

Reliability of Rubrics in Mini-CEX

Anam Arshad, Muhammad Moin, Lubna Siddiq

Pak J Ophthalmol 2017, Vol. 33, No. 1

See end of article for
authors affiliations

Purpose: To study the reliability of rubrics in mini clinical exercise (CEX) in Ophthalmic examination.

Study Design: Observational cross sectional study.

Place and Duration of Study: Our study was conducted at the ophthalmological society of Pakistan, Lahore branch on Sep 17, 2015.

Material and Methods: 16 raters were recruited from the candidates eligible for fellowship exit exam. All these raters were provided with a rubric to evaluate the clinical performance of cover/uncover (squint assessment) test. Every rater gave scores (2-5) for 12 steps of the clinical examination. All scores were entered into SPSS version 20 and Cronbachs' alpha coefficient of inter rater reliability and internal consistency of scores was determined.

Results: 16 raters having age range from 26-35 years with mean age of 29.4 SD \pm 1.99 took part in this study. Out of them 7 were male and 9 were female. The Cronbach Alpha (0.972) was found to be very significant after analyzing the scores of the sixteen raters in SPSS. The intra class correlation co-efficient was found to be .967. Descriptive statistics showed that sixteen raters gave a rating between 3.3 to 4.0 for each step of the rubric.

Conclusion: Rubrics are effective in achieving a high inter rater reliability in mini-CEX and make it a very useful tool in assessment of clinical skills.

Keywords: Rubrics, mini-CEX, inter rater reliability, variability.

Correspondence to:

Anam Arshad
Postgraduate Trainee,
Postgraduate Medical Institute
Lahore.
Email: anam_1038@hotmail.com

Clinical Skills of residents in many specialty training programs have been assessed by using mini-clinical evaluation exercise (mini-CEX). This tool provides both assessment and

education for residents in training¹ and its validity has been established². The mini-CEX is also a feasible and reliable evaluation tool for post graduate residency training³. The number of feedback comments make the

mini-CEX a useful assessment tool⁴. To some extent, such a tool may predict the future performance of medical students⁵. The mini-CEX has been well received by both learners and supervisors⁶.

Resident performance which is valid is required by all program directors for certification of competence of all trainees completing their residency^{7,8}. However, assessments which are valid in assessing clinical skills can be challenging⁹. Long case clinical evaluation exercise (CEX) has been proven to be unreliable in a research conducted by the American Board of Internal Medicine (ABIM) because the inter-rater and inter-case reliability is quite high^{10,11,12}. Validity of mini-CEX scores could be better if the inter rater reliability was improved which would also lead to reduction in resident-patient encounters¹³. Consistency of examiner ratings is necessary to improve reliability of assessment¹⁴.

Use of topic-specific analytical rubrics can improve the reliability of performance scoring of assessments especially with examples and/or training of raters¹⁵. Introduction of Rubrics in assessment make the criteria and expectations very clear and also facilitate self-assessment and feedback. This is the reason why learning is promoted and instruction is enhanced by the use of rubrics¹⁵. We undertook this

study to find out the reliability of rubric in mini-CEX as a reliable tool of assessment.

MATERIALS AND METHODS

Our study was conducted at the ophthalmological society of Pakistan, Lahore branch on Sep 17, 2015. It was observational cross sectional study by randomized non-probability consecutive convenient sampling technique. Sixteen raters were recruited from the candidates eligible for fellowship exit exam, who were attending a pre examination preparatory course on clinical ophthalmology. A consent was signed by the raters and their names and all other details were kept confidential. All these raters were provided with a rubric set to evaluate the clinical performance of cover/uncover (squint assessment) test, figure 1. All the raters gave scores to the steps of single clinical performance by junior resident. Every rater gave scores (2-5) for 12 steps of the clinical examination method. All scores were entered into SPSS version 20 and Cronbachs' alpha coefficient of inter rater reliability and internal consistency of scores was determined. Raters with incorrectly filled forms were excluded from the study. A demonstration about how to fill the rubric was given to all the participants before the actual test.

Figure 1: Resident Assessment Form (cover/uncover test).

Skill	Novice (Score 2)	Beginner (Score 3)	Advanced Beginner (Score 4)	Competent (Score 5)	Total Score
Introduction	Not introduced	Introduced as doctor Didn't ask patient name	Introduced as doctor Ask patient name	Inquired patients name and well being	
Informed Consent	No consent	Didn't explain procedure	Didn't insist on fixation Didn't ask about refractive error	Fully explained the procedure	
Examination level	Didn't adjust	Inaccurate adjustment	Awkward adjustment	Accurate proper adjustment	
Visual acuity	Not assessed	Assessed for near only	Assessed for far and near	Asked for snellens. Assessed unaided and aided VA Recorded VA	
Hirschberg	Didn't perform	Didn't ask patient to look at spot light	Asked to fixate at light but light not held properly and	Asked to fixate light held centrally and stable	

			centrally		
Near Target	Didn't given	Target not held at working distance	Target held at working distance	Target held at working distance with stability	
Cover test	Didn't cover	Covered deviating eye	Covered fixating eye	Completely covered fixating eye with occluding	
Uncover test	Didn't perform	Observed uncovered eye	Observed covered eye	Observed covered eye and measured secondary deviation	
Alternate cover test	Didn't perform	Performed but too rapidly or slowly	Performed with proper time for cover and uncover	Performed with proper time	
Repetition of steps for Far targets	Didn't perform	Didn't gave specific target	Gave specific target Steps incomplete	Gave specific target and Completed examination steps	
Repetition of steps with glasses	Didn't inquire about glasses	Repeated with glasses for far only or near only	Repeated with glasses for far and near	Repeated with glasses and explain completely	
Thank the patient	Didn't thank the patient	Thanked the patient	Thanked the patient with smile	Thanked the patient and shook hand	

RESULTS

The study included 16 raters having age range from 26 – 35 years with mean age of $29.4 \text{ SD} \pm 1.99$. Out of them 7 were male and 9 were female. There are 12 steps to be scored by the raters, every step carried 5 marks, missing a particular step by the candidate was recommended by the rubric to be scored as zero. If the step was performed by the candidate its proficiency was scored guided by the rubric from one to five score. The Cronbach Alpha (0.972) was found to be significant after analyzing the scores of the sixteen raters in SPSS, table 2. The intra class correlation coefficient was found to be .967, table 3. Descriptive statistics showed that sixteen raters gave a rating between 3.3 to 4.0 for each step of the rubric, table 4.

Table 1: Demographic Data.

Characteristics	Groups	Number
AGE	< 28	4
	28 – 32	9

	> 32	3
GENDER	Male	7
	Female	9
Experience in ophthalmology	< 4 years	2
	4 – 6 years	10
	> 6 years	4
NUMBER		16

Table 2: Reliability statistics.

Cronbachs' Alpha	Number of Raters
0.972	16

Table 3: Intra class correlation coefficient.

		95% Confidence Interval	
Intra Class Correlation (ICC)	Lower Bound	Upper Bound	

Average measures	.967	.932	.989
------------------	------	------	------

One-way random effect

Table 4: Inter rater reliability: Mean and Standard deviation.

Rater	Mean	Standard Deviation	Number
1	3.3	± 0.77	12
2	4.0	± 1.1	12
3	4.2	± 1.1	12
4	3.4	± .90	12
5	3.7	± 1.1	12
6	3.5	± 1.0	12
7	3.5	± 1.0	12
8	3.2	± .75	12
9	3.8	± .93	12
10	3.3	± .88	12
11	3.4	± .79	12
12	3.4	± 90	12
13	4.0	± 1.2	12
14	3.5	± 1.0	12
15	3.6	± 1.1	12
16	3.7	± 1.2	12

DISCUSSION

High reliability of assessment of medical examiners has been shown by several researchers when rubric is introduced^{15,16}. On the other hand the reliability has never been found to decrease when rubrics are used. Therefore, rubrics are being used by a lot of teachers on the assumption that grading objectivity is enhanced, especially regarding the performance of the students. This leads to the postulation that when rubrics are not used in assessment, there is more subjectivity because of the examiner's only subjective judgment of the performance of the students. Consequently teachers usually prefer to incorporate a rubric in all their assessments¹⁷. But there are cases where inconsistent scores are produced even when rubrics are used due to many problems. Inter-rater reliability scores can be affected by many factors, including "the objectivity of the task/item/scoring,

the difficulty of the task/item, the group homogeneity of the examinees/raters, speediness, number of tasks/items/raters, and the domain coverage". Poor reliability of the raters has been seen when there is poor training of raters, insufficient detail in the rubric, or "failure of the examiners to internalize the rubrics"¹⁸. Raters with diverse levels of scoring capacity do not look at different results or performance features, but their understanding about the criteria of scoring has many levels¹⁹. Injustice and bias is removed in assessments by using rubrics because criteria for scoring a student performance are clearly defined. The details given in the various score levels of the rubrics act as a guide in the process of evaluation. Designing a good rubric scoring can eliminate the occurrence of discrepancies between different raters²⁰. The reliability of scoring across students is enhanced by rubrics, along with the consistency between different raters. Another advantage of using a rubric is that a valid decision of performance assessment is achieved which is not possible with rating done conventionally. Complex competencies can be assessed according to the desired validity by using rubrics²¹.

In our study, the Cronbach's alpha coefficient for 16 raters was found to be 0.972, showing that there is a relatively high internal consistency of the raters. Reliability coefficient of 0.70 or higher is considered "acceptable" in most research situations according to the institute for digital research and education UCLA-Los Angeles.

D'Antoni et al; calculated inter rater reliability of 3 examiners that judged 66 first year medical students using MMAR(mind mapping assessment rubric) and calculated cronbachs' alpha coefficient of 0.38²².

Fallatah et al assessed the reliability and validity of sixth year medical students at king Abdulaziz University by four examiners (2 seniors and 2 juniors) and Internal-consistency reliabilities for the total assessment scores were calculated. Cronbachs' alpha for the four parts of the total assessment score on both long and short cases (2012) or OSCE (2013) was 0.63 and 0. 83 for 2012 and 2013²³.

Daniel et al studied inter-rater reliability in evaluating the micro surgical skills of ophthalmology residents and alpha Cronbachs' found to be 0.72²⁴.

Golnik et al observed that Ophthalmic Clinical Evaluation Exercise (OCEx) is a reliable tool for the faculty to assess clinical competency of residents, alpha Cronbachs' reliability coefficient was 0.81²⁵.

CONCLUSION

Rubrics are effective in achieving a high inter rater reliability in mini-CEX and make it a very useful tool in assessment of clinical skills.

Author's Affiliation

Dr. Anam Arshad
Postgraduate Trainee,
Postgraduate Medical Institute, Lahore.

Prof. Muhammad Moin
Prof of Ophthalmology,
Postgraduate Medical Institute Lahore.

Dr. Lubna Siddiq
Senior Registrar,
Department of Ophthalmology,
Postgraduate Medical Institute Lahore.

Role of Authors

Dr. Anam Arshad
Collection of Data and manuscript writing.

Prof. Muhammad Moin
Study Design, Manuscript Review.

Dr. Lubna Siddiq
Statistical Analysis.

REFERENCES

1. **Malhotra, S., Hatala, R., and Courneya, C.A.** Internal medicine residents' perceptions of the Mini-Clinical Evaluation Exercise. *Med Teach.* 2008; 30: 414-419.
2. **Kogan, J.R., Holmboe, E.S., and Hauer, K.E.** Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA.* 2009; 23: 1316-1326.
3. **Durning, S.J., Cation, L.J., and Jackson, J.L.** The reliability and validity of the American Board of Internal Medicine Monthly Evaluation Form. *Acad Med.* 2003; 78: 1175-1182.
4. **Pernar, L.I., Peyre, S.E., Warren, L.E. et al.** Mini-clinical evaluation exercise as a student assessment tool in a surgery clerkship: lessons learned from a 5-year experience. *Surgery.* 2011; 150: 272-277.
5. **Ney, E.M., Shea, J.A., and Kogan, J.R.** Predictive validity of the mini-Clinical Evaluation Exercise (mCEX): do medical students' mCEX ratings correlate with future clinical exam performance? *Acad Med.* 2009; 84: S17-S20.
6. **Nair, B.R., Alexander, H.G., McGrath, B.P. et al.** The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Med J Aust.* 2008; 189: 159-161.
7. **Holmboe ES, Hawkins RE, Huot SJ.** Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med.* 2004; 140: 874-81.
8. **Norcini JJ, Blank LL, Duffy FD, Fortna GS.** The mini-CEX: a method for assessing clinical skills. *Ann Intern Med.* 2003; 138: 476-81.
9. **Kogan JR, Bellini LM, Shea JA.** Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med.* 2003; 78 (10 Suppl): S33-5.
10. **Herbers JE Jr., Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ.** How accurate are faculty evaluations of clinical competence. *J Gen Intern Med.* 1989; 4: 202-8.
11. **Kroboth FJ, Hanusa BH, Parker S, et al.** The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med.* 1992; 7: 174-9.
12. **Noel GL, Herbers JE Jr., Caplow MP, Cooper GS, Pangaro LN, Harvey J.** How well do internal medicine faculty members evaluate the clinical skills of residents. *Ann Intern Med.* 1992; 117: 757-65.
13. **Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS.** Effect of Rater Training on Reliability and Accuracy of Mini-CEX Scores: A Randomized, Controlled Trial. *Gen Intern Med.* 2009 Jan; 24 (1): 74-79.
14. **Ogunbanjo GA.** Adapting mini-CEX scoring to improve inter-rater reliability. 2009; 43 (5): 484-485.
15. **Johnsson A, Svingby G.** The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review.* 2007; 2 (2): 130-144.
16. **Silvestri, L., & Oescher, J.** Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education.* 2006; 9: 25-30.
17. **Spandel, V.** In defense of rubrics. *English Journal.* 2006; 96 (1): 19-22.
18. **Colton, D. A., Gao, X., Harris, D. J., Kolen, M. J., Martinovich-Barhite, D., Wang, T., et al.** Reliability Issues with Performance Assessments: A Collection of Papers. *ACT Research Report Series.* 1997; 97-3.
19. **Wolfe, E. W., Kao, C., & Ranney, M.** Cognitive differences in proficient and nonproficient essay scorers. *Written Communication.* 1998; 15 (4).
20. **Moskal, B. M., & Leydens, J. A.** Scoring rubrics development: Validity and reliability. *Practical Assessment, Research, and Evaluation.* 2000; 7 (10).
21. **Morrison, G. R., & Ross, S. M.** Evaluating technology-based processes and products. *New Directions for Teaching and Learning.* 1998; 74.
22. **D'Antoni et al;** *BMC Medical Education* 2009 9: 19: 10.1186/1472-6920-9-19.
23. **Fallatah et al;** *BMC Medical Education* 2015 15:10. 10.1186/s12909-015-0295-4.

24. **Daniel et al;** Skills Acquisition and Assessment after a Microsurgical Skills Course for Ophthalmology Residents. *Ophthalmol.* 2009; 116 (2): 257-262.
25. **Golink KC et al;** The Ophthalmic Clinical Evaluation Exercise: Reliability Determination. 2005; 112 (10): 1649-1654.