

Faculty and patient evaluations of radiology residents' communication and interpersonal skills

Naila Nadeem,¹ Abdul Mueed Zafar,² Rukhsana Wamiq Zuberi,³ Muhammad Nadeem Ahmad⁴

Department of Radiology,^{1,2,4} Department of Educational Development,³ Aga Khan University, Karachi, Pakistan,

Department of Diagnostic Imaging, Alpert Medical School of Brown University, Providence, RI, USA,²

Corresponding Author: Naila Nadeem. Email: naila.nadeem@aku.edu

Abstract

Objective: To assess communication and interpersonal skills (CIS) of radiology residents through faculty and standardised patients (SP).

Methods: In this day-long objective structured clinical examination (OSCE) in January 2009, 42 radiology residents took part at six stations in Karachi, each with a standardised patient and a faculty evaluator. Each encounter lasted 15 minutes followed by independent assessments of the residents by both the evaluators.

Results: Based on rating-scale evaluations, all cases had satisfactory internal consistency (Cronbach's alpha 0.6 to 0.9). The alpha values were comparatively diminutive against the checklist scores. Correlation among faculty was 0.6 ($p < 0.001$) with the use of both the checklist and the rating scale. Among standardised patient, intra-class correlation was 0.6 ($p < 0.001$) for checklists and 0.7 ($p = 0.001$) for rating scales. Moderate to strong correlations ($r = 0.6$ to 0.9) existed between checklist and rating scores by the same type of evaluator. Correlations among the faculty and standardised patient using the same assessment tool were unimpressive.

Conclusion: Both checklists and rating scales can serve as satisfactory assessment tools for communication and interpersonal skills using objective structured and clinical examination with the assistance of faculty and standardised patients.

Keywords: Radiologists, Communication, Interpersonal skills, Standardised patients. (JPMA 62: 915; 2012)

Introduction

Radiologists play a key role in the delivery of healthcare by effective provision of diagnostic information to the patients and the care-providers.¹⁻³ Though such information is mainly transmitted in writing, situations demanding communication and interpersonal skills (CIS) are commonly encountered.^{1,2} These include issues bearing scientific, ethical and legal implications such as differences of opinion among colleagues, ominous findings during an antenatal ultrasound or urgent findings pertinent to patient safety.^{1,4} A deep understanding of CIS is imperative for a radiologist as s/he has the unique responsibility of communicating with the patients, their families and the clinicians.⁵

CIS is among the core competencies mandated by the Accreditation Council for Graduate Medical Education.⁶ The Joint Commission also stipulates proficiency in CIS for the credentialing of physicians.⁷ The American College of Radiology has been providing standards of non-written communication for over a decade.¹ Despite such strong emphasis, CIS receives little attention in radiology residency programmes. Detection of the disease and

performance of the procedure are emphasised much more than effective communication of information thus collected.⁸ In one study, 80% clinical radiologists perceived themselves inadequately trained in communication skills.⁹ Many programmes have introduced workshops or courses in CIS. Nevertheless, mere exposure does not correlate with performance in a CIS exam.¹⁰ At the same time, exit examinations in Radiology are generally deficient in CIS component and, unfortunately, learning tends to be focused on areas of assessment.¹¹

The complexity involved in objective measurement of CIS is a major roadblock to its inclusion in competency examinations.^{12,13} A multitude of methods are available: oral examinations, direct observation by faculty and peers, objective structured clinical examination (OSCE) employing standardised patients (SP), self-assessment scales, patient surveys, computer-assisted simulations, and 360 degree assessments.¹³ Of these, SP-OSCE has been studied extensively and appears to be the most promising.¹⁴

Selection of the appropriate tool is also critical for gauging CIS.¹² Both checklists and rating scales have been utilised for the psychometric assessment of CIS through SP-

OSCE.¹⁴ While proponents of checklists emphasise objective assessment based on performance of specific tasks, others point towards the potential of rating scales for capturing subjectivity inherent to CIS.^{11,14,15} We compared the utility of checklists and rating scales for assessment of Radiology residents' CIS by two types of evaluators: Faculty and SP.

Subjects and Methods

A day-long SP-OSCE was conducted at Aga Khan University Hospital, Karachi, Pakistan in January 2009. The study was approved by the institution's Ethical Research Committee. Written informed consent was a pre-requisite for participation in the study. Participants wore random number identification tags throughout the OSCE. No identifying information of participating individuals or institutions was collected.

All post-graduate trainees enrolled in Radiology departments of Karachi, accredited by the College of Physicians and Surgeons Pakistan (CPSP) at the time of study and had attended the workshop on communication skills at the CPSP were eligible for participation.

A minimum sample of $n=34$ was required for achieving greater than 80% power to detect a difference of 0.5 in Pearson's coefficient (significance level 0.05, two-tailed) for correlation between scores using the two evaluation tools. A two-stage sampling strategy was employed. Of the nine (five public, four private) eligible training programmes, three public and two private institutes were identified by random draw. After permission from each institution, a random sample of residents was selected from these programmes. The sample size was inflated by 25% and $n=42$ residents were invited to participate.

Incident reports and patient complaints filed with the department of Radiology, where the study was conducted, were reviewed to identify six case scenarios. Subsequently, extensive literature review and expert discussions were conducted to develop items within each scenario. The Kalamazoo consensus statement issued at the Bayer-Fetzer conference¹⁶ served as the scaffolding during this process. The number of items was tailored to each scenario and ranged from 9 to 18. Two assessment tools were designed for each

case: a binary checklist (0=Not done, 1=done) and Likert-type rating scale (1=Unsatisfactory, 2=Acceptable, 3=Fair, 4=Good, 5=Very Good, 6=Excellent, 7=Outstanding).

The university's medical school maintains a group of SPs for training and evaluating medical students in clinical as well as communication skills. Six of these experienced SPs participated in the current study. The faculty evaluators were selected from among the examiners of post-graduate certification in Radiology. All raters were provided the study protocol, cases and methods of evaluation one week prior to the OSCE. In addition, SPs completed a 5-hour session with two master trainers one day prior to the OSCE. Case items were finalised based on the feedback received after the training session.

Prior to OSCE, each resident completed a short questionnaire documenting his/her gender, age, type of training institute and the year of training. The OSCE comprised six stations, each with a permanently stationed SP and a faculty member. Each encounter spanned 15 minutes, followed by independent assessment of the resident's performance by faculty and SP using both the checklist and the rating scale. Seven minutes were allocated for this step.

Data was dually entered and validated using Epi-Data v3.2. Data were then exported to SPSS v16.0 for further analyses. Frequencies (percentages) were computed for categorical and mean $\pm$ SD for continuous variables. Cronbach's alpha was calculated as a measure of internal consistency for each scenario. Cumulative scores of individual residents at each station were converted to percentage to allow for meaningful comparisons. Paired t-test and Pearson's correlation coefficient were used to compare scores on the checklist and the rating scale. Separate analyses were conducted for evaluations done by the faculty and the SPs. For all analyses, $p < 0.05$ was considered significant.

Results

A total of 42 radiology residents (69% females) with a mean age of 30.5 ± 3.4 years participated in the study. Of the residents, 23 (54.8%) were being trained at private

Table-1: Description of cases included in the study and alpha values of each vignette based on evaluation by faculty and standardized patients (SP).

Case	Description	No. of Items	Faculty		SP	
			Checklist	Rating Scale	Checklist	Rating Scale
A	Wrong diagnosis - foetal gender on prenatal ultrasound	18	0.6	0.9	0.8	0.8
B	Breaking bad news - Intrauterine demise in primigravida	15	0.4	0.9	0.4	0.9
C	Disagreement with colleague - post nephrectomy obstruction	9	0.4	0.9	0.6	0.7
D	Missed diagnosis - pneumo-peritoneum	9	0.7	0.8	0.8	0.8
E	Complication of procedure - Pneumothorax during a diagnostic pleural tap	15	0.4	0.6	0.8	0.9
F	Informed consent - Administration of Iodinated contrast medium	14	0.2	0.7	0.9	0.9

Table-2: Comparison of the use of evaluation tools (checklist and global rating scales) by Faculty and Standardized Patients (SP).

Case	% Score		Difference*		Correlation	
	Checklist Mean (SD)	Rating Scale Mean (SD)	Mean (SEM)	p	r	p
Faculty evaluations						
A	51 (14)	39 (12)	12 (2)	0.000	0.6	0.000
B	78 (9)	62 (14)	16 (1)	0.000	0.8	0.000
C	73 (15)	48 (14)	25 (1)	0.000	0.9	0.000
D	74 (23)	59 (15)	16 (2)	0.000	0.7	0.000
E	80 (8)	68 (7)	12 (1)	0.000	0.8	0.000
F	72 (10)	51 (9)	21 (1)	0.000	0.7	0.000
Overall†	72 (17)	54 (16)	17 (1)	0.000	0.8	0.000
SP evaluations						
A	65 (18)	68 (14)	-3 (1)	0.047	0.9	0.000
B	78 (9)	59 (14)	18 (2)	0.000	0.8	0.000
C	60 (18)	59 (17)	1 (1)	0.467	0.9	0.000
D	68 (27)	68 (17)	0 (3)	0.932	0.8	0.000
E	75 (16)	48 (14)	28 (2)	0.000	0.7	0.000
F	42 (21)	42 (16)	0 (1)	0.927	0.9	0.000
Overall†	65 (23)	57 (18)	7 (1)	0.000	0.7	0.000

*Paired t-test, two tailed, level of significance <0.05

†based on comparison of 252 ratings (42 residents x 6 stations).

Table-3: Comparison of evaluators (Faculty and Standardized Patients (SP)) on the use of checklist and rating scales.

Case	% Score		Difference*		Correlation	
	Faculty Mean (SD)	SP Mean (SD)	Mean (SEM)	p	r	p
Checklists						
A	51 (14)	65 (18)	-14 (2)	0.000	0.6	0.000
B	78 (9)	78 (9)	0 (1)	0.700	0.8	0.000
C	73 (15)	60 (18)	13 (2)	0.000	0.6	0.000
D	74 (23)	68 (27)	6 (3)	0.066	0.7	0.000
E	80 (8)	75 (16)	5 (2)	0.031	0.5	0.000
F	72 (10)	42 (21)	30 (3)	0.000	0.4	0.018
Overall†	72 (17)	65 (23)	7 (1)	0.000	0.5	0.000
Global Rating Scales						
A	39 (12)	68 (14)	-29 (2)	0.000	0.4	0.003
B	62 (14)	59 (14)	3 (0.5)	0.000	1.0	0.000
C	48 (14)	59 (17)	-12 (2)	0.000	0.5	0.001
D	59 (15)	68 (17)	-10 (2)	0.000	0.8	0.000
E	68 (7)	48 (14)	20 (2)	0.000	0.5	0.002
F	51 (9)	42 (16)	10 (2)	0.000	0.5	0.000
Overall†	54 (16)	57 (18)	-3 (1)	0.023	0.3	0.000

*Paired t-test, two tailed, level of significance <0.05

†based on comparison of 252 ratings (42 residents x 6 stations).

institutions, whereas 19 (45.2%) at public institutions. At the time of the study, 14.3% (n=6) participants were in the 1st year, 28.6% (n=12) in the 2nd, 33.3% (n=14) in the 3rd and 23.8% (n=10) in the 4th year of training.

Based on faculty ratings, all cases were found to have acceptable internal consistency (Table-1) with alpha values ranging from 0.6 to 0.9 for rating scale scores. The alpha values were comparatively diminutive when calculated against the checklist scores (range 0.2 to 0.7), suggesting that inter-item covariance improved enough with the use of rating scales to offset the effect of increased average item variance.

Average Intraclass correlation coefficient (ICC) was 0.6 (95% CI 0.4 to 0.8, $p<0.001$) for faculty ratings using the checklist. ICC was 0.6 (95% CI 0.4 to 0.8, $p<0.001$) for faculty ratings using the rating scale. Two-way random effects model was used. Statistics for individual stations were noted (Table-2).

The trends for Cronbach's alpha were very similar to that observed with faculty evaluations. Average ICC was 0.6 (95% CI 0.3 to 0.7, $p<0.001$) with the use of checklists and 0.7 (95% CI 0.5 to 0.8, $p<0.001$) with rating scales.

Analyses of the use of the same tool by different evaluators - faculty and SP - did not reveal any consistent

trends. Though there were some stations with excellent correlation between the two evaluators, overall, correlation coefficients of 0.5 and 0.3 for checklists and rating scales, respectively, were seen (Table-3).

Discussion

CIS is recognised as an important set of skills for radiology residents and consultants.⁶ Limitations of human resources and assessment tools are critical roadblocks to formal assessment of CIS in exit examinations of radiology residency programmes. Our study demonstrated moderate to strong correlation between checklist and rating scales for assessing CIS of radiology residents. However, if we consider absolute difference, it appears that a significantly higher score is awarded with the use of checklists compared to the use of rating scales. Both these observations remain valid regardless of the type of evaluator (faculty or SP).

The above observations, combined with slightly higher alpha values obtained with the use of rating scales, suggest a better ability of rating scales to evaluate CIS. van der Vleuten et al have argued that subjective methods have greater capacity for subtler assessment of skills as opposed to 'objectified' methods.¹⁵ They further opined that too much emphasis on 'objectification' may even have a negative impact on learning and assessment. Cohen et al. reported an excellent correlation between scores awarded by SPs on checklists and rating scales. In fact, subjective ratings were found to have better reliability for CIS assessment than the checklists.¹⁴ Cohen et al organised 26 items into five sections, with each section designed to represent a different aspect of CIS. At the end of every section, a single global rating scale was used. We used a slightly different layout with a one-to-one matching of items on checklists and rating scales but have similar findings supporting the rating scales.

The ease of administration and objectivity make checklist an attractive option.¹³ On the other hand, global rating scales are deemed limited in terms of reliability and comparability. However, the notion that objectivity and reliability are inseparable is fading fast.^{11,17} In fact, given enough sample size or sampling time (e.g 8-hours of testing), the reliability of diverse formats such as MCQs, Oral exam, OSCE and Mini-CEX (mini clinical evaluation exercise) tends to be very similar.¹¹ The rating scale, such as the one used in the current study, appears to be a good amalgam, with the objectivity of checklists and the discriminative ability of a graded scale.^{14,18} Cohen et al also reported that, compared to checklists, rating scales offer the same degree of reliability with a smaller number of cases.¹⁴

Regardless of the type of rating scale used, both faculty and SPs were capable of evaluating residents' CIS

with satisfactory fidelity. The intra-class correlation coefficient - a measure of agreement between raters — was significant for both checklist and rating scale. On the same note, internal consistency of individual vignettes was found to be satisfactory for both tools of evaluation, though slightly better numbers were seen for the rating scales. This, again, suggests that a rating scale allows a subtler differentiation of examinees' CIS while maintaining the reliability traditionally considered inherent in a dichotomous checklist.

Consistency in the type of evaluator (faculty or SP) is also an important consideration for meaningful comparisons across different exam sessions.¹⁹ Although, both evaluators were able to satisfactorily use the checklist and the rating scales, the overall correlation between SP and faculty evaluations was merely 0.5 and 0.3 for checklists and rating scales, respectively. The choice of evaluator would in turn have bearing on the choice of the assessment tool. Rating scale can be an excellent tool in the hands of an experienced faculty evaluator, but the use of dichotomous checklists might be more convenient for SPs having limited background in medical education, training and assessment. In the current study, all faculty members were experienced in OSCE evaluations, checklists and rating scales; SPs had no such experience. Considering SPs' current performance, coupled with the fact that training can enhance it, SPs come forward as a valuable resource for CIS training and assessments.²⁰

In a detailed analysis of the performance of SPs and expert raters in an OSCE, Han et al concluded that intensely trained SPs outperform experts on both checklists and rating scales for evaluating medical students' clinical skills.²⁰ Residents represent a higher level of expertise and it may be argued that a connoisseur would be able to better discern the taste that residents' CIS are likely to leave.^{19,21} However, Donnelly et al. also observed similar performance of faculty and SP raters for assessing CIS of surgery residents.¹⁷ Our study also suggested that adequately trained SPs can be satisfactory alternatives to faculty examiners. If possible, this will immensely improve the utilisation of financial and human resources.

There are certain limitations of this study that must be considered before drawing any direct or indirect implication from its results. Firstly, although all faculty members were qualified enough to be examiners for the post-graduate certification in radiology, the level of expertise in the particular areas of SP-OSCE may not be the same across the board. Secondly, the training SPs was limited to a single five-hour session. The SPs were experienced in undergraduate medical students' OSCEs. Arguably, a more intense training could have led to a

superior performance of SP evaluators. Thirdly, the faculty evaluators had an edge over SPs with respect to the time available for assessment, as they could complete their assessment in real time which was not possible in the case of SPs.

Conclusion

The study demonstrated that both checklists and rating scales can serve as satisfactory CIS assessment tools in SP-OSCE. The actual choice of assessment tool and the evaluator would depend on the context, available resources, goals of the exam (e.g formative vs summative), and the institutional environment.

Acknowledgments

The authors would like to thank all the participating institutions and their faculty members for their invaluable help in the organization and conduct of the study, which was funded by a grant (08101<RAD) from the Research Council of the Aga Khan University.

References

1. American College of Radiology. ACR Guidelines for communication of diagnostic imaging findings. Res 11. (Online) 2005 (Cited 2010 Oct 21). Available from URL: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/dx/comm_diag_rad.aspx.
2. Berlin L. Communicating findings of radiologic examinations: whither goest the radiologist's duty? *AJR Am J Roentgenol* 2002; 178: 809-15.
3. Williamson KB, Steele JL, Gunderman RB, Willkin TD, Tarvas RD, Jackson VP, et al. Assessing radiology resident reporting skills. *Radiology* 2002; 225: 719-22.
4. Schreiber MH. Communicating with the referring physician. The standard of care. *AJR Am J Roentgenol* 1997; 169: 343-5.
5. Olijeski SA, Homer MJ, Krackor WS. Incorporating ethics education into the radiology residency curriculum: a model. *AJR Am J Roentgenol* 2004; 183: 569-72.
6. Accreditation Council for Graduate Medical Education. Toolbox of assessment methods. (Online) 2007 (Cited 2007 December 29). Available from URL: www.acgme.org.
7. Joint Commission. Comprehensive accreditation manual for hospitals: The official handbook. Oakbrook Terrace, IL: Joint Commission, 2007.
8. Gunderman RB. Patient communication: what to teach radiology residents. *AJR Am J Roentgenol* 2000; 177: 41-3.
9. Graham J, Ramirez AJ, Field S, Richards MA. Job stress and satisfaction among clinical radiologists. *Clin Radiol* 2000; 55: 182-5.
10. Yudkowsky R, Downing SM, Ommert D. Prior experiences associated with residents' scores on a communication and interpersonal skill OSCE. *Patient Educ Couns* 2006; 62: 368-73.
11. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39: 309-17.
12. Boon H, Stewart M. Patient-physician communication assessment instrument: 1986-1996 in review. *Patient Educ Couns* 1998; 35: 161-75.
13. Hobgood CD, Riviello RJ, Jouriles N, Hamilton G. Assessment of communication and interpersonal skills competencies. *Acad Emerg Med* 2002; 9: 1257-69.
14. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. *Acad Med* 1996; 71: s87-9.
15. Van der Vleuten CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ* 1991; 25: 110-8.
16. Makoul G. Essential elements of communication in encounters: the Kalamazoo Consensus Statement. *Acad Med* 2001; 76: 390-3.
17. Donnelly MB, Sloan D, Plymale M, Schwartz R. Assessment of residents' interpersonal skills by faculty proctors and standardized patients: a psychometric analysis. *Acad Med* 2000; 75: s93-5.
18. Yudkowsky R. Should we use standardized patients instead of real patients for high stakes exams in Psychiatry? *Acad Psychiatr* 2002; 26: 187-92.
19. Finlay IG, Stott NC, Kinnersley P. The assessment of communication skills in palliative medicine: a comparison of the scores of the examiners and simulated patients. *Med Educ* 1995; 29: 424-9.
20. Han JJ, Kreiter CD, Park H, Ferguson KJ. An experimental comparison of rater performance on an SP-based clinical skills exam. *Teach Learn Med* 2006; 18: 304-19.
21. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklist do not capture increasing levels of expertise. *Acad Med* 1999; 74: 1129-34.