PEER ASSESSMENT: AN INNOVATIVE TOOL FOR ASSESSMENT IN HIGHER EDUCATION

Maneeza Kishwar, Amina Ahmad, Siraj ul Haque Sheikh, Javeria Ilyas

College of Physicians and Surgeons Pakistan, Karachi

ABSTRACT

Objective: To determine the correlation between peer and instructor assessment.
Study Design: Correlation study.
Place and Duration of Study: College of Physicians and Surgeons Pakistan (CPSP), in department of Medical Education (DME) from November 2011 to March 2012.
Material and Methods: All participants enrolled in MCPS-Health Professions Education program of 2011-2013 (n=21 in 1st session and n=21 in 2nd session) consented for inclusion in the study. During contact sessions the participants gave lecture presentations of 10 minutes duration, on topics of their choice but preferably related to medical education. The lecture was appraised by both the instructors and the fellow participants using evaluation forms with 5 point Likert scale. Each aspect of the presentation was assessed and scored.
Results: The Pearson's coefficient of correlation, between scores awarded by participants and scores awarded by instructors was found to be 0.63 (p= 0.002) in the 1st contact session. By 2nd contact session the correlation improved from moderate to high turning out to be 0.80 (p<0.001).
Conclusion: The study provides a sound evidence that; Peer assessment is comparable to faculty assessment and hence can be used more frequently to steer students' learning towards learning outcomes.

Keywords: Assessment, Peer assessment, Peer evaluation, Peer review

INTRODUCTION

With the introduction of competency based model of education, international medical education organizations e.g. CANMEDS, ACGME & GMC have defined varied set of competencies expected in a medical graduate including the ability to take necessary measures to improve learning process and assess self and peer learning[1-2]. Consequently practices which would inculcate these outcomes, need to be incorporated in medical education.

Assessment has been part and parcel of learning process since time immemorial. It creates the need for learning activity, scaffolds the process of learning and regulates different aspects of learning behavior; and at times it has more impact on learning than does teaching[3]. 'Assessment for learning' has to fulfill dual functions: assess the content and prepare learner for future learning.This involves a joint effort of the teacher and the learner in identifying learning goals according to the laid down standards, assessing the problems, reviewing and reflecting on data generated through feedback and peer assessment. This mode of assessment builds up the learners insight regarding their performance by means of information provided to them through two sources; instructors and peers. This in turn develops motivation and self-esteem, as they are able to assess themselves and hence improve learning[3-4].

Student involvement in assessment is usually seen in the form of peer assessment or self-assessment. Peer assessment is grounded in the philosophies of active learning, andragogy and social constructivism as it involves joint construction of knowledge through discourse[5].

In recent years much work has been done on the utility of peer assessment in learning, across the world. Peer assessment is a strategy in which students observe and judge the work of their contemporaries in accordance with previously defined criteria[1,6,7]. It can be used in different formats:quantitative (marks or score) or qualitative (feedback). The qualitative format may consist of written or verbal feedback, which may be unstructured or structured, according to a pre-defined rubric. Peer assessment now is being used for both formative and summative purposes[1].

Correspondence: Dr Maneeza Kishwar, 137 Block 3, Bahadurabad, Karachi, 74800
Email: maneezakishwar137@gmal.com
Received: 30 May 2014; Accepted: 02Jul 2014

The evidence shows that peer assessment enhances learning and contributes to learning efficiency and quality by promoting deep learning[7,8]. The clear advantage of this activity is an increased ability in the learner to reflect & make independent judgments regarding work done by others and self, and provide constructive feedback[9-10]. It promotes sense of involvement, responsibility and ownership of the assessment process[10]. It also builds critical thinking, self assessment, and confidence, therefore peer learning is so very necessary for effective learning[8].

Even though compatibility is often absent between two individuals of equal standing, a good correlation has often been found between peers and instructors despite the difference in their level of knowledge and experience. As a result peer assessment is now being employed in medical education to evaluate technical ability, clinical knowledge, judgment, interpersonal skills, dependability, personal appearance, reaction to pressure, ambulatory care skills, management of complex problems, management of hospitalized patients, empathy, responsibility and teaching skills[6].

In view of changing scenario of the world, deliberation is required regarding the feasibility of incorporating peer assessment in our setting, and as experienced by others; the foremost hurdle would be the question of credibility[7]. Hence this study was carried out in order to generate statistical evidence supporting or rejecting the credibility of peer assessment in relation to faculty assessment.

MATERIAL AND METHODS

This correlational study was carried out at College of Physicians & Surgeons Pakistan (CPSP) in the Department of Medical Education (DME) from November 2011 to March 2012.

The Department of Medical Education (DME), conducts MCPS- Health Professions Education program (MCPS-HPE),which is a 2 years' post graduate program for medical teachers in the context of medical education. It is a blended learning (hybrid) course, comprising of distance learning and three on campus learning sessions lasting for a period of 12 to 14 days each.

After obtaining permission from institutional and departmental heads, informed consent of study participants and acquisition of approval from institutional ethical review committee, data collection was done during the two contact sessions of MCPS-HPE program using a structured pro forma for evaluation of teaching skills.

The study sample consisted of n=21+21 in both first and second contact sessions. Out of these 21 participants; 14 were male & 7 were females. The ages ranged between 36–55 years and all had procured post graduate qualifications.

During the contact sessions, participants gave lecture presentation of 10 minutes duration, on topics of their choice but preferably related to medical education. These presentations were assessed formatively by fellow participants and two instructors. Each aspect of the presentation was appraised and scored on the content, organization and delivery of lecture along with methods employed to encourage students' participation and time management. A 16 item lecture rating evaluation proforma with 5 point Likert scale was used for scoring. The proforma was the modified version of Lecture Evaluation Form used for the HPE Program. For construct validation the proforma was vetted by a pair of subject specialists and pilot testing was done.

Participants also gave verbal feedback at the end of each lecture to justify very low and high scores given.

Prior to the activity, the participants were briefed about peer assessment per se and how the lecture presentations are to be observed and documented. Instructions regarding provision of constructive feedback was not given as the participants were well versed with the process.

Each participant was observed and rated by 20 peers and 2 instructors in both first and second contact sessions.

The individual student's peer evaluation score was the average of scores awarded by all teammates[11-13]. Similarly instructor's scores

were averaged to form an aggregate score for each student.

Statistical analysis was done with the help of Statistical Package for Social Sciences (SPSS) 20.

Descriptive statistics such as mean and standard deviation were calculated. Pearson's Correlation Coefficient analysis was calculated to investigate the strength of relationship between the mean scores of facilitator-based-assessment and peer-based-assessment. A *p*-value < 0.05 was considered as significant.

RESULT

Data analysis of the 1st contact session revealed; a mean of 69.93 with standard deviation of 4.896 on scores awarded by peers and a mean of 62.38 with standard deviation of 9.484 on scores awarded by instructors. Pearson's Correlation Coefficient demonstrated moderate correlation between scores awarded by participants and those awarded by instructors (table-1).

Data analysis of the 2nd contact session demonstrated a mean of 72.96 with standard deviation of 4.794 on peer awarded scores and a mean of 74.74 with standard deviation 7.034 on scores awarded by instructors. Pearson's Correlation Coefficient showed significant correlation between scores awarded by participants and those awarded by instructors (table-1).

DISCUSSION

Peers have been appraising the competence of their colleagues by various methods and this has formed the basis of referral in medicine throughout history. However systematic study of this form of assessment has been undertaken recently, with initial report being published by Topping in 1998[6].

In a systematic review done by Speyer, Pilz, Der Kruis & Brunings, reference has been made to multiple studies which state that faculty ratings are usually considered to be the gold standard in educational settings[14]. Authenticating learners scoring against teachers scoring implies establishing the concurrent validity of the instrument[5,12]. Our study establishes the evidence of concurrent validity as statistically significant ($p < 0.002$) correlation between the scores awarded by peers and faculty members using the same tool.

The result of correlation between instructor and peer rating was found to be 0.63 in the 1st contact session; which is statistically significant at $p < 0.002$. Similar result was reported in a study carried to determine the possibility of employing peers as examiners in OSCE[15]. Meta-analysis done by Falchikov & Goldfinch (2000) reported a mean overall value of r= 0.69 (0.14-0.99). Several studies have quoted comparable results[8,16,17].

Peer assessment has the ability to provide accurate and valid information regarding assessment but it gets influenced by several factors, which in turn affect the results, these factors include reliability, relationships, stakes and standardization[6].

Whether employed in summative or formative perspective; no statistically significant difference has been found between peer-assessed and instructor-assessed scores[9,10,18] even then it has been recommended that peer assessment should be introduced gradually, starting with low stake-formative assessments. Reason for this proposal being; less stressful approach encourages open discussion and reflection, which in turn contributes to learning and hence improves the reliability of the process[6,19]. Keeping this aspect in mind the two sessions in which formative assessments were carried out, were selected for this study.

The reliability of peer assessment is also influenced by number of other factors which include; number of performances observed, number of peers involved and the number of aspects of competence evaluated[6].

This means that assessment should be based on observations made in a variety of different situations; however this study is founded on information collected from observations carried out in two situations.

There were 20 peer assessors in both, first and second sessions; while 7 to 28 peers have been estimated to provide a reasonable reliability coefficient, any further increase in number produces little change in reliability[6,16].

For determining the concurrent validity, we correlated peer awarded scores with scores awarded by two instructors, as single teacher scores are considered unreliable assessment measurement[12]. Nevertheless, in the study conducted by Iqbal & Mehmood scores awarded by single teacher were employed for determining the correlation[18].

As recommended, reliability can be increased by assessing several aspects of competence under consideration. It has been suggested that approximately 10 questions covering different aspects of competence provide reasonable information and also keep the questionnaire manageable[6]. The proforma employed for assessing the lecture presentation in this study had 16 items covering various aspects of lecture presentation. Furthermore it had been modified from the document already in use by the department, which further validated it.

Despite fulfilling several criteria essential for establishing the reliability of results e.g. structured teacher evaluation proforma, and conducive learning environment, a few aspects were over looked such as in peer assessment, assessors' anonymity was not maintained with the view that it will help in improving their critiquing skills and managing issues that can be detrimental to group functioning[13]. Research does recommend that maintaining anonymity improves reliability by ensuring that evaluators score their peers without hesitation[6,7,16,20] . Furthermore it mitigates the effects of competition and friendship on the peer awarded scores[6,8]. Despite this, the results of this study demonstrate moderate to high correlation between scores assigned by peers and instructors in the two contact sessions.

Previously done studies endorse that one of the requirements of this mode of assessment is development of criteria/rubrics which are clearly understood by participants, written and verbal instructions along with training should be given in using this method; this creates familiarity and ownership of criteria which in turn contributes to the reliability and validity of the results[4-7,15,21]. The participants in this study were well versed in providing verbal feedback, but the concept of assessing their peers according to a given criteria and that too in a non-confidential manner was new to them. A brief explanation of the exemplar was given beforehand but, as the results of the 2nd contact session demonstrate; practice of using the methodology improves the correlation between the scores of participants and instructors[1,4,20].

The results of 2nd contact session showed correlation of 0.807 with *p* value of <0.001, which is highly significant. This improvement could have resulted from gaining experience of carrying out peer assessment and also

Table-1: Comparison of Instructors vs Peer assessment in two contact sessions.

	r	*p*
1st Contact Session	0.634	0.002
2nd Contact Session	0.807	< 0.001

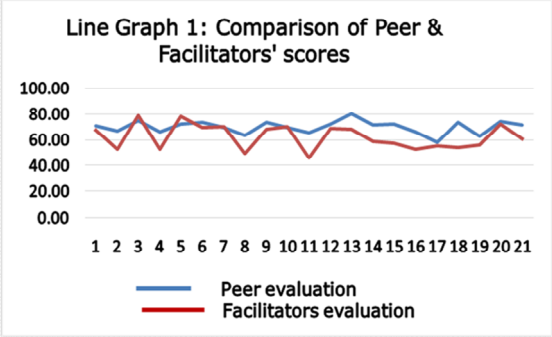

Figure-1: Line graph depicting comparison of scores awarded by peers and instructors in 1st contact session.

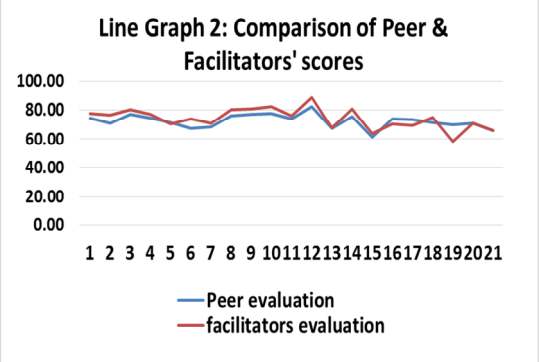

Figure-2: Line graph in depicting comparison of scores awarded by peers and instructors in 2nd contact session.

providing feedback without getting pressurized by interpersonal relationships. High correlation between ratings by peers and supervisors with correlation coefficient (r) ranging from; 0.86–0.98 and 'p' less than 0.001, has been reported in many studies[6,9,10,18].

In 1st contact session the scores awarded by peers were more 'bunched' than instructor awarded scores (Fig-1). Other studies have found similar tendency[4,12]. This may arise from students' hesitancy to use a scoring range similar to that employed by instructors, or it is due to students having lesser ability to discriminate between differing levels of given criteria. The use of restricted marking range lowers the reliability of the scores. However by 2nd contact session the participants had overcome the diffidence (Fig-2).

CONCLUSION

Peer evaluation is comparable to instructor assessment and can produce valid and reliable results. Hence it can prove to be a valuable resource for frequent utilization, which can complement instructor assessment in steering students learning towards the desired learning outcomes.

There are many reasons for the use of peer-based assessment. It not only acts as a meaningful learning activity, but also provides opportune feedback. Although student examiners have established themselves as viable alternative to faculty examiners in formative assessments, they however, need training in assessing a fellow student's performance in an objective way and provide meaningful and accurate feedback. Even though peer assessment is not as yet common in medical education, it is showing signs of becoming one and could easily lend itself to unethical conduct on the part of the assessor that is the peer. Therefore when implementing this format of assessment it is judicious to take note of the associated potential risks by keeping checks on reliability & validity and monitoring the process of implementation and the evaluators.

Further studies are now needed to evaluate and report aspects, which require development of a robust and feasible approach with a desirable educational impact.

CONFLICT OF INTEREST

This study has no conflict of interest to declare by any author.

REFERENCES

1. Chen J.Y. Why peer evaluation by students should be part of the medical school learning environment. Medical Teacher. 2012; 34: 603–606. DOI: 10.3109/0142159X.2012.689031
2. Creswell, J.W .Research Design: Qualitative, Quantitative and Mixed Methods Approaches. (2nd Ed).Thousand oaks, California: Sage Publication 2003: Pg 87-104.
3. Tillema HH. Appraisal of Practice Teaching Lessons by Mentors, Supervisors, and Student Teachers. Journal of Teacher Education. March/April 2009. Volume 60 Number 2. Pg 155-167. http://online.sagepub.com
4. Juwah C. Using Peer Assessment to Develop Skills and Capabilities. USDLA Journal. January 2003; Vol.17: No.1. ISSN 1537-5080
5. Falchikov N, Goldfinch J. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. Review of Educational Research. Fall 2000, Vol. 70, No. 3, pp. 287-322.
6. Norcini JJ. Peer assessment of competence. Medical Education. 2003; 37: 539–543. doi: 10.1046/j.1365-2923.2003.01536.x
7. Bay E. The Opinions of Prospective Teachers about Peer Assessment. GaziantepÜniversitesiSosyalBilimlerDergisi. 2011 10(2):909 -925.
8. Şahin S. An Application of Peer Assessment In Higher Education. The Turkish Online Journal of Educational Technology-TOJET. April 2008. volume 7 Issue 2 Article 1
9. Stefani L A J. Peer, self and tutor assessment: Relative reliabilities. Studies in Higher Education.1994; Vol19:1, 69-75. DOI: org/10.1080/03075079412331382153
10. Judy. R. Harris. Peer assessment in large undergraduate classes: an evaluation of a procedure for marking laboratory reports and a review of related practices. Advan in Physiol Edu. June2011.Vol 35,No.2:178–187; doi:10.1152/advan.00115.2010
11. Dannefer E F, Henson L C, Bierer S B, Grady-Weliky T A, Meldrum S, Nofziger A C et al. Peer assessment of professional competence. Medical Education. 2005; 39: 713–722.
12. DOI:10.1111/j.1365-2929.2005.02193.x
13. Magin D, Helmore P. Peer and Teacher Assessments of Oral Presentation Skills: How reliable are they?, Studies in Higher Education.2001.Vol26:3,287-298: http://dx.doi.org/10.1080/03075070120076264
14. Peterson C H, Peterson N A. Impact of Peer Evaluation Confidentiality on Student Marks. International Journal for the Scholarship of Teaching and Learning. July 2011;Vol. 5, No. 2. http://www.georgiasouthern.edu/ijsotl
15. Speyer R, Pilz W, Der Kruis J V, Brunings J W. Reliability and validity of student peer assessment in medical education: A systematic review. Medical Teacher.2011;33:e572–e58
16. Chenot J F, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B et al. Can student tutors act as examiners in an objective structured clinical examination? Medical Education 2007: 41: 1032–1038. DOI:10.1111/j.1365-2923.2007.02895.x
17. Kristi J Ferguson K J, Kreiter C D. Assessing the relationship between peer and facilitator evaluations in case-based learning. Medical Education 2007: 41: 906–908. DOI:10.1111/j.1365-2923.2007.02824.x
18. Moineau G, Power B, Pion A J, Wood T J, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. Medical Education 2011: 45: 183–191. DOI:10.1111/j.1365-2923.2010.03800.x
19. Iqbal Z, Mahmood N. Compatibility of Peer Assessment and Teacher Assessment in Observational Situations: An Emerging Assessment Tool in Higher Education. Bulletin of Education and Research December 2008, Vol. 30, No. 2 pp. 61-77
20. Evans A W, Leeson R M A, Petrie A. Medical Education 2007: 41: 866–872. DOI:10.1111/j.1365-2923.2007.02819.x
21. Gukas S D, Miles S, Heylings D J, Leinster S J. Medical students' perceptions of peer feedback on an anatomy student-selected study module. Medical Teacher. 2008;30: 812–814.
22. Schönrock-Adema J, Heijne-Penninga M, A J van Duijn M, GeertsmaJ, Cohen-Schotanus J. Medical Education 2007: 41: 836–842.