# Impact of Task-based Checklist Scoring and Two Domains Global Rating Scale in Objective Structured Clinical Examination of Pharmacy Students

**Sajesh Kalkandi Veettil*** **and Kingston Rajiah**

Department of Pharmacy practice, International Medical University, Kuala Lumpur-57000, Malaysia.

## ABSTRACT

**Introduction:** Global ratings are station-independent scales identifying general areas of competence, such as communication, rapport and similar constructs that may not be well captured in a checklist item. Global ratings seem to have psychometric possessions that are as noble as or healthier than those of checklists, whether used in conjunction with a checklist scoring system or on their own. **Methods:** This was a retrospective study. The results of second year pharmacy students' end of semester froma private University in Malaysia have been used. **Results:** There were 164 participating students.There was a significant positive Pearson correlation between the two scales ($p < 0.05$); however $R^2$ value was not satisfactory. The $R^2$ coefficient is the proportional change in the dependent variable (checklist score) due to change in the independent variable (global grade). This allowed us to determine the degree of linearity between the checklist score and the global rating score for each station, with the expectation that higher global ratings should generally correspond with higher checklist scores. **Conclusion:** Since the global rating scale exactly represents the overall criterion in the checklists, the reasons for the unsatisfactory correlation may be due to improper standardization of global scale and checklist among markers or poor understanding of criteria to use in the global rating system. It is mandatory to re-look on the mentioned problems as well as re-writing the stations or checklists or criteria for global rating.

**Key words:** Global rating scale, Malaysia, OSCE, Pharmacy students, Pharmacy curriculum, Task-based checklist.

## INTRODUCTION

Recently several opinions have forwarded to sustenance an amplifiedusage of global ratings.[1] Global ratings are station-independent scales identifying general areas of competence,[2] such as communication, rapport and similar constructs that may not be well captured in a checklist item.[3] Global ratings seem to have psychometric properties that are as gallant as or better than those of checklists,[4] whether used in conjunction with a checklist scoring system or on their own.[5] The psychometric properties is a psychological test relate to the data that has been collected on the test to determine how well it measures the construct of interest.[6] Moreover, related works proposing that clinicians with greaterechelons of proficiency do not decipherissues in clinical sceneriesby tactics reproduced in a checklist rating system.[7] There is a clear evidence that global rating scales or combination of global rating scale and checklist scales may be a reliable and valid method of rating.[8] In part, these evidences have directed to thinking on utilization of global approach to scoring such as global rating scales, would be better in generating some benefits in OSCE administration.[9]

## METHODS

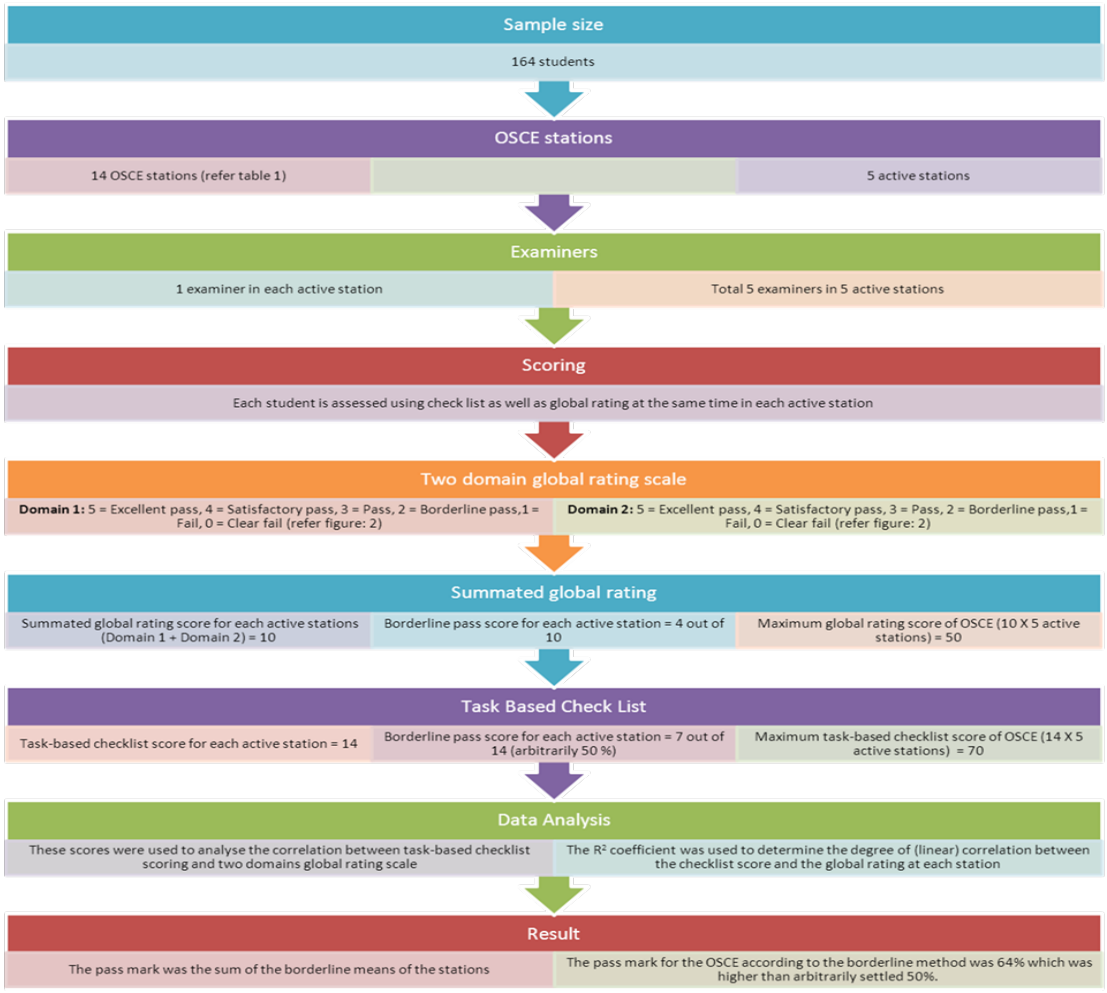
This was a retrospective study aimedto examine the correlation between task-based scoring rating and two domain global rating in a Malaysian private university. The hypothesis of this study was: There will be significant positive relationship between checklist scoring and two domains global

Submission Date : 11-05-2015
Accepted Date : 19-07-2015

DOI: 10.5530/ijper.50.1.3

**Correspondence Address**
***Mr. Sajesh Kalkandi Veettil***
Lecturer,
Department of Pharmacy Practice, International Medical University,
No. 126, Jalan Jalil Perkasa 19, Bukit Jalil, 57000 Kuala Lumpur, Malaysia.
Email: sajesh_kalkandi@imu.edu.my

www.ijper.org




**Graphical Abstract**

rating scale. This allows us to determine the degree of linearity between the checklist score and the global rating score for each station, with the expectation that higher global ratings should generally correspond with higher checklist scores.

**Participants:** 164 pharmacy students' results were recruited from of the OSCE examination conducted in 2013 during their second year end of semester examination.

**Procedure of OSCE:** 14 stations (table 1) total, End of semester OSCE contains a track of 14 stations total, which are linked in sequence. There were 5 active stations, devoted to evaluate the skills in various scenarios. The time allotted for each station was 5 minutes. All the scenarios generated are new and hence would not be formerly faced by the students.

**Development of Check list and global rating scale:** Clinical scenarios and task-based checklists for each station were formulated by faculty of pharmacy practice based on the education outcomes of the module and the students' flat of knowledge following a standardized protocol. Checklists were created and combined into OSCE to upsurge the impartiality and consistency of valuation by dissimilar assessors. The test content were planned against the learning objectives through basic "blueprinting". In the second year, students will find the higher order process in the cognitive and affective domains include valuing and organizing process. The students will begin to incorporate application and analysis in their learning activities. Based on this, as well as the module outcomes and the task-based checklists, key competencies were identified and developed into a two domains global rating scale. This generally represents the overall criterions in the checklists for all the scenarios. In two domains global rating scale, for each domain, a set of six pointscales (0 to 5) was used to reflect high and low divisions within the pass, borderline and fail categories (refer Figure 1). Scores on the two separate global scales were added to generate a 'summated global rating'. Similarly task-based checklists for individual stations were summed for a total score (refer Figure 1). Standardized simulated patients were elected from the established clinical skill center ofthis University. Both checklists as well as global rating scale were validated and

**Global rating scale**

Domain 1: Communication skills — Domain 2: Integrative problem solving skills

5 = Excellent pass
4 = Satisfactory pass
3 = Pass
2 = Borderline pass
1 = Fail
0 = Clear fail

Summated global rating score for each active stations (Domain 1 + Domain 2) = 10
Borderline pass score for each active station = 4 out of 10
Maximum global rating score of OSCE (10 X 5 active stations) = 50
Criteria for marking global rating scale was validated and standardized among markers before examination.

**Task-based checklist score**

Items in the checklist overall represent the criteria present in domain 1 and 2.
Task-based checklist score for each active station = 14
Borderline pass score for each active station = 7 out of 14
Maximum task-based checklist score of OSCE (14 X 5 active stations) = 70
Task-based checklists were standardized among markers prior to examination

**Figure: 1 Task-based checklist scoring and two domain global rating scale[14]**

standardized among 5 markers in each active stationbefore using in the exam. Standardized clinical faculty from a variety of disciplines served as markers.Inter-rater reliability was overcome by absolute agreement method. In this study, an absolute agreement level was 73% which would be acceptable, but exact and adjacent agreement was 80%. However, this is a fairly substantial proportion as per U.S. Department of Education, the Center for Educator Compensation Reform.[10]

**Data Analysis:** We performed the analysis using SPSS version 18. Person's correlation was used to determine the correlations between task-based checklist scoring and the two domains global rating scale. Level of significance was fixed at $p<0.05$.

**Table 1: General information on stations**

| Station | Scenario | Station | Scenario |
|---|---|---|---|
| 1 | Rest | 8 | Preparation |
| 2 | Preparation | 9 | Lifestyle modification |
| 3 | Responding to Symptoms | 10 | Preparation |
| 4 | Rest | 11 | Device counselling |
| 5 | Preparation (Prescription Screening 1) | 12 | Rest |
| 6 | Preparation (Prescription Screening 2) | 13 | Preparation |
| 7 | Prescription Screening | 14 | Medication counselling |

## Ethical approval

As the study is an evaluation, it did not undergo ethical approval. There is, however, an assessment committee for the evaluation methods at our university that was aware of our evaluation. The committee was also asked for an opinion about potential implications. The head of the department and programme coordinator for undergraduate pharmacy was also aware of the study and saw no issue in publishing it. The standards of the Declaration of Helsinki were maintained.

## RESULTS

There were 164 participating students and among them

**Table 2: Correlation between task-based checklist scoring and two domains global rating scale**

| Station | r | $R^2$ | P value |
|---|---|---|---|
| Station 3 | 0.613 | 0.376 | <0.05 |
| Station 7 | 0.692 | 0.479 | <0.05 |
| Station 9 | 0.569 | 0.323 | <0.05 |
| Station 11 | 0.554 | 0.307 | <0.05 |
| Station 14 | 0.491 | 0.241 | <0.05 |

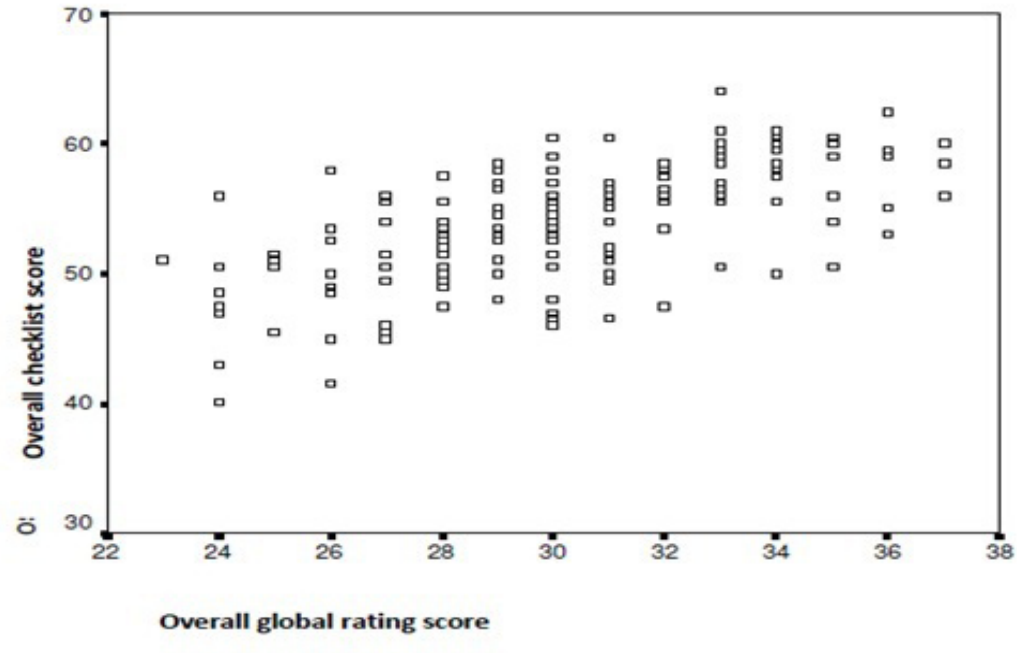



**Figure 2: Scatter diagram showing relationship between overall global rating score and overall checklist score amongall active stations of OSCE**

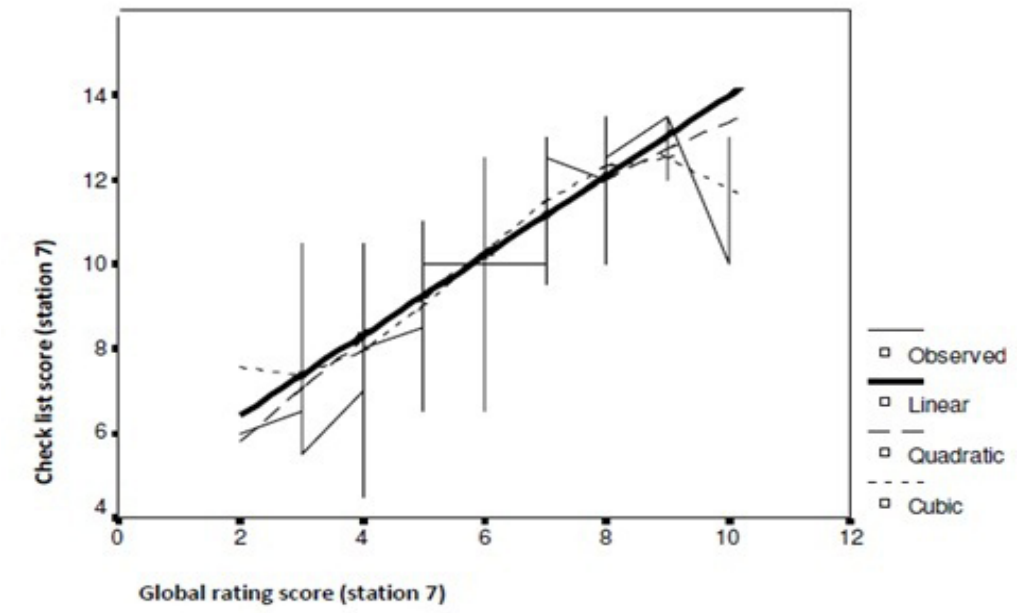



**Figure 3: Curve estimation for station 7**

126 were female and 38 were male. The reliability coefficient (Cronbach's alpha) for overall global rating score showed a value of 0.72-0.78 across all active stations, which was higher than task-basedchecklist scoring 0.602-0.686 across items for all active stations. The Pearson's correlation between task-based checklist scoring and two domains global rating scale were moderate and significant (refer Table 2). Variation between overall global rating score and overall checklist score among all active stations of OSCE are showed in Figure 2.

Station 7 showed a significant correlation (r=0.692) whereas, station 14 (r=0.491) showed the least correlationamong all active stations. Station 7 which was a prescription screening station has a comparatively good $R^2$ value of 0.479 (refer Table 2), implying that 47.9% of variation in the students' global ratings are accounted for the variation in their check list scores. In contrast, station

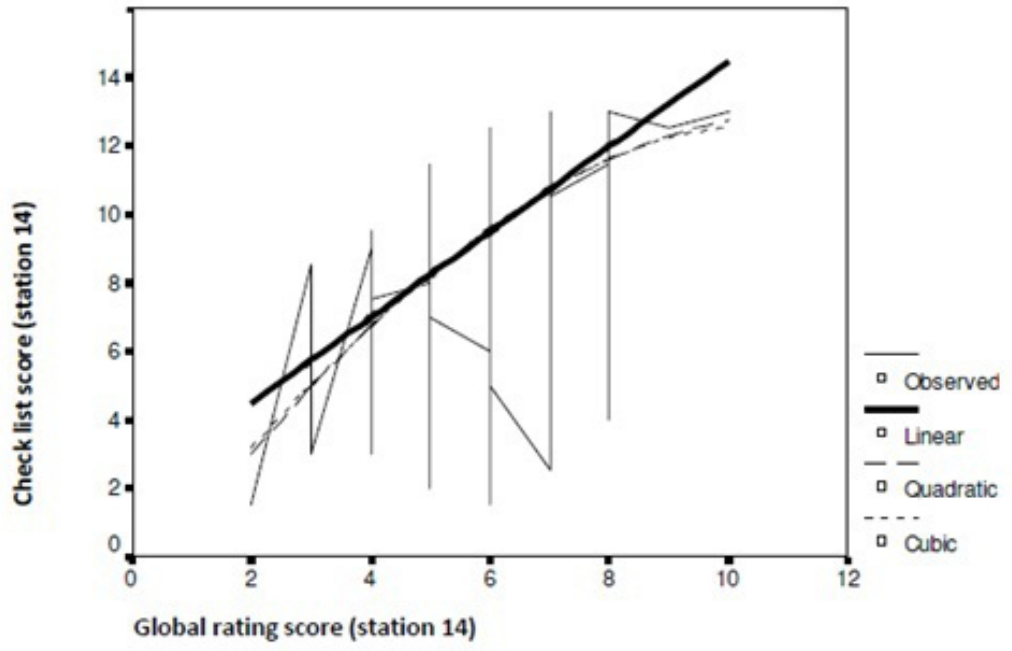



**Figure 4: Curve estimation for station 14**

| Table 3: Curve estimation table for station 7 | | | | | |
|---|---|---|---|---|---|
| **Polynomial fitted** | **$R^2$** | **F** | **Df1** | **Df2** | **P value** |
| **Linear** | 0.479 | 149.09 | 1 | 162 | 0.000 |
| **Quadratic** | 0.485 | 75.71 | 2 | 161 | 0.000 |
| **Cubic** | 0.512 | 55.93 | 3 | 160 | 0.000 |

| Table 4: Curve estimation table for station 14 | | | | | |
|---|---|---|---|---|---|
| **Polynomial fitted** | **R2** | **F** | **Df1** | **Df2** | **p** |
| Linear | 0.241 | 51.33 | 1 | 162 | 0.000 |
| Quadratic | 0.249 | 26.72 | 2 | 161 | 0.000 |
| Cubic | 0.249 | 17.71 | 3 | 160 | 0.000 |

14 is less satisfactory with an $R^2$ value of 0.241. Graphical representation of curve estimation (Figure 3 and 4) as well as curve estimation values (Table 3 and 4) abetted to investigate the exact nature of the association between checklist score and global score for stations 7 and 14.

Even though there was a significant positive correlation between global rating scale and the task based check list ($p<0.05$), $R^2$ value was not satisfactory. This allowed us to determine the degree of linearity between the checklist score and the global rating score for each station, with the expectation that higher global ratings should generally correspond with higher checklist scores.[9] It was helpful to observe the association graphically to examine the precise nature of the association between checklist and global rating that have conquered a low $R^2$ value for station 14 (refer Figure 4). In station 14; there were two main issues–widespread of marks for global rating, and a low spread of marks for 'borderline pass' grade has been awarded (refer Figure 4). The'overall 'borderline pass score' had been set as 4 in two domains global rating scale for all active stations (refer Figure 1). However in station 7, a low spread of marks for global rating can be observed compare to the station 14.

The unsatisfactory association between checklist marks and global ratings can be seen in all stations that cause some degree of non-linearity, as demonstrated in the Table 3, where it is graphically clear that the best fit is clearly cubic. In station 7, the fit of the cubic polynomial is significantly better than that of the linear one.

## DISCUSSION

The reason for $R^2$ value not satisfactory was because of the proportional change in $R^2$ coefficient of the dependent variable (checklist score) and the independent variable (global rating score).[11] $R^2$ coefficient was however, quite small with a lowest value across the 5 active stations being 0.241, implying that only 24.1% of the variation in global rating were clarified by the variation in checklist. This indicates that some students have acquired more marks from the two domains of the global rating, but their checklist marks had not stretched to an expected level.

The reason for unsatisfactory association between checklist marks and global ratings is,mathematically a cubic will always produce a better fit, but stingily dictates that the dissimilarity among the two fits has to be statistically substantial for a higher order model to be preferred.[11] The reason for the fit of the cubic polynomial is significantly better in station 7 is better explained in a previous study[12] which states that, the key point to note is whether the cubic expression is the result of an underlying relationship or as a result of outliers, resulting from inappropriate checklist design or unacceptable assessor behavior in marking.

### Limitations

Though the global rating scale exactly represents the overall criterions in the checklists, the reasons for the unsatisfactory correlation may be due to improper standardization of global rating scale and checklist among markers or poor understanding of criteria to use in the global rating system. Some factors in this study make us to be cautious about generalizing the results. One factor is final decision-making was grounded on the overall score (across all stations) which is a fully compensatory model, however it is always debatable.[13] Another factor is, the correlation ofcheck list scoring system compared to the global rating scale can be improved through statistically driven weightage of checklist items which is not done in this study.

## CONCLUSION

It has been noted that studentswere not able to acquire good marks in the check list scoring system compared to the global rating scale in some OSCE stations though the overall comparison was good. Further study should be done with statistically driven weightage checklist items may be with defined proportion of scores for each station.

## ACKNOWLEDGMENT

The authors wish to thank Dr. Oommen P. Mathew for the interpretation of the results and statistics for this study. The authors also wish to thank Mr. Razman Shah Mohd Razali, Reference Librarian, International Medical University for providing the full text articles whenever needed.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## ABBREVIATION

OSCE : Objective structured clinical examination

SPSS : Statistical Product and Service Solutions

US : United states

## SUMMARY

- The reliability coefficient (Cronbach's alpha) for overall global rating score showed a value of 0.72–0.78 across all active stations.
- The Pearson's correlation between task-based checklist scoring and two domains global rating scale were moderate and significant.
- The unsatisfactory association between checklist marks and global ratings can be seen in all stations that cause some degree of non-linearity.
- The reason for R2 value not satisfactory was because of the proportional change in R2 coefficient of the dependent variable (checklist score) and the independent variable (global rating score).

### About Authors

**Sajesh K Veettil** of International Medical University (IMU), Kuala Lumpur with expertise in Renalpharmacy, educational assessments and pharmacoeconomics.

## REFERENCES


1. Bunmi S, Malau-Aduli *et al*. Inter-rater reliability: comparison of checklist and global scoring for OSCEs. Creative Education 2012; 3(06): 937-42.
2. Wilkinson T, Newble D, Frampton C. Standard setting in an objective tructured clinical examination: use of global ratings of borderline performance to determine the passing score. Med Educ. 2001; 35(11): 1043-9.
3. Boursicot K, Roberts T. How to set up an OSCE. Clin Teach. 2005; 2(1): 16-20.
4. Jodi HeroldMcIlroy, Brian Hodges, Nancy McNaughton, Glenn Regehr. The effect of candidates' perceptions of the evaluation method on reliability of check list and global rating scores in an objective structured clinical examination. Acad Med. 2002; 77(7): 725-28.
5. Regehr G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med. 1998; 73(9): 993-7.
6. Kristalyn SP. Psychometric Properties definition September 03, 2013 http://bpd.about.com/od/glossary/g/Psychometric-Properties.htm
7. Norman GR, Tugwell P, Feightner W, *et al*. Knowledge and clinical problem solving. Med Educ. 1985; 19(5): 344-56.
8. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. Med Educ. 2003 Nov; 37(11): 1012-6.
9. Regehr G, Freeman R, Robb A, Missiha N, Heisey R. OSCE performance evaluations made by standardized patients: comparing checklist and global rating scores. Acad Med. 1999; 74(10): S135-7.
10. Center for Educator Compensation Reform, Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings February 2012. http://cecr.ed.gov/pdfs/Inter_Rater.pdf
11. Rothman AI, Blackmore D, Dauphinee WD, Reznick R. The use of global ratings in OSCE station scores. Advances in Health Sciences Education 1997; 1(1): 215-9.
12. Pell G, *et al*. How to measure the quality of the OSCE: A review of metrics–AMEE guides. Medical teacher 2010; 32(49): 802-11.
13. David N. Techniques for measuring clinical competence: objective structured clinical examinations. Med Edu. 2004; 38(2): 199-203.
14. Kingston R, Sajesh KV, Suresh K. Standard setting in OSCEs: a borderline approach. The clinical Teacher 2014; 11(7): 551-6.