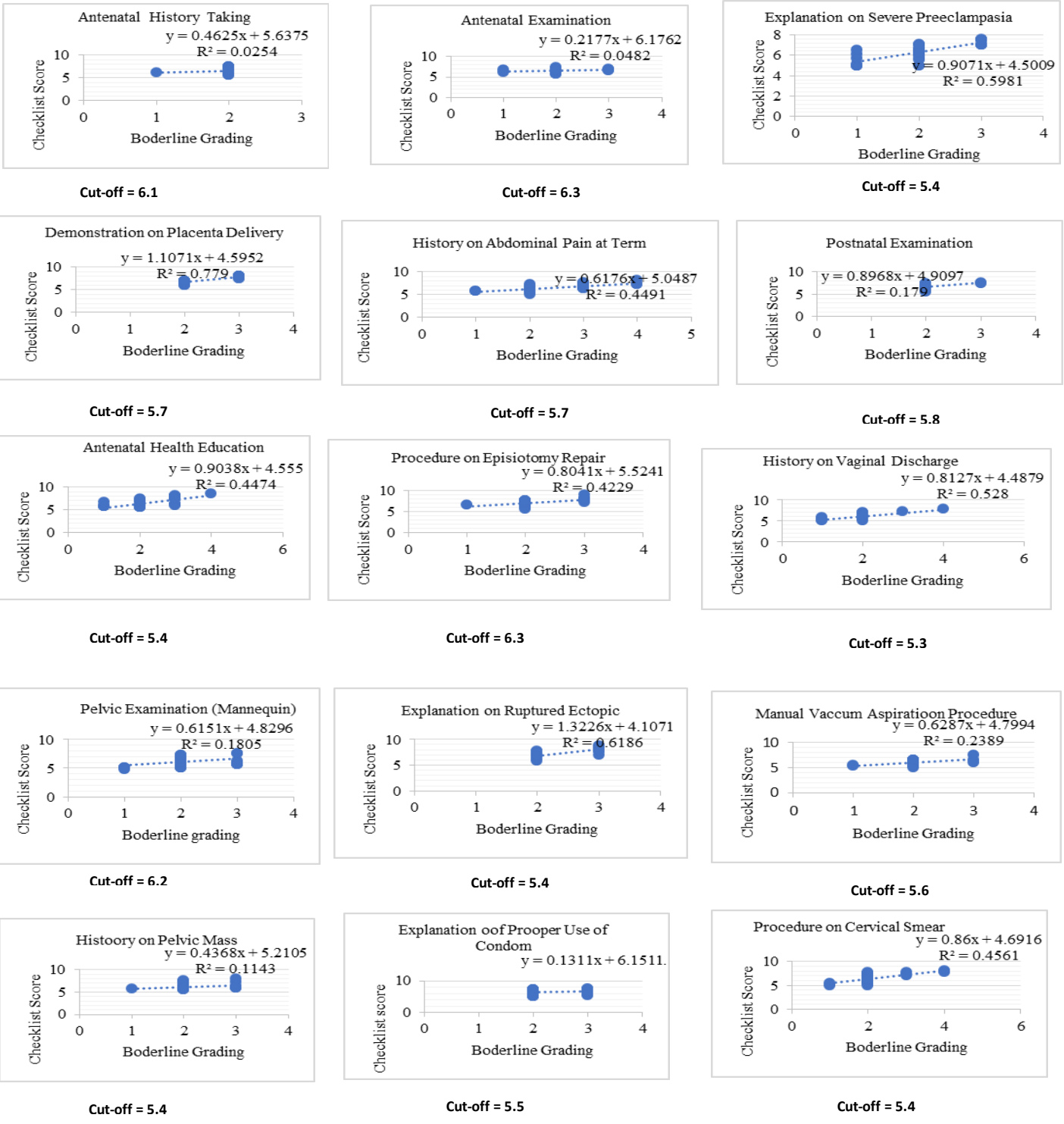

Figure 1 Borderline Regression Analysis on 15 OSCE Stations

3.2 Paired T-Test Results

The Paired T-test was conducted to compare the scores from the checklist (CL) method and the Borderline Regression method (BRM) for each station. The results for all 15 stations are summarized in the table below:

Table 2: Comparison of Borderline Regression Method (BRM) and Checklist Score Method (CL)

No.	Type of Station	Stastion Number	Mean (CL)	SD (CL)	Mean (BRM)	SD (BRM)	P-value
1	History	1	6.5485	0.50505	6.5485	0.08051	1.000
2	Examination	1	6.6182	0.50505	6.6182	0.08051	1.000
3	Explanation	1	6.3152	0.77585	6.3151	0.59999	1.000
4	Procedure	1	7.2121	0.61276	7.2120	0.54082	0.998
5	History	1	6.5833	0.69582	6.5833	0.46631	1.000
6	Examination	2	6.7576	0.51358	6.7577	0.21730	0.999
7	Explanation	2	6.6364	0.98388	6.6365	0.65816	0.999
8	Procedure	2	7.3273	0.62062	7.3272	0.40357	1.000
9	History	3	6.0394	0.64756	6.0394	0.47054	1.000
10	Examination	3	6.0970	0.88193	6.0971	0.37476	0.999
11	Explanation	3	7.2333	0.82146	7.2332	0.64610	0.999
12	Procedure	3	6.0758	0.50685	6.0759	0.24776	0.999
13	History	4	6.1636	0.60043	6.1635	0.20296	0.999
14	Explanation	4	6.4848	0.80161	6.4848	0.06629	1.000
15	Procedure	4	6.5939	0.88669	6.5940	0.59883	0.999

The p-values for all stations are either **1.000** or **close to 1.000**, indicating that there is no statistically significant difference between the scores from the checklist and the Borderline Regression method. This suggests that both scoring systems provide very similar evaluations of student performance across all 15 OSCE stations.

3.3 Interpretation of Results

These results provide strong evidence that combining both checklist scoring and Borderline Regression methods in OSCE assessments can enhance the reliability of clinical evaluations at

Vol.9, No. 04; 2025

ISSN: 2581-3366

DSMA. The consistent results between the two methods suggest that assessors may confidently use either or both methods for accurate and fair student assessments.

The results of the Borderline Regression method and the paired t-test indicate that both the checklist scoring system and the Borderline Regression method provide consistent and reliable evaluations of student performance. The absence of significant differences between the two scoring methods across all stations suggests that the Borderline Regression method does not alter the students' overall performance evaluation when compared to the checklist method.

This finding is significant because it implies that using the Borderline Regression method could be a useful strategy to improve the passing score computations without sacrificing the assessments accuracy or fairness. In addition the comparatively low standard deviations found in the Borderline Regression scores (in contrast to the checklist scores in certain stations) imply that the Borderline Regression method might provide a marginally more reliable grading strategy especially with regard to scoring stability. These findings offer compelling evidence that the reliability of clinical evaluations at Defence Services Medical Academy(DSMA can be improved by integrating the Borderline Regression and checklist scoring techniques in OSCE assessments. Given the consistent outcomes of the two approaches assessors can use either or both with confidence to provide fair and accurate student evaluations.

4. Discussion

The comparison of two assessment methods—checklist scoring (CL) and Global Rating with Borderline Regression (BRM)—in the context of Objective Structured Clinical Examinations (OSCEs) at Defence Services Medical Academy provides significant insights into their effectiveness for evaluating clinical competencies. The analysis suggests that both methods produce comparable results, contributing to the overall reliability and fairness of the assessment process.

4.1 Consistency Between Checklist Scoring and Borderline Regression Method

There were no statistically significant differences between the scores generated by the checklist technique and the borderline regression approach across all OSCE sites. The borderline regression approach, which modifies pass scores according on student performance, does not significantly alter the assessment of pupils. Both methodologies ultimately assess the same abilities, but via somewhat different procedures. This outcome is not really unexpected, to be frank. This aligns with prior research; Boulet et al. (2003) shown that integrating global ratings with checklist scoring results in more equal and balanced evaluation findings. In summary, both scoring techniques perform comparably well in this setting (Boulet et al., 2003). This study advances prior research by establishing that Borderline Regression is a technique as valid as existing methodologies, without disrupting the current system. This study advances prior research by establishing that Borderline Regression is a technique as valid as existing methods, without disrupting the current system (Sim et al., 2015). Irrespective of the particular

Vol.9, No. 04; 2025

ISSN: 2581-3366

methodology, dependability in this context signifies a fundamental agreement in evaluation, indicating that both strategies may be modified to more accurately depict student achievement. Furthermore, Van Der Vleuten and Schuwirth (2005) assert that while comparable score results across methodologies should not inherently instill trust in an evaluation system, such systems need to be transparent and amenable to change. Van Der Vleuten and Schuwirth (2005). Read et al. (2015) endorse this notion, suggesting that the integration of global rating scales with checklists enhances fairness in OSCE assessments (Read et al., 2015). Hodges (2003) emphasizes the need of balancing subjectivity and objectivity in clinical skills assessments to guarantee that evaluations are both justifiable and accurately indicative of ability (Hodges, 2003).

4.2 Impact of Borderline Regression on Passing Scores

When determining passing scores, the Borderline Regression method provides a flexible approach that differs depending on the OSCE station. These cutoffs were adjusted for each station's degree of difficulty, with higher passing scores needed for more difficult tasks like physical exams and procedures. Thus, this approach offers a more sophisticated way to evaluate students' clinical skills. To illustrate the disparities in task difficulty, Procedure 2 had a cutoff score of 6 points, while History 3 had a cutoff of 5 points.

The findings are consistent with Elabd et al.'s work (2023), who proposed modifying passing thresholds using data-informed techniques like Borderline Regression to create a more trustworthy standard for clinical evaluations (Elabd et al., 2023). This approach takes task complexity into account, resulting in an assessment framework that more accurately captures the diversity of clinical settings where different circumstances call for different skill sets. Besar et al. (2012) also support adaptive assessments, suggesting that flexible methods better accommodate the nuances of clinical training (Besar et al., 2012). Further support for adaptive standard setting is found in the work of Yudkowsky(2020), who argued that pass/fail decisions should reflect task difficulty and variability in rater judgments (Yudkowsky, 2020).

4.3 Implications for Student Performance and Fairness

Flexible scoring systems like Borderline Regression really do a better job at capturing students' real clinical abilities than sticking to rigid cutoffs. Students benefit from fairer evaluations when assessment tools account for varying task complexity and rater expectations. In addition, this method may motivate students to improve in areas that are more clinically demanding, fostering a growth mindset. Furthermore, Hamdy and colleagues (2006) made it clear: fairness in assessments isn't just theory; it actually boosts student learning and ramps up engagement(Hamdy et al., 2006). Bringing in more flexible, dynamic standards could seriously improve both how students perform and how much they trust the whole evaluation process.

Vol.9, No. 04; 2025

ISSN: 2581-3366

4.4 Clinical Competency Evaluation and OSCE Standards

OSCEs are considered a key element in medical education because they allow students to experience realistic clinical scenarios, which is an integral component of effective education. Even with OSCEs as a standardized method of assessment, the assessment value comes down to how they are scored. While checklist scoring is objective it may not adequately capture the complexity involved in clinical reasoning and adaptability which is often required in practice. The borderline regression method provides an alternative method of determining passing standards based on the statistical analysis of performance data.

Hejri et al. (2013) identified that even though OSCEs are more structured than real world clinical environments, they are inherently more variable with more uncertainty than OSCEs. Investigating the use of borderline regression and checklist scoring jointly presents an opportunity to reduce the gap between OSCEs as assessments, the variability of clinical tasks, and the actual competency of students (Hejri et al., 2013).

Furthermore, reflecting the principles of OBE in assessment systems could significantly enhance the relationship between assessment and OBE principles through the use of Borderline Regression. Martin et al. (2020) explain that OBE is more than just a process that adheres to regulations; it is a commitment to ensuring that students actually achieve the given outcomes (Martin et al., 2020). Using dynamic passing scores across a range of student performance data supports OBE by making sure students are judged based on their own degree of competence, rather than a fixed standard. Pell et al. (2010) and colleagues confirm the importance of criterion-referenced standard setting as a means of ensuring fairness for different students with distinct backgrounds, both in their initial education and in terms of their capabilities (Pell et al., 2010). Norcini et al. (2011) argue that OSCE assessments must also be allowed to develop in line with growing complexity, so that competence can be validly and reliably shown. In terms of the above, Borderline Regression is hardly just a technique: it is a solid lever for valid, fair and outcome based assessment (Norcini et al., 2011).

4.5 Limitations and Areas for Further Research

There are many limitations to the findings of the study that may affect dependency on the information presented. First, the sample size of 33 was small and thus impacted the generalizability of results, and larger cohorts will be needed to evaluate the reproducibility of the findings and the use of the borderline regression approach in other settings.

Second, although the paired t-test found no significant difference in scoring systems, other statistical analyses might be better suited to display the peculiarity in differential performance outcomes reflected in the t-test. Also, future research could employ inter-rater reliability measures and more advanced regression modelling to evaluate the effectiveness of using both the OG and SMD methods of evaluation in future studies. Studies by Sim et al. (2015) have shown

that using a mix of different scoring types can help reduce differences between raters and improve the reliability of OSCES evaluations (Sim et al., 2015).

In summary, future research should look at how borderline regression over time can impact student results, the fairness of exams, and overall academic performance. This longitudinal study would be particularly beneficial in determining the longer-term sustainability of this method and heralding its methods to bolster the medical education aspects of education in various political systems across the globe.

5. Conclusion

The research shows that both checklist scoring and the Borderline Regression Method (BRM) work equally well for Objective Structured Clinical Exams (OSCEs) at the Defence Services Medical Academy. Both methods achieve similar results and support decision-making in assessment, leading to fairness and reliability in assessing students' clinical competencies. BRM has the potential for flexibility, given its ability to support a modified passing score for different levels of difficulty and its representation of more relevant clinical circumstances. Although there is a small sample, BRM also has the ability to make traditional models of assessment potentially more meaningful and to better accommodate the complexities of variability in clinical reasoning and performance.

The Defence Services Medical Academy (DSMA) should incorporate the Borderline Regression Method (BRM) into OSCE evaluations to ensure reliable and equitable assessments. BRM's flexibility and alignment with outcome-based education make it suitable for modern medical training. Faculty should be trained on BRM usage to ensure uniform use. Integrating BRM with systematic examiner feedback can improve assessment quality and clinical reasoning. Increased sample size and ongoing research can further validate BRM's efficacy across various therapeutic fields. The broader use of BRM could enhance DSMA's goal of producing competent, reflective, and clinically adept graduates.

6. References

Besar, M., Siraj, H., Manap, R., Mahdy, Z., YDSMAn, M., KDSMArudin, M., & MohDSMAd, N. (2012). Should a Single Clinician Examiner be used in Objective Structure Clinical Examination? *Procedia - Social and Behavioral Sciences*, *60*, 443–449. https://doi.org/10.1016/j.sbspro.2012.09.404

Boulet, J. R., De Champlain, André F., & and McKinley, D. W. (2003). Setting defensible performance standards on OSCEs and standardized patient examinations. *Medical Teacher*, *25*(3), 245–249. https://doi.org/10.1080/0142159031000100274

Elabd, K., Abdul-Kadir, H., Alkhenizan, A., & Alkhalifa, M. K. (2023). A Comparison of the Checklist Scoring Systems, Global Rating Systems, and Borderline Regression Method for an Objective Structured Clinical Examination for a Small Cohort in a Saudi Medical School. *Cureus*. https://doi.org/10.7759/cureus.39968

Hamdy, H., Prasad, Kameshwar, Anderson, M. Brownell, Scherpbier, Albert, Williams, Reed, Zwierstra, Rein, & and Cuddihy, H. (2006). BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher*, *28*(2), 103–116. https://doi.org/10.1080/01421590600622723

Hejri, S. M., Jalili, M., & Muijtjens, A. M. M. (2013). Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. *Journal of Research in Medical Sciences*.

Hodges, B. (2003). OSCE! Variations on a theme by Harden. *Medical Education*, *37*(12), 1134–1140. https://doi.org/10.1111/j.1365-2923.2003.01717.x

Malau-Aduli, B. S., Mulcahy, S., Warnecke, E., Otahal, P., Teague, P.-A., Turner, R., & Vleuten, C. V. D. (2012). Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. *Creative Education*, *03*(06), 937–942. https://doi.org/10.4236/ce.2012.326142

Martin, R. D., Ngo, N., Silva, H., & Coyle, W. R. (2020). An Objective Structured Clinical Examination to Assess Competency Acquired During an Introductory Pharmacy Practice Experience. *American Journal of Pharmaceutical Education*, *84*(4), 7625. https://doi.org/10.5688/ajpe7625

Norcini, J., Anderson, Brownell, Bollela, Valdes, Burch, Vanessa, Costa, Manuel João, Duvivier, Robbert, Galbraith, Robert, Hays, Richard, Kent, Athol, Perrott, Vanessa, & and Roberts, T. (2011). Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, *33*(3), 206–214. https://doi.org/10.3109/0142159X.2011.551559

Pell, G., Fuller, Richard, Homer, Matthew, & and Roberts, T. (2010). How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. *Medical Teacher*, *32*(10), 802–811. https://doi.org/10.3109/0142159X.2010.507716

Read, E. K., Bell, C., Rhind, S., & Hecker, K. G. (2015). The Use of Global Rating Scales for OSCEs in Veterinary Medicine. *PLOS ONE*, *10*(3), e0121000. https://doi.org/10.1371/journal.pone.0121000

Setyonugroho, W., Kennedy, K., & Kropmans, T. (2015). Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. *Patient Education and Counseling*, *98*. https://doi.org/10.1016/j.pec.2015.06.004

Sim, J. H., Abdul Aziz, Y. F., Vijayanantha, A., Mansor, A., Vadivelu, J., & Hassan, H. (2015). A Closer Look at Checklist Scoring and Global Rating for Four OSCE Stations: Do the Scores Correlate Well? *Education in Medicine Journal*, *7*(2). https://doi.org/10.5959/eimj.v7i2.341

Van Der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

Yudkowsky, R. (Ed.). (2020). *Assessment in health professions education* (2nd edition). Routledge.