

Compatibility of Peer Assessment and Teacher Assessment in Observational Situations: An Emerging Assessment Tool in Higher Education

Zafar Iqbal* & Nasir Mahmood**

Abstract

This study involved students enrolled in the 3rd semester of a Master of Educational Research and Assessment program at a teacher training institute during their teaching practice. The purpose was to explore the potential of peer assessment as a dependable alternative to teacher assessment in higher education. The student in the sample observed their class fellows during their teaching practice. Each student was observed by twenty two class fellows and supervising teacher at least once during teaching practice. This paper reports about eleven students selected on the basis of performance in course work (five students with highest and six with the lowest grades). An observation schedule developed by researchers was used for recording observational information. Mann-Whitney U test was used to find compatibility in peer and teacher assessment on each of the ten aspects addressed in observation. Academically high performing student's observation of peers was relatively similar to the observation made by the supervising teacher. A greater degree of agreement in assessment on traits involving tangible (directly measurable) characteristics was observed while more variation existed on measurement of abstract traits.

Introduction

According to Dochy (2001) assessment not only refers to measuring, but also to involvement of students, application of knowledge and skills, integration in the learning environment, knowledge construction instead of knowledge reproduction in real life situation. In this concept, students are regarded as independent, autonomous and exploring individuals who direct their own learning processes so their role in the assessment process is different and depends less on the teacher and more on themselves. This assessment concept is the shift of responsibilities from the teacher to the student. This shift leads towards a model of sharing responsibilities among teacher and students for assessing the progress of the students. The teacher is no longer the only person responsible for the assessment process.

*Ph.D. Scholar, IER, University of the Punjab, Lahore

** IER, University of the Punjab, Lahore

The idea of involving students in their own and peer assessment gained reasonable acceptance among the professionals keeping in view increasing access to higher education, greater value placed in appreciating and nurturing diversity among students enrolled in higher education, and dealing with challenges for teachers in providing multiple, meaningful feedback to individual students.

By engaging students directly in the assessment process, peer assessment allows the provision of feedback to be shared among teachers and students, with potential learning benefits (Vu & Dall'Alba, 2007).

Peer assessment practices have been applied in many institutions for over 50 years. Although Kane and Lawler in their publication of 1978 addressed the failure of recognition of the use of peer assessment, many studies since then have proven the importance to involve students in the assessment process through peer assessment (Arter, 1996; Boud *et al.*, 1999; Dochy & McDowell, 1997; Dochy, 2001).

Peer assessment may prove to be a valuable tool for evaluating colleagues. Peer assessment may guarantee a more objective evaluation (Steensels, et al., 2006). Also, Peer assessment reinforces the message to students that not only individual achievement is evaluated, thereby inciting students to work together. Finally, Peer assessment has been recognized as a skill necessary for professional practice (Thomas, 1997).

Peer assessment involves students directly in the assessment process which was traditionally in the hands of teachers only. Peer feedback has a different quality to feedback from people in positions of authority like teachers. Potential benefits and limitations of peer assessment have been identified in the literature (Vu & Dall'Alba, 2007).

Effect of Peer Assessment on Student Learning

One of the greatest benefits of peer assessment is the increased likelihood of learning from peers and the assessment process. Peer assessment requires students to closely scrutinize their peers' work, guided by criteria of desired performance. This activity helps them to diversify their own approaches and strategies in undertaking a learning task and can deepen understanding about high- or low-quality performance (Gibbs, 1999). With insight into performance quality, students can better understand their own learning and the feedback they receive from assessment.

An additional benefit of peer assessment is that it can be an appropriate arena for independent learning. Peer assessment requires students to make independent judgments and provide comments on the work of their peers (Brown & Knight, 1994; Brown & Glasner, 1999). Student enjoyed carrying out the peer assessment and considered it beneficial for their learning.

Peer assessment has been found to promote learning. Falchikove (1995) employed peer assessment in oral presentations to enhance learning processes and sharpening critical. Peer assessment methods have the

potential to strengthen the link between tutors' feedback and students' learning (Orsmond et al., 2000). Therefore, peer assessment has been used not only as an assessment method, but also as a method to enhance learning.

Issues in Peer Assessment

Peer assessment can be intended to either supplement or substitute teacher assessment. Peer assessment was not meant to be substitutional, but the relation to teacher assessment differed. The process of peer assessment is new and maximum students find it difficult and feel uneasy to do it. Students found that criticizing their friends was difficult. Ellis (2001) reported student voices in his research by quoting example of comments made by students on assessing their peers as, "You don't want to mark a fellow student too harshly" or "I feel uncomfortable assessing other student" (p.292). The reluctance of students in assessing their peers has made assessment by students a "risky and unfair" (Kwan & Leung, 1996) endeavor despite being very useful in many respect.

Several studies have reported poor correlations between marks awarded by students and by tutors (Falchikov, 1986; Orsmond, Merry, & Reiling, 2000). Issues concerning the use of students as assessors, including their lack of ability to discriminate between levels of performance, favoring mutually and their reluctance to judge their peers (Falchikov, 1995; Li, 2001) are most widely reported reasons in the literature.

Researchers have pointed out several potential limitations in the use of peer assessment (Brown & Knight, 1994; Brew, 1999). The most frequently reported limitations include; the use of peer assessment may involve increased time for the purpose, due to lack of familiarity with procedures and skills as assessors, students may be unrealistic or biased in assessing peers, and possibility of friction among peers. It will take a reasonable time for teachers to trust the peer assessment and for students to accept the assigning of marks by peers as uncontroversial (Brew, 1999). Even if it previously mentioned both concerns are dealt with amicably, asking students to assess their peers require very skilful handling otherwise "peer assessment can destroy a group's morale and working relationships" (Ellis, 2001).

Peer Assessment as Emerging Assessment Tool in Higher Education.

There is an increasing amount of attention being given in higher education to the concept of peer assessment, which can be understood as an educational arrangement in which students assess the quality of their fellow students' work and provide each other with feedback (Berg, Admiraal & Pilot, 2006). The teachers are very cautious in approving peer assessment as reliable and valid source of assessment because of the issues mentioned earlier. This cautious behavior has limited the scenarios in which peer assessment is practically used. Peer assessment in higher education is used

either with the assessment of individual contributions to group work (Earl, 1986; Freeman, 1995) or with the extent to which such marks may be valid for grading purposes.

A crucial message emerging from literature debating the use of peer assessment in group work situation recommended some pre-conditions to using peer assessment as a tool for evaluating student contributions to group work. First, the group size cannot be too large. Second, students have to be aware of the criteria used for assessment. Finally, students need to be skilled in carrying out peer assessment. In other words, students have to learn to assess one another. Brown and Knight (1994) claim that students', who participate for the first time in peer assessment, need the tutor to provide them with the performance criteria on which they have to assess their peers. When they have mastered this skill, students are capable of assessing their peers in an accurate way (Segers & Dochy, 2001).

The broad agreement in the peer assessment as a potential authentic tool, although in very limited assessment scenarios, and alternative to teacher assessment is widely acknowledged. The concern is about determining the capacity of the students to skillfully conduct it and its acceptability among teachers and students. Moreover, exploring the variety of situations where peer assessment can be used as alternative source of assessment. This study took the challenge of exploring the compatibility of teacher and peer assessment in observational situation, generally open to subjectivity in assessment even when different teachers are assessing students. The result will explicitly report on degree of variation in teacher and peer assessment in observational situation during teaching practice course in a teacher education program. Keeping in view, some of the issues raised in relation to competence and knowledge of students in literature, student's academic excellence as a potential source of variation in peer assessment will also be examined.

Context of the Study

This paper is based on part of the data collected through multiple sources during an eight week teaching practice activity carried out by twenty five students of a master's degree program at a teaching training institute for partial fulfillment of their degree program. The students who participated in this teaching practice had already completed courses in educational psychology, curriculum and instruction and methods of teaching. These courses were aimed at providing them with sufficient theoretical understanding of the how students learn, suitability of various teaching methods in context of the content to be taught, the art of effective planning and execution of lessons and techniques of becoming effective teachers. They were also trained in techniques of observational assessment in their course of Educational Measurement and Assessment. These courses along

with their individual experiences help them in imagining themselves as teachers of their own choice.

Method

This section of the paper will describe design of the study, nature of participants, development of instruments and procedure of data collection

Design of the Study

The study encompasses various aspects of the eight week teaching practice resulting in variety of forms of data collection through questionnaire, observations, class discussions, group reflection and individual interviews.

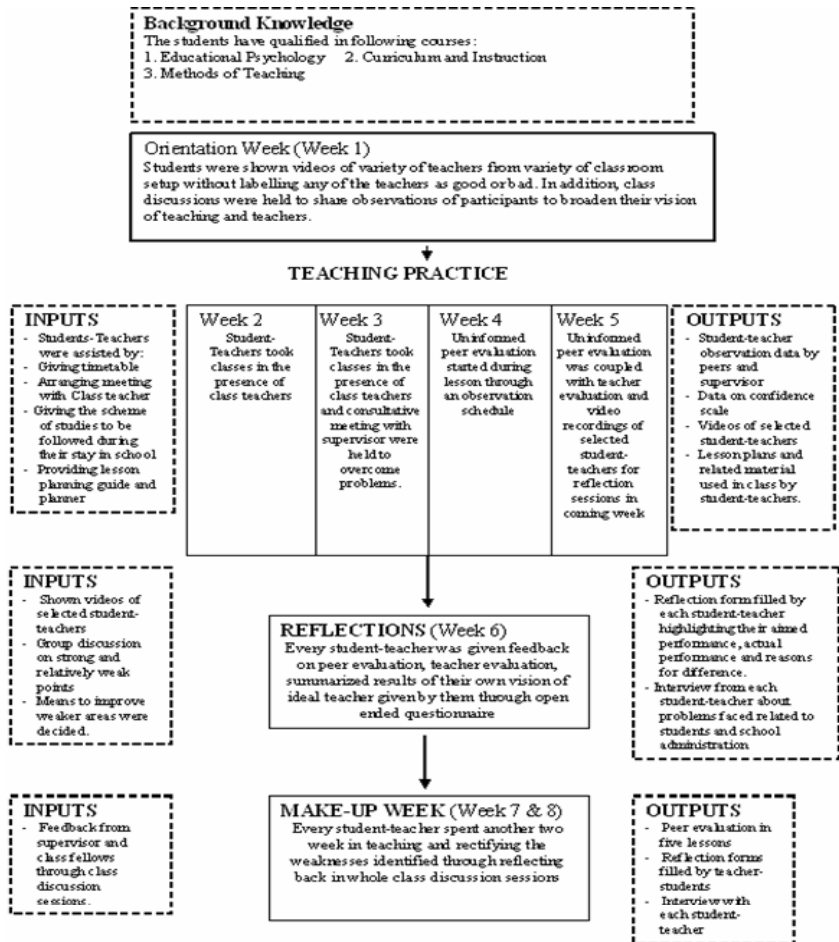


Figure 1: Design of the eight week teaching practice course/ activity

Figure 1 displays the design of the study to conceptualize the environment in which the study was conducted and time space in which peer assessment was conducted.

The peer observations were carried out during week four, five, seven and eight of the study. An observational schedule was used by supervising teacher and peers to teaching practices of student teachers. Every student teacher was observed by his/her peers at least once during four weeks. Thus, every practicing teacher was observed for almost 22 times during his/her teaching practice alongside at least one observation by the supervising teacher using the same observation schedule.

Nature of Participants

Participants of the study were students enrolled in the third semester of a Master of Educational Research and Assessment program at a teaching training institute during their teaching practice. Keeping in view the research questions set for this study eleven students (five students having highest grades and six students of lowest grades in selected subjects) were selected out of the total of twenty five students on the basis of their performance in the relevant subjects during their course work aiming at developing teaching skill among them.

Instrument

The instrument used for measuring compatibility of peer assessment and teacher assessment was an observation schedule. It was based on various sub-constructs deemed important in classroom teaching by various researchers and information given by the student -teachers through an open ended questionnaire about variables potentially important for good teaching. Table 1 explains the nature, scope and reliability of the sub-constructs of the observation schedule.

The observer subjectivity has always remained a debatable issue (Falchikov, 1995; Li, 2001; Ellis, 2001) in observational measurement. Being cognizant of the fact, researchers used two step scales for recording observational data. The first part of the scale required observers to record tallies in the provided five minute time slots to represent the frequency of happening of the sub-construct being observed during the lesson. As a second step of the observation, the observers were required to count the frequency of happening and its suitability by the time slot in which it happened, keeping in view the need of the lesson, to decide the appropriate coding on six-point scale ranging from strongly agree to strongly disagree.

Table 1

Description of sub-constructs and reliability of the observation schedule

Sub construct	Scope	items	Example	α
Content expertise	Teacher's command on content being delivered in classroom	3	Used examples from real life situations beyond textbook to explain concept studied	0.320
Classroom activities	Teacher's preparedness of lecture	4	Adjusted lesson plan when class proceeding demanded	0.449
Nature of permitted involvement of students	Type and organization of student involvement during classroom	5	Willingly accepted students' interruption during lesson for asking lesson related questions	0.538
Teacher response to students questions as indicator of students centered teacher	Motivate students during the class	5	Teacher praised student for asking question	0.561
Effect of student involvement on their learning	Type of learning occurred through interaction	5	Students crossed-questioned frequently	0.361
Student teacher interaction	Nature of interaction in the classroom	5	Builds class talk on students' argument	0.196
Instructional methodology	Method of instruction being used in the class	4	Method of teaching was same as stated in this stage of lesson plan	0.247
Teacher behavior towards students	Teacher's fairness and sincerity	4	Bias free	0.519
Effective use of A.V. aids	Proper and suitable usage of A.V. aids	3	Shift from one to another A.V aid was appropriate with respect to content taught	0.418
Assessment	Classroom assessment	5	Teacher visited students while they were working in groups	0.659

Reliability of the observation schedule = 0.774

Procedure of Data Collection

Training for all students involved in peer observation was conducted by the supervising teacher in a whole group situation. The trainer explained the nature, scope and indicators for measuring each sub-construct in the observation schedule in a lecture-cum-discussion session of about four hours on day one of the training. On the second day of training, all students were shown a forty minute lesson video already recorded for this purpose and they were required to assess the lesson using the observation schedule. This hands-on experience was meant to give experience of real classroom like situation where they will be actually using it during the course. A summary of the rating by different students was discussed with the whole group, particularly the ratings by the students where the difference was quite large. The discussion aimed at developing skill of critically analyzing basis of one's own observational judgment without actually criticizing any individual observer. The exercise helped the students in bringing clarity about sub-constructs and their scope. The discussion on each item in the scale assisted in evolving a uniform mechanism for utilization Of this schedule in the classroom. The training produced an acceptable degree of consensus and clarity in the process of recording observations.

The selection of students to be videoed was made on the basis of their performance in pedagogy courses on campus. The students having average of more than 75% marks in the courses mentioned in Figure 1 were classified as high performers and students between 50-74% marks were classified as low performers. The students closest to an average of 50% marks in selected courses were included in the sample. The lowest five were to be selected but the average marks of fifth and sixth students were same, therefore both were selected thus raising the number of low performing students to six. Similarly five students with highest average score in the selected courses were selected.

This was followed by four weeks of data collection. The data comprised of classroom observation of eleven selected students by twenty two peers at least once by each. The lessons delivered by the selected eleven students during which they were observed by their peers and supervising teacher were videoed too. The video recordings were later on used for cross verification of homogeneity in the process of observational recordings when each of the observers reviewed at least one of their observations jointly with supervising teacher. The process of reviewing resulted in some modifications in the observational data in consultation with the observer. Although number of such changes in the coding was not very high but it ensured a certain minimum degree of uniformity in the observational process.

Table 2

An overall comparison of peer and teacher assessment in observational situation

Sub-constructs	Assessor	N _{obs}	Mean per item	Std. Dev.	Negative Rank (N _n)		Positive Rank (N _p)		Z-value
					N	Mean	N	Mean	
Content expertise	Peer	242	3.09	0.60	6	6.50	5	5.40	-0.533
	Teacher	11	2.93	0.55					
Classroom activities	Peer	242	4.74	0.58	11	6.00	0	-	-2.934*
	Teacher	11	3.11	0.93					
Nature of permitted involvement of students	Peer	242	3.54	0.31	9	6.89	2	2.00	-2.578*
	Teacher	11	2.45	1.04					
Teacher response to student's question as indicator of student centered teacher	Peer	242	4.02	0.37	10	6.40	1	2.00	-2.756*
	Teacher	11	2.64	0.94					
Effect of student involvement on their learning	Peer	242	4.33	0.30	11	6.00	0	-	-2.934*
	Teacher	11	3.00	1.21					
Student teacher interaction	Peer	242	4.13	0.35	10	6.40	1	2.0	-2.758*
	Teacher	11	3.18	0.74					
Instructional methodology	Peer	242	3.36	0.54	6	7.83	5	3.80	-1.245
	Teacher	11	3.09	0.90					
Teacher behavior towards students	Peer	242	5.29	0.262	11	6.00	0	-	-2.934*
	Teacher	11	3.95	0.850					
Effective use of A.V. aids	Peer	242	3.45	0.308	7	6.71	4	4.75	-1.245
	Teacher	11	2.84	1.345					
Assessment	Peer	242	4.65	0.868	9	6.89	2	2.00	-2.578*
	Teacher	11	3.05	1.1387					
Total	Peer	242	4.05	0.309	0	-	11	6.00	-2.934*
	Teacher	11	3.03	0.772					

*p<0.05, Number of peers involved (N_p) = 11 Number of Teacher involved N_T = 1 N_{obs} = Number of Observations

Results and Conclusion

The difference of peer and teacher assessment was analyzed in observational situation. This was examined through finding the proximity between student and teacher assessment. It will be analyzed whether student academic excellence is a factor in determining the proximity between student and teacher assessment.

Table 2 represents the overall comparison of compatibility between teacher and peer assessment. The results showed statistically significant difference in the assessment of peers and teacher on seven of the ten sub-constructs used in class observation. The data was analysed using Mann-Whitney U test for non-parametric data due to less number of subjects involved in this study. Content expertise, instructional methodology and effectiveness of AV aids were the observed sub-constructs where teacher and peer assessment was compatible with each other. The indicator measured under the above mentioned sub-constructs included command of the teacher on content being delivered in the classroom, pedagogy used to convey the content to the students and use of helping material and their suitability for the purpose. These indicators were directly measurable and tangible in nature in terms of their direct demonstrability in the class during teaching. The sub-constructs on which peer and teacher assessment were different, involved indicator which were intangible requiring indirect inferences to be concluded. The intangibility of these sub-constructs involved element of self-judgement on the part of the observer which opened the possibility of a certain degree of subjectivity which became source of variability in the measurement.

Despite the fact that number of teacher observation is considerably less than the number of peer observations, the mean difference in peer and teacher assessment is worth considering. The mean score of peer assessment is higher than teacher assessment in all the sub-constructs. The higher mean score of peer assessment does not necessarily leads to intentional subjectivity by the peers but may also be caused by relative shallowness of ideas possessed by students (Li, 2001). The students concept of teaching and ideals of good teaching practice may not be same as teacher because they are just entering the field and students usually don't have a clear philosophy of teaching in which they belief.

The analysis for compatibility between academically high performing students' peer assessment and teacher assessment is shown in Table 3 and 4. It is very evident that in observational situation the students with academically better performance have greater degree of compatibility with teacher assessment as compared to academically less competent students. The assessment of academically high performing students and teachers was similar in seven sub-constructs out of ten measured while there were only four sub-constructs where academically low performing students had similar assessment as the teacher.

Table 3

Comparison of high performing peers and teacher evaluation in observational situation

Sub-constructs	Assessor	N _{obs}	Mean	Std. Dev.	Negative Rank (N _n)		Positive Rank (N _p)		Z-value
					N	Mean	N	Mean	
Content expertise	Peer	110	2.96	0.60	2	2.50	3	3.33	-0.674
	Teacher	5	3.13	0.51					
Classroom activities	Peer	110	4.58	0.47	5	3	0	.00	-2.023*
	Teacher	5	3.25	1.13					
Nature of permitted involvement of students	Peer	110	3.51	0.36	3	4	2	1.5	-1.214
	Teacher	5	2.52	1.18					
Teacher response to student's question as indicator of student centered teacher	Peer	110	4.02	0.42	4	3.50	1	1.00	-1.753
	Teacher	5	2.72	0.96					
Effect of student involvement on their learning	Peer	110	4.28	0.34	5	3	0	.00	-2.023*
	Teacher	5	3.12	0.94					
Student teacher interaction	Peer	110	4.09	0.44	4	3.50	1	1.00	-1.753
	Teacher	5	3.12	0.79					
Instructional methodology	Peer	110	3.53	0.43	4	3.50	1	1.00	-1.753
	Teacher	5	2.90	0.95					
Teacher behavior towards students	Peer	110	5.23	0.25	5	3.00	0	.00	-2.023*
	Teacher	5	4.10	0.65					
Effective use of A.V. aids	Peer	110	3.43	0.26	4	3.25	1	2.00	-1.483
	Teacher	5	2.33	1.45					
Assessment	Peer	110	4.26	0.70	4	3.50	1	1.00	-1.753
	Teacher	5	2.68	0.94					
Total	Peer	110	3.99	0.33	0	.0	5	3.0	-2.023
	Teacher	5	2.99	0.71					

*p<0.05, Number of peers involved (N_p) = 5 Number of Teacher involved N_T = 1 N_{obs} = Number of Observations

It is interesting to note that students, irrespective of their academic performance were equally compatible with teacher assessment in assessing content expertise, instructional methodology used, effectiveness of AV aids and class assessment techniques used of class teacher. The sub-constructs mentioned above were all physical in nature in terms of their happening in the classroom. Therefore, it leads to the conclusion that assessment of physical/tangible traits is relatively objective in its measurement. Thus teacher can rely on peers’ assessment in measuring such traits.

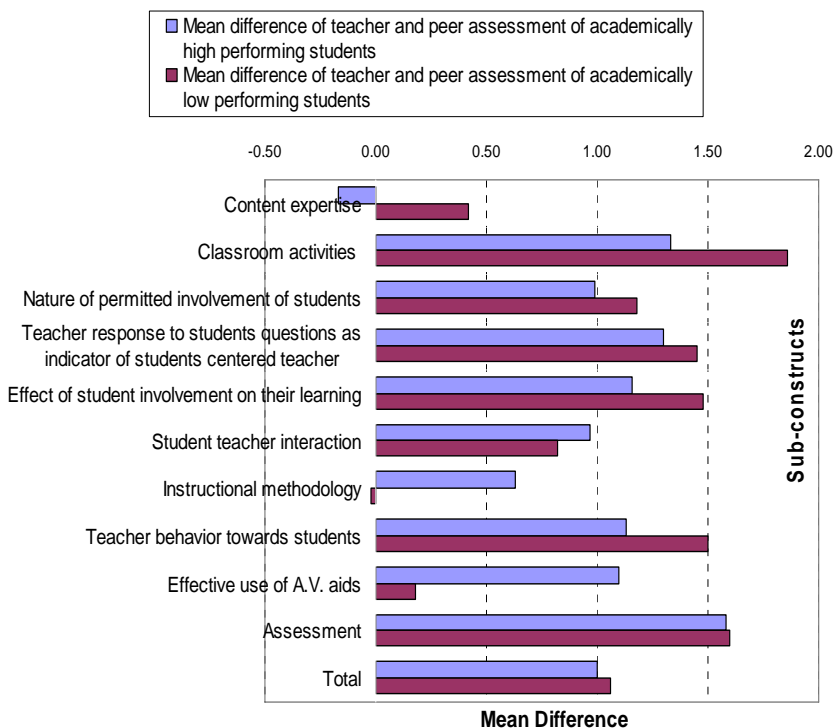


Figure 2: Mean Difference of teacher assessment and assessment by academically high and low performing peers

The sub-constructs on which both high and low performing students had significant difference with teacher assessment included nature of permitted involvement of students, teacher response to student’s question as indicator of student centered teacher and effect of student involvement on their learning.

Table 4

Comparison of low performing peers and teacher evaluation in observational situation

Sub-constructs	Assessor	N	Mean	Std. Dev.	Negative Rank (N _n)		Positive Rank (N _p)		Z-value
					N	Mean	N	Mean	
Content expertise	Peer	132	3.20	0.64	4	3.75	2	3.00	-0.943
	Teacher	6	2.78	0.58					
Classroom activities	Peer	132	4.86	0.67	6	3.50	0	-	-2.201*
	Teacher	6	3.00	0.82					
Nature of permitted involvement of students	Peer	132	3.58	0.29	6	3.50	0	-	-2.201*
	Teacher	6	2.40	1.02					
Teacher response to student's question as indicator of student centered teacher	Peer	132	4.02	0.36	6	3.50	0	-	-2.201*
	Teacher	6	2.57	1.00					
Effect of student involvement on their learning	Peer	132	4.38	0.28	6	3.50	0	-	-2.201*
	Teacher	6	2.90	1.48					
Student teacher interaction	Peer	132	4.16	0.30	6	3.50	0	-	-2.201*
	Teacher	6	3.34	0.78					
Instructional methodology	Peer	132	3.23	0.63	2	4.50	4	3.00	-0.314
	Teacher	6	3.25	0.91					
Teacher behavior towards students	Peer	132	5.34	0.23	6	3.50	0	-	-2.201*
	Teacher	6	3.84	1.03					
Effective use of A.V. aids	Peer	132	3.46	0.37	3	4.00	3	3.00	-0.314
	Teacher	6	3.28	1.20					
Assessment	Peer	132	4.97	0.92	5	3.80	1	2.00	-1.782
	Teacher	6	3.37	1.27					
Total	Peer	132	4.12	0.31	0	-	6	3.50	-2.201*
	Teacher	6	3.06	0.89					

*p<0.05, Number of peers involved (N_p) = 6 Number of Teacher involved N_T = 1 N_{obs} = Number of Observations

These sub-constructs required peers to infer judgment indirectly through set of happenings of varied nature in classroom which needed experience and vision (Brown & Knight, 1994) where teacher have a clear edge over students.

Nature of involvement of students in lesson, teacher questioning, and student teacher interaction were the sub-constructs where assessment by high performing students and teacher was compatible but assessment by low performing students and teacher was different. This indicates that student's own understanding of the relevant content makes significant difference in their competence to assess peers. Academically better students were those having higher achievement in courses like educational psychology, methods of teaching and curriculum and instruction which is an evidence of knowing more about the content of sub-constructs being measured helped in bringing compatibility between peer and teacher assessment.

Figure 2 shows a comparison of mean difference in assessment by teacher and assessment by academically high and low performing students. The gap between assessment by teacher and assessment by low performing peers is wider than assessment by teacher and high performing peers on almost all sub-constructs except instructional methodology.

There is sufficient evidence in the results that peer assessment has potential to be trusted as alternative to teacher assessment, particularly, when the constructs to be measured are tangible in nature. In case of intangible constructs, assessment by academically better students is more reliable than academically less competent students.

Discussion

There has been a continuous debate on the potential of relying on peer assessment as a reliable alternative to teacher assessment. This research intended to add some clarity to this debate by empirically investigating the possibilities and issues in compatibility of peer and teacher assessment in observational situation. It is generally perceived that it is harder to have compatibility even between two equally experienced and qualified assessment experts because of the judgmental subjectivity which is part of observational assessment. Therefore, it was not expected to have absolute compatibility between teacher and peer assessment when it is obvious that other than complexities involved in observational process the two also differ in their experience, exposure and knowledge of the constructs being assessed.

The major findings of the study were encouraging to the extent that similarities were found in the peer and teacher assessment on number of sub-constructs. Especially, in the assessment of constructs involving observation of physically verifiable characteristics but a tendency of scoring leniently as compared to teachers was observed in constructs of relatively abstract

nature. The reason might be less exposure, limited knowledge or inter personal relation with the individual being observed/ assessed.

The evidence further suggested that academic excellence of peers involved in observational assessment also helps in brining proximity between teacher and peer assessment. If teacher is assumed as authentic source of student assessment then it can be said that academically high performing students are more realistic and evaluate their peers according to their level of performance in the classroom. Another related finding suggested that students with lesser academic performance also had certain degree of similarity in their observational assessment on directly observable sub-constructs but their assessment on intangible/indirectly measurable sub-constructs was different than the teacher. This has two implications for using student assessment as alternative to teacher assessment. Firstly, when it is intended to involve students in assessment process, their involvement should be limited directly observable measures and should not be trusted where inferential judgments are to be made on the basis of constructs being observed. Secondly, the assessment made by peers having sufficient content knowledge about construct being observed and self-exposure to the activity to be assessed should carry more weight-age as compared to other peers.

One more important trend in peer assessment was the higher rating by peers as compared to teacher irrespective of their academic performance. This is a commonly perceived reservation hindering recognition of peer assessment's compatibility with teacher assessment. This argument has its merit but at the same time it is also important to recognize that relatively shallow knowledge, limited exposure and experience of assessment lead to tendency of giving higher rating in general.

The results also highlighted a disagreement between peer and teacher assessment in areas where decision involved fair degree of subjectivity in assessment i.e. the constructs which have no hard and fast way of assessment but depend on inference deduced from set of data collected. The compatibility can also be increased by structuring the tool of observation to the extent where it leaves less room for subjective inferences. While considering this option it should also be considered that bringing structure at the cost of compromising the meaning of the data is not suggested. It is recommended to keep a balance between structure and meaningfulness of the assessment to be made.

References

- Arter, J. (1996). Using assessment as a tool for learning, in: R. Blum & J. Arter (Eds) *Student performance assessment in an era of restructuring* (Alexandria, VA, Association for Supervision and Curriculum Development).
- Berg, I. V. D., Admiraal, W. & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education*, 31(3), pp. 341-356.
- Boud, D., Cohen, R. & Sampson, J. (1999). Peer learning and assessment, *Assessment and Evaluation in Higher Education*, 24 (4), pp. 413-426.
- Brew, A. (1999). Towards autonomous assessment: using self-assessment and peer assessment, in: S. Brown & A. Glasner (Eds) *Assessment matters in higher education: choosing and using diverse approaches* (Buckingham & Philadelphia, SRHE & Open University Press), 159-171.
- Brown, S. & Glasner, A. (Eds) (1999) *Assessment matters in higher education: choosing and using diverse approaches* (Buckingham & Philadelphia, SRHE & Open University Press).
- Brown, S. & Knight, P. (1994) *Assessing learners in higher education* (London & Philadelphia, Kogan Page).
- Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Res. Dial Learning Instruction*, 2(11) pp. 11-20.
- Dochy, F. J. R. C. & Mcdowell, L. (1997) Assessment as a tool for learning, *Studies in Educational Evaluation*, 23 (4), pp. 279-298.
- Earl, S. E. (1986). Staff and peer assessment – measuring an individual's contribution to group performance. *Assessment and Evaluation in Higher Education*, 11, pp. 60-69.
- Ellis, G. (2001). Looking at ourselves- self assessment and peer assessment: Practice examples from New Zealand. *Reflective Practice*, 2(3) pp. 289-302.
- Falchikov, N. (1986). Product comparison and process benefits of collaborative peer group assessment. *Assessment & Evaluation in Higher Education*, 11, pp. 146-166.

- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32(2), pp. 175-186.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education*, 20, pp. 289-300.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn, in: S. Brown & A. Glasner (Eds) *Assessment matters in higher education: choosing and using diverse approaches* (Buckingham & Philadelphia, SRHE & Open University Press), 41–53.
- Kwan, K. & Leung, R. (1996) Tutor versus peer group assessment of student performance in a simulation training exercise, *Assessment and Evaluation in Higher Education*, 21 (3), pp. 205–214.
- Li, L. K. (2001). Some refinements on peer assessment of group projects. *Assessment & Evaluation in Higher Education*, 26, pp. 6-18.
- Orsmond, P., Merry, S. & Reiling, K. (2000). The use of students derived marking criteria in peer and self assessment, *Assessment and Evaluation in Higher Education*, 25(1), pp. 23-38.
- Segers, M. & Dochy, F. (2001). New assessment forms in problem based learning. The value added of the students;’ *Perspective Studies in Higher Education*, 26, pp. 327-343.
- Steensels, C., Leemans, L., Buelens, H., Laga, E., Lecoutere, A., Laekeman, G. & Simoens, S. (2006). Peer assessment: a valuable tool to differentiate between student contributions to group work? *Pharmacy Education: An International Journal for Pharmaceutical Education*, 6(2), pp. 111-118.
- Thomas, R. E. (1997). Problem based learning: measurable outcomes. *Medical Education*, 31, pp. 320-329.
- Vu, T. T. & Dall’Alba, G. (2007). Students’ experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, 32(5), pp. 541-556.