



RELIABILITY OF PEER FEEDBACK IN ESL CONTEXT

Sevwandi Ganesha Pathmaperuma
Sri Lanka Institute of Advanced Technological Education,
Colombo, Sri Lanka

Abstract-Reliability can be considered as one of the crucial factors in the assessment process. However, with regard to peer assessment, reliability has become controversial due to many reasons. Hence, this study aims to investigate the extent to which peer assessment is reliable in ESL context. Irrespective of the fact that peer assessment is popular and wide spread, in the context the study has been carried out, peer assessment is still a new concept which needs more exploration and practice. The sample of this study consisted of 45 science students who were studying technology in a higher educational institute. The students received peer feedback for the oral presentations they made. The feedback forms of the students were the research instruments of this study. Here, the attention was paid on the inter-rater reliability of the students. According to the findings, six, out of nine groups were identified to have fair to moderate reliability with regard to rating of their peer groups.

Key words: Peer assessment, reliability, ESL, feedback forms

I. INTRODUCTION

Since learners-centered education has emerged with the new educational reforms, there was a need of new method of assessment. In such circumstances, peer assessment has become popular in many educational contexts. Karami & Rezaei (2015) point out that "Peer assessment is considered to be one of the main forms of alternative assessment. The importance of peer assessment is highlighted in different educational learning and educational research" (p. 94). On the other hand, the students are expected to bear more responsibility in learning and it has been identified that peer assessment is able to enhance their critical thinking and they become more aware about the assessment criteria.

Students learn to become assessors through peer assessment. This is leaning on a meta-level beyond the immediate leaning gains from receiving feedback and assessing someone else's work....They will benefit from knowing how to define criteria and to determine themselves whether or not they meet these (Gielen et al. 2011 p.729).

Similarly, the students are engaged in the assessment process to involve them with giving and receiving feedback which is important not only for their academic life but also for the

carrier life. Black et al. (2003) highlight that "the ultimate aim of peer (and self) assessment is not that students can give each other levels and grades...the real purpose-the identification of learning needs and the means of improvement" (as cited in White, p.3). Hence, it is obvious that peer assessment contributes to the teaching learning process making learners autonomous. Due to this reason, peer assessment is popular in higher education sector.

However, as in any assessment tool, the reliability of peer assessment has become imperative. Many studies have been carried out on reliability of peer assessment but it is apparent that the word reliability was confused in many studies. At the same time, it is obvious that reliability of peer assessment incorporated with oral presentations need more exploration. In the study on reliability of teachers and peer assessment, Magin (2010) points out that the findings "...lend support to the common-held belief that students are quite poor at judging oral presentations skills"(p.295). Presentation skills are essential for today learners and those skills cannot be acquired within one night. Therefore, learners should bear more responsibility on enhancing their presentation skills by understanding the learning needs.

II. LITERATURE REVIEW

Peer assessment has been studied in many educational sectors. According to Topping (2009) "Peer assessment is an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" (p.20-21). Unlike in traditional assessment, here the students are expected to assess their peers. At the same time, peer assessment improves the learner involvement. Puegphrom et al (2011) specify that peer assessment "is an alternative of assessment process that involves the learner's participation" (p.2). However, reliability of peer assessment has become problematic. Here, the maintaining reliability is crucial as the assessors in this assessment process are not teachers but peers who have less experience in assessing. Many studies have been carried out on reliability of peer assessment and it is imperative to identify what reliability is. In fact, in teaching learning context, traditionally teachers' reliability has been studied and maintained well. The reliability of marking can be achieved when all the markers



have given consistent marks for a particular performance of a learner or learners. According to Aryadoust (2016), “If student raters marked the performance of another sample of student presenters drawn from the same population, they would achieve high consistency in rating” (p.19). In fact, the reliability in marking will appear when the assessor and the assessee are in the same population.

However, with reference to literature, there are many studies on peer assessment based on reliability. Cho (2006) points out that “Regarding the reliability of the peer-generated grades, students focus on the distribution of grades they receive: The greater the spread of grades, the less reliable is the grading” (p.893). He further says that the “actual problem of reliability i.e. did the students agree with one another” and at the same time, he points out some previous studies in which validity and reliability have been misunderstood (p.892). It is observed that in some studies, validity and the reliability have been misinterpreted. According to, Falchikov and Goldfinch (2000) “validity in assessment is the ‘degree of agreement between [the teachers’] marks and those awarded by their students’, whereas reliability is about agreement in ratings between and among peers”. Hence, considering the above statements, it can be concluded that reliability means the agreement of peer ratings.

Meanwhile, considering the validity and the reliability, Topping (1998) pointed out the difference between validity and reliability and mentioned that there is a gap in literature since these two areas have not been explored adequately. “Many purported studies of ‘reliability’ appear actually to be studies of validity. That is, they compare peer assessments with assessments made by professionals rather than with those of other peers or the same peers over time” (p.257). This statement, itself suggests the need for further studies on reliability. The purpose of the present study is to fill the gap existing in the literature. Hence, this study attempts to investigate to what extent is peer assessment reliable in ESL context.

III. METHODOLOGY

The sample

The sample of this study consisted of 45 students who were studying in a higher educational Institute. During their Advanced Levels, they have studied in science stream. The age group of the participants is 21-23 and all of them belonged to the lower middle socio-economic background. At the same time, they are a mixed ability group. To maintain the homogeneity of the sample, only the students who have obtained a “C” pass (Credit pass) for “General English” offered in G.C.E. A/L examination (General Certificate of Education Advanced Level) at schools, were included in this sample. It was observed that the majority of the participants were female students (based on the data obtained from the bio-data questionnaire).

Peer Feedback Forms

The feedback form which consisted of a five-point Likert scale was adopted from White (2009) and Yamashiro & Johnson (1997) to assess presentation skills of adult learners. It contained four major sections as voice, body language, content of the presentation and effectiveness and with 14 sub sections. The language of the feedback form was English.

The procedure

The study was carried out for a period of 6 weeks. During this period they were exposed a session on peer assessment and how to provide peer feedback. In the next stage, students were grouped as five members per a group. Then the students were supposed to make two group presentations to which they receive peer feedback respectively.

Analysis of data

The data collected from the feedback forms given to the peer assessment groups were analyzed using One-way Analysis of variance (ANOVA) to measure the inter-rater reliability of using peer assessment in the assessment process as ‘decisions are made by comparing averages which come from different group of raters’ (Ebel, 1951,p.412)

IV. RESULTS

In order to find out the reliability of the peer assessment the responses given by the peer groups on the peer feedback forms were used. First the inter-rater reliability of the responses given for the Group Presentation 1 was measured, secondly, the reliability of the responses given for the Group Presentation 2 was measured and finally the inter-rater reliability of the responses given for both Group Presentation 1 and 2 was measured.

Here, the statistical tool one-way analysis of variance (ANOVA) was used as it determined whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. This analysis compares the means between the groups which are involved and determines whether any of those means are statistically significantly different from each other.

Inter rater reliability is the estimation based on the correlation of scores between/among two or more raters who rate the same item, scale, or instrument i.e. the level of agreement between raters or judges. The reliability of the responses of this study was discussed according to the Landis and Koch-Kappa’s Benchmark Scale.

Group Presentation 1

The Table 4.1 depicts the inter-rater reliability of the feedback received for the Group presentation 1. According to the above mentioned table, the reliability of peer groups 4, 5 and 7 is poor as their inter rater index (designated r_{nn}) is < 0.0 while the peer groups 1 and 2 have slight reliability with regards to



the peer responses they have given and their r_{nn} is between 0.0 and 0.20 . The peer groups 8 and 9 which have gained $0.21 < r_{nn} < 0.40$ have fair reliability. Moderate reliability is seen in groups 3 and 6 as they have $0.41 < r_{nn} > 0.60$.

Rater	Inter rater Index	Mean value of given marks	N	Benchmark Scale
1 Group	0.002997951	2.26546479 865429	8	slight
2 Group	0.06845815	1.28998982 721741	8	slight
3 Group	0.54148868	2.06289875 778577	8	moderate
4 Group	-0.48255984	2.287959	8	poor
5 Group	-0.65644172	2.52848427 949434	8	poor
6 Group	0.416485246 1	2.31124846 185191	8	moderate
7 Group	-0.162780	2.35554588 888852	8	poor
8 Group	0.25492628	3.62699523 26462	8	fair
9 Group	0.255	2.58963254 88625	8	fair

Table 4.1: Inter-rater Reliability of Group Presentation 1

It is significant to note that though this is their first peer assessment session, six groups are reliable in assessing their peers. Therefore, it is evident that utilizing peer assessment in this context is 66.67% successful with regard to reliability.

Group Presentation 2

The inter-rater reliability of the feedback received for the Group presentation 2 is shown in table 4.2. According to the table, in Group presentation 2, the same inter-rater reliability is seen as in the Group presentation 1, in all the groups except Group 3. In Group 3, the reliability has reduced from moderate (Group presentation 1) to fair (Group presentation 2). This is an unexpected result considered to the Group presentation 1.

Rater	Inter rater Index	Mean value of given marks	N	Benchmark Scale
1 Group	0.002997951	2.265464798 65429	8	slight
2 Group	0.005879478 1	2.336949771 62	8	slight
3 Group	0.326146872	2.288998745 1612	8	fair
4 Group	-0.42784652	2.355862	8	poor
5 Group	-0.65644172	2.528484279 49434	8	poor
6 Group	0.416485246 1	2.311248461 85191	8	moderate
7 Group	-0.29317577	2.032548695 254	8	poor
8 Group	0.25492628	3.626995232 6462	8	fair
9 Group	0.255	2.589632548 8625	8	fair

Table 4.2: Inter-rater Reliability of Group Presentation 2

Group Presentation 1 and 2

Finally, the inter rater reliability of both group presentation was measured. However, when the inter-rater reliability of both presentations is considered, the reliability has increased in the three groups. This is significant in this study with regard to the reliability of peer assessment. According to the table 4.3, the reliability has increased from moderate (Group presentation 1) and fair (Group presentation 2) to Substantial in group 3, as $0.61 < r_{nn} > 0.80$.

At this point it is imperative to note that the peer group 3, which shows a decrease of inter-rater reliability in group presentation 2 compared with group presentation 1, has shown a “substantial” reliability when considering the reliability of both group presentations. The reliability of group 9 has increased from fair to Substantial as $0.61 < r_{nn} > 0.80$. The outstanding i.e. “almost perfect” reliability is apparent in group 6 as in group 6 inter rater reliability is $0.61 < r_{nn} > 0.80$.



Rater	Inter rater Index	Mean value of given marks	N	Benchmark scale
1 Group	0.0218 54305	3.298076923 07692	16	slight
2 Group	0.0804 52546	2.889423076 92308	16	slight
3 Group	0.7652 77343	3.158653846 15385	16	Substantial
4 Group	- 0.7033 05456	3.0625	16	poor
5 Group	- 0.9700 5988	3.221153846 15385	16	poor
6 Group	0.9322 69874	3.288461538 46154	16	Almost perfect
7 Group	- 0.3348 83721	3.240384615 38462	16	poor
8 Group	0.3237 92487	3.105769230 76923	16	fair
9 Group	0.608	3.475961538 46154	16	substantial

Table 4.3: Inter-rater Reliability of Group Presentation 1 & 2

V. DISCUSSION

Considering the reliability of peer assessment on presentation skills, it was identified that six out of nine groups have shown considerable to moderate reliability and the study showed that there were differences between different groups in how the marking was done even with the same criteria sheet. It appeared that peer assessment had not enabled all the groups in the study to make judgments about the performance of others in a consistent manner. The reliability of the responses of this study was identified as lower than that of the findings of Magin and Helmore (2001) who have reported a reliability of 0.79-0.84 based on the study carried out for four years with peer and teacher assessment around 100 student presentations in each year.

Similarly, based on a study carried out with staff and students, Liow (2008) has reported a lower r value of students' marks and has indicated that students have given a higher score (5-10%) than the staff. Further, literature presents evidence of a study at Hong Kong Polytechnic University in which Cheng and Warren (2003) reported that their students were having misgivings about awarding grades to peers with some grades which were "unfair and risky" (p.268) Similarly, in the study of Peng (2009), she drew the attention to the effect of group competition which has caused the under-marking in the group to group peer assessment. Therefore, these findings prove that the irrespective of the context, the reliability with reference to peer assessment is still a matter consider. However, in the study of Kappe (2008), he had found that students were able to provide a reliable overall assessment but he added that reliability could be gained by additional training on specific criteria of oral presentations.

VI. CONCLUSION

There is no doubt that the peer assessment will dominate the assessment system in the higher education sector. Hence, it is imperative to consider the reliability of peer feedback. The awareness and usage of peer assessment is rare in the Sri Lankan context and only a few educators have attempted to practice it in the real scenario. Hence, the focus of this study was to examine the reliability of the peer feedback on presentation skills. The students in this study neither had experience with peer assessment nor heard of the peer assessment process. Yet, five groups were able to achieve considerable reliability in their grading. The students in the sample belong to the science stream and the results of this study may not be on par with another sample that belongs to other streams like Arts, Commerce etc. However, irrespective of its limitations, it is expected that the important pedagogical implications of this study will contribute to the education system in Sri Lanka and stimulate the researchers and practitioners to research and utilize peer assessment. On the other hand the result of this study could be enhanced if these participants had experience and practice in peer assessment. Hence, in future studies, it is vital to provide further training for the participants.

VII. ACKNOWLEDGEMENT

The author acknowledges the students, the director and the family for their continuous support, and encouragement to improve the assessment methods in this context.

VIII. REFERENCES

- [1] Aryadoust, V. (2016) Gender and Academic Major Bias in Peer Assessment of Oral Presentations, *Language Assembly Quarterly*, 13:1, 1-24
- [2] Cheng,W & Warren, M (2003) Peer assessment of language proficiency, *Language Testing*, 22:1,93-121, DOI: 10.1191/0265532205lt298oa .



[3] Cho, K., Schunn, C.D. and Wilson, R.W. (2006) Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives, *Journal of Educational Psychology*, 98: 4, 891–901, DOI: 10.1037/0022-0663.98.4.891

[4] Falchikov, N. & Goldfinch, J (2000) Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks, *Review of Educational Research*, 70: 3, 287-322, Retrieved from <http://www.jstor.org/stable/1170785>

[5] Gielen, S., Tops, L., Dochy, F., Onghena, P. and Smeets, S. (2010), A comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum, *British Educational Research Journal*, 36: 1, 143-162, Retrieved from <http://www.jstor.org/stable/27823591>

[6] Karami, A. & Rezaei, A (2015) An Overview of Peer-Assessment: The Benefits and Importance, *Journal for the Study of English Linguistics*, 3:1, 93-99, doi: 10.5296/jsel.v3i1.7889

[7] Liow, J. (2008) Peer assessment in thesis oral presentation, *European Journal of Engineering Education*, 33:5-6, 525-537

[8] Magin, D. & Helmore, P. (2001) Peer and Teacher Assessment of Oral Presentation Skills: How reliable are they?, *Studies in Higher Education*, 26: 3, 287-298, doi: 10.1080/03075070120076264

[9] Peng, J.F. (2009) Peer Assessment of Oral Presentation in an EFL context, UMI Dissertation Publishing, Ann Arbor

[10] Puegphrom, P. & Chiramanee, T. (2011) The Effectiveness of Implementing Peer Assessment on Students' Writing Proficiency, The 3rd International Conference on Humanities and Social Sciences, 1-17

[11] Topping, K.J. (1998), Peer Assessment between Students in Colleges and Universities, *Review of Educational Research*, 68 : 3 , 249-276 , Retrieved from <http://www.jstor.org/stable/1170598>

[12] Topping, K.J. (2009) Peer Assessment, *Theory into Practice*, 48:1, 20-27, doi: 10.1080/00405840802577569

[13] White, E. (2009) Student Perspectives of Peer Assessment for Learning in a Public Speaking Course, *Asian EFL Journal*, 33, 1-55, https://www.asian-efl-journal.com/pta_January_09.pdf