

# An Analysis of the Extent of Intra and Inter-rater Variability in OSCE

Swanand Shrikant Pathak<sup>1</sup>, Gaurav Vedprakash Mishra<sup>2</sup>, Rupesh Anna Warbhe<sup>1,\*</sup>

<sup>1</sup>Department of Pharmacology, JNMC, DMIHER, Sawangi, Wardha, Maharashtra, INDIA.

<sup>2</sup>Department of Radiology, AVBRH, Sawangi, Wardha, Maharashtra, INDIA.

## ABSTRACT

**Background:** OSCE is an objective structured clinical examination, though the tool is designed to increase the objectivity in the examination of the clinical cases, the inter and intra rater variability of the tool needs to be analysed. **Aim:** To determine the extent of intra and inter-rater variability in OSCE. **Settings and Design:** The study was conducted in a medical college and is a cross sectional study. **Materials and Methods:** 5 OSCE stations were designed and videos recorded of the candidates performing the clinical exercise, 9 teachers were selected, three of assistant professor grade (junior level), three of associate grade (middle level) and three of professor grade (senior level), all the teachers were shown the videos on day 1 and the scores recorded, the same videos were shown to the teachers on the second day and the scores recorded, each teacher graded all the 5 OSCE recordings. **Statistical Analysis Used:** Student paired *t* test, One-Way ANOVA and multiple comparison to Tukey test and Cronbach alpha. Software used for analyses was SPSS 27.0 version. **Results:** when the scores were segregated according to level of Assessors it was observed that there was significant intra-rater variability at senior level, while there is significant inter-rater variability between junior and mid-level assessors. **Conclusion:** There is intra and inter-rater variability observed in OSCE assessment when assessors were segregated in groups as per their seniority levels.

**Keywords:** OSCE, Variability, Inter rater, Intra rater.

## Correspondence:

**Dr. Rupesh Anna Warbhe**

Department of Pharmacology, JNMC,  
DMIHER, Sawangi, Wardha, Maharashtra,  
INDIA.

Email: warbherupesh5@gmail.com

**Received:** 27-06-2022;

**Revised:** 08-03-2023;

**Accepted:** 08-06-2023.

## INTRODUCTION

In medical education, Performance Assessment (PA) has been widely employed Objective Structured Clinical Examination (OSCE) being an excellent example. An OSCE usually comprises a circuit of short stations, where candidates are examined on a one-to-one basis with one or two impartial examiner(s) and either real or simulated patients.<sup>1</sup> An OSCE is designed to be *objective* – all candidates are assessed using the same stations and marking scheme. In an OSCE, candidates get mark for each step on the mark scheme which they perform correctly, which therefore makes the assessment of clinical skills more objective, rather than subjective.<sup>2</sup>

## MATERIALS AND METHODS

5 OSCE stations were designed and videos recorded of the candidates performing the clinical exercise, 9 teachers were selected, three of assistant professor grade (junior level), three

of associate grade (middle level) and three of professor grade (senior level), all the teachers were shown the videos on day 1 and the scores recorded, the same videos were shown to the teachers on the second day and the scores recorded, each teacher graded all the 5 OSCE recordings.

## Statistical analysis used

Student paired *t* test, One-Way ANOVA and multiple comparison to Tukey test and Cronbach alpha. Software used for analyses was SPSS 27.0 version.

## RESULTS

The intra rater and inter-rater variability at day 1 and day 2 is not statistically significant, but when we compare intra rater variability between day 1 and 2 we observed statistical significance difference between the scorings of day 1 and day 2 in the senior level. When we compare for inter rater variability of day 1 and day 2 between junior and mid-level it is significant, mid-level and senior level it is not significant, junior level and senior level it is not significant. When we compare all the three it is observed to be significant.

when the scores were segregated according to level of Assessors it was observed that there was significant intra-rater variability at senior level, while there is significant inter-rater variability



DOI: 10.5530/ijper.57.3s.58

### Copyright Information :

Copyright Author (s) 2023 Distributed under  
Creative Commons CC-BY 4.0

**Publishing Partner :** EManuscript Tech. [www.emanuscript.in]

between junior and mid-level assessors. Inter-rater reliability is the degree of agreement among raters. It is a score of how much homogeneity or consensus exists in the ratings given by various judges. Whereas, intra-rater reliability is a score of the consistency in ratings given by the same person across multiple instances.

## DISCUSSION

Statistical analysis was done by using descriptive and inferential statistics using student's paired *t* test, one way ANOVA, multiple comparison: Tukey Test and Cronbach Alpha method of reliability and software used in the analysis was SPSS 27.0 version and  $p < 0.05$  is considered as level of significance. The purpose of this study was to evaluate the inter rater and intra rater variability among the various levels of medical teachers. The focus of the

study examined the consistency in grading between same examiners on different day and different level.

The method in our study making use of the interrater and intra-rater variability done on consecutive day and incorporation of different level of examines has been done for the first time (Tables 1-8) (Figures 1-3).

There was no study found doing the intra rater variability study involving the same examiner on different day which is unique in this study<sup>3-11</sup> (Tables 1-8). Although there was no such study which has made use of the above criteria, we found few studies related to interrater variability which is compared with. Weaver *et al.* showed interrater variability which was found to be similar in our study when it was compared on different days.<sup>12</sup> Mortsiefer *et al.* showed inter significant interrater variability among the clinical vs non clinical which was found to be similar with our

**Table 1: Comparison of score at three levels at Day 1-Descriptive Statistics.**

Level	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Junior Level	5	17.13	1.44	0.64	15.33	18.92	15.33	19.33
Mid-Level	5	17.13	2.03	0.91	14.60	19.66	14.33	19.00
Senior Level	5	17.00	1.33	0.59	15.34	18.65	15.33	18.67
Total	15	17.08	1.51	0.39	16.25	17.92	14.33	19.33

**Table 2: Comparison of score at three levels at Day 1-One way ANOVA.**

Source of variation	Sum of Squares	$d_f$	Mean Square	F	<i>p</i> -value
Between Groups	0.059	2	0.030	0.011	0.989
Within Groups	32.044	12	2.670		NS, $p > 0.05$
Total	32.104	14			

**Table 3: Comparison of score at three levels at Day 1-Multiple Comparison: Tukey Test.**

Level		Mean Difference (I-J)	Std. Error	<i>p</i> -value	95% Confidence Interval	
					Lower Bound	Upper Bound
Junior Level	Mid-Level	0.00	1.03	1.000, NS	-2.75	2.75
	Senior Level	0.13	1.03	0.991, NS	-2.62	2.89
Mid-Level	Senior Level	0.13	1.03	0.991, NS	-2.62	2.89

**Table 4: Comparison of score at three levels at Day 2- Descriptive Statistics.**

Level	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Junior Level	5	17.20	1.65	0.42	16.28	18.11	15.00	20.00
Mid-Level	5	17.00	1.55	0.40	16.13	17.86	14.00	19.00
Senior Level	5	18.20	1.26	0.32	17.49	18.90	15.00	20.00
Total	15	17.46	1.56	0.23	16.99	17.93	14.00	20.00

**Table 5: Comparison of score at three levels at Day 2-One way ANOVA.**

Source of variation	Sum of Squares	$d_f$	Mean Square	F	p-value
Between Groups	12.40	2	6.20	2.74	0.076 NS, $p>0.05$
Within Groups	94.80	42	2.25		
Total	107.20	44			

**Table 6: Comparison of score at three levels at Day 2-Multiple Comparison: Tukey Test.**

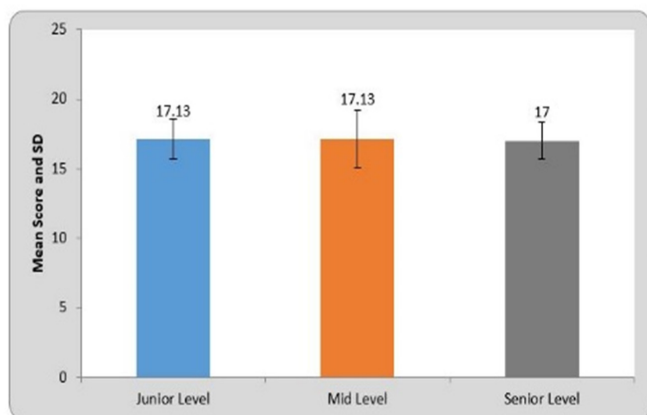
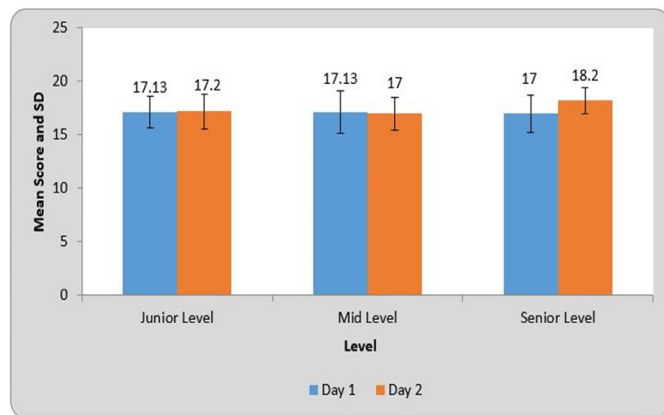
Level		Mean Difference (I-J)	Std. Error	p-value	95% Confidence Interval	
					Lower Bound	Upper Bound
Junior Level	Mid-Level	0.20	0.54	0.929, NS	-1.13	1.53
	Senior Level	-1.00	0.54	0.175, NS	-2.33	0.33
Mid-Level	Senior Level	-1.20	0.54	0.085, NS	-2.53	0.13

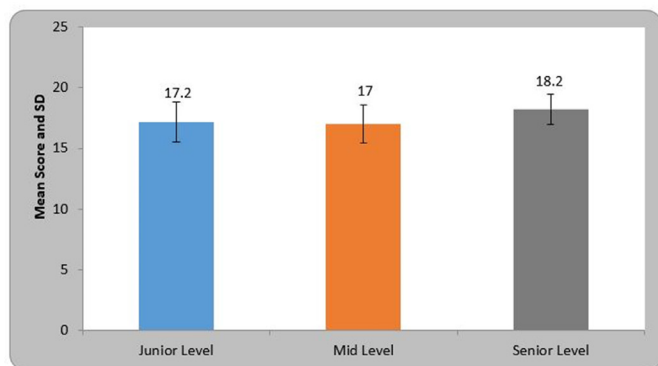
**Table 7: Comparison of score at Day 1 and at Day 2 -Student's paired t test.**

	Day 1	Day 2	Mean Difference	t-value
Junior Level	17.13±1.45	17.20±1.65	0.06±0.79	0.32 $p=0.75$ , NS
Mid-Level	17.13±1.95	17±1.55	0.13±1.68	0.30 $p=0.76$ , NS
Senior Level	17±1.73	18.20±1.26	1.20±1.42	3.26 $p=0.006$ , S

**Table 8: Inter rater Variability.**

	Junior vs Mid	Mid vs Senior	Junior vs Senior	Junior vs Mid vs Senior
Cronbach Alpha	0.891	0.202	0.134	0.640
p-value	0.0001, S	0.56, NS	0.63, NS	0.0001, S

**Figure 1:** Comparison of score at three levels.**Figure 3:** Comparison of score at day 1 and at day 2.



**Figure 2:** Comparison of score at three levels at day 2.

study.<sup>13</sup> Our study has some similarity with the Faherty *et al.* where they have obtained less variability when the examiners are paired as compared to the individual examiner.<sup>14</sup>

## CONCLUSION

There is intra and inter-rater variability observed in OSCE assessment when assessors were segregated in groups as per their seniority levels. This knowledge can be used for further application in the medical curriculum for the non-biased assessment of the students.

## ACKNOWLEDGEMENT

The authors acknowledge the contribution of the statistician of the institute for contribution in analysis of the data in this study.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## ABBREVIATIONS

**ANOVA:** Analysis of Variance; **OSCE:** Objective Structured Clinical Examination. **PA:** Performance Assessment; **SPSS:** Statistical Package for Social Sciences; **STD:** Standard.

## SUMMARY

The purpose of conducting and undertaking this study was to truly assess and analyse the implication of the assessment of the students through the different teachers and how they interpret the performance of the students under different time.

## REFERENCES

- Brenner E. Human body preservation – old and new techniques. *J Anat.* 2014;224(3):221-76. Harden:316-44. doi: 10.1111/joa.12160, PMID 24438435. Downie M, WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ.* 1975;1(5955):447-51. doi: 10.1136/bmj.1.5955.447, PMID 1672423, PMID 1115966.
- Ross M, Carroll G, Knight J, Chamberlain M, Fothergill-Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. *J Adv Nurs.* 1988;13(1):45-56. doi: 10.1111/j.1365-2648.1988.tb01390.x, PMID 3372885.
- Calls to ensure end-of-year exam fairness. [retrieved 2017-4-13]. British Medical Association - Connecting doctors.
- Yeung E. Sharing of information in objective structured clinical examinations - does no evidence of difference mean evidence of no difference? *Open: BMJ Publishing Group;* 2017.
- Parks R, Warren PM, Boyd KM, Cameron H, Cumming A, Lloyd-Jones G. The Objective Structured Clinical Examination and student collusion: marks do not tell the whole truth. *J Med Ethics.* 2006;32(12):734-8. doi: 10.1136/jme.2005.015446, PMID 2563344, PMID 17145916.
- Fleming PR, Manderson WG, Matthews MB, Sanderson PH, Stokes JF. Evolution of an examination: M.R.C.P. (U.K.). *Br Med J.* 1974;2(5910):99-107. doi: 10.1136/bmj.2.5910.99, PMID 4596404.
- Wilson GM, Lever R, Harden RM, Robertson JI. Examination of clinical examiners. *Lancet.* 1969;1(7584):37-40. doi: 10.1016/s0140-6736(69)90998-2, PMID 4178777.
- Hubbard JP, Levit EJ, Schumacher CF, Schnabel TG. An objective evaluation of clinical competence. New technics used by The National Board of Medical Examiners. *N Engl J Med.* 1965;272:1321-8. doi: 10.1056/NEJM196506242722505, PMID 14299142.
- Marshall VR, Ludbrook J. The relative importance of patient and examiner variability in a test of clinical skills. *Br J Med Educ.* 1972;6(3):212-7. doi: 10.1111/j.1365-2923.1972.tb01660.x, PMID 5077435.
- Lever RS, Harden RM, Wilson GM, Jolley JL. A simple answer sheet designed for use with objective examinations. *Br J Med Educ.* 1970;4(1):37-41. doi: 10.1111/j.1365-2923.1970.tb01806.x, PMID 5432638.
- Penta FB, Kofman S. The effectiveness of simulation devices in teaching selected skills of physical diagnosis. *J Med Educ.* 1973;48(5):442-5. doi: 10.1097/00001888-197305000-00005, PMID 4699421.
- Weaver SB, Daftary M, Wingate L, Turner M. Evaluation of faculty inter-variability OSCE grade scoring on overall student performance in a laboratory course. *Pharm Educ.* 2022;22(1):48-53. doi: 10.46542/pe.2022.221.4853.
- Mortsiefer A, Karger A, Rotthoff T, Raski B, Pentzek M. Examiner characteristics and interrater reliability in a communication OSCE. *Patient Educ Couns. patient ed. patient education and counselling.* 2017;100(6):1230-4. doi: 10.1016/j.pec.2017.01.013, PMID 28139274.
- Faherty A, Counihan T, Kropmans T, Finn Y. Inter-rater reliability in clinical assessment: do examiner pairings influence candidate ratings? *BMC Med Educ.* 2020;20(1):147. doi: 10.1186/s12909-020-02009-4, PMID 32393228.

**Cite this article:** Swanand PS, Gaurav M, Rupesh WA. An Analysis of the Extent of Intra and Inter-rater Variability in OSCE. *Indian J of Pharmaceutical Education and Research.* 2023;57(3s):s507-s510.