



The Effect of Peer Assessment on Oral Presentation in an EFL Context

SaedeH Ahangari (Ph.D.)

Department of English Language, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

Behzad Rassekh-Alqol (Corresponding Author)

Department of English Language, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

Tel: +98-914-867-9761 E-mail: Behzad_rassekh@yahoo.com

Leila Ali Akbari Hamed

Member of Scientific Association of English, Tabriz branch, Islamic Azad University, Tabriz, Iran

Received: 11-02-2013

Accepted: 16-03-2013

Published: 01-05-2013

doi:10.7575/aiac.ijalel.v.2n.3p.45

URL: <http://dx.doi.org/10.7575/aiac.ijalel.v.2n.3p.45>

Abstract

The purpose of this study was to examine the effect of peer assessment on oral presentation of Iranian English non-major students. To this end, 52 students of Azad and State universities of Tabriz were selected as participants. Their ages ranged from 25 to 38. The peer assessment was incorporated into the experimental group's course to explore whether and to what extent their oral presentation skills may enhance. Data were obtained through a Likert scale questionnaire of peer assessment. Data were analyzed using SPSS Software. The results specified a statistically significant difference among the groups. The findings of the present study also suggest that, when assessment criteria are definitely established, peer assessment empowers students to evaluate the performance of their peers in a manner comparable to those of the teachers.

Keywords: Peer assessment, oral presentation, ELT, feedback, EFL

1. Introduction

An essential feature of education is assessment and the significance and popularity of student-oriented learning demand alternative techniques of assessment to evaluate teaching and learning. A large number of novel approaches of assessment have hence been suggested which meant to develop the integration of learning and assessment by enhancing the engagement of students in the assessment tasks (Slujsmans, Brand-Gruwel, vanMerrienboer, & Bastiens, 2003). According to these approaches, learners are regarded as active agents who have equal responsibilities, collaborate, think, and carry out a continuous conversation with the teacher and/or peers. Furthermore, assessment does not totally grade students' activities, rather it is applied to check out learners' achievement and promote their learning tasks by involving them in the assessment tasks. Learner engagement in the process of assessment has been enormously rising in tertiary educational settings in various fields such as medicine, science, engineering, business and writing (Falchikov, 1995a; Freeman, 1995). Not long ago, studies have proposed the use of assessment as a learning method (Purchase, 2000; Rust, Price, & O'Donovan, 2003; Smyth, 2004).

Few studies of peer assessment appear to have involved EFL students being required to evaluate their peers' proficiency of spoken English, and to judge peers' assessment of oral production contrasting nonlinguistic criteria in both written and oral tasks. The researchers of this study thought that in an EFL context, the inclusion of peer assessment of oral presentation skills would benefit students by enhancing their understanding of the type and diversity of errors made and this would also serve to indicate the potential obstacles to communicate effectively which are posed by these errors. If students could be successfully trained to judge their peers' oral presentation, this would have a positive impact on self-correction. The present study sought to add to the growing body of knowledge relating to the application of peer assessment in EFL contexts.

2. Review of Literature

A technical shift which has recently taken place in tertiary education has been from a teacher-centered perspective to a perspective in which student engagement is more central (Boud, 1995). Peer assessment is the process in which students assess their peer(s) (Falchikov, 1995b; Freeman, 1995). Eisner (in Boud, 1995) acknowledged the characteristics of the new assessment in education as follows:

- Assessment tasks need to reveal the tasks that learners will face in the settings outside schools, not just those restricted to the schools themselves;
- Assessment tasks should not be constrained to the solutions that students articulate, but also demonstrate how students perform solving a problem;

- Assessment tasks should mirror the values of the intellectual society from which tasks are originated;
- Assessment tasks need not be confined to a single performance;
- Assessment tasks should have more than one conventional solution to a problem and more than one acceptable response to a request;
- Assessment tasks should have curricular application, but not be limited to the curriculum as taught;
- Assessment tasks should allow the learner to select a form of picture that he or she chooses to display what has been learned (pp.93-95).

Somervell (1993) found that peer assessment involves learners in making evaluations about other students' work or the performance. Whereas many similarities exist in the oral assessment activities which are described in literature, Joughin (1998) witnessed three specific sources of variation; the exclusion or inclusion of a written paper integrated with the oral presentation to constitute the overall score, individual or group presentations, and the person/s evaluating the assessments. In addition, he also recognized four particular advantages of oral assessment: developing good learning; authenticity; promoting student strengths; and fighting plagiarism.

Sambell & McDowell (1998), for instance, examined the students' perceptions towards various aspects of novel assessment. Many students who were discussing innovative assessment believed that success is, more or less, affected by continuous application and hard work, and this achievement did not depend on sheer luck or a last-minute burst of effort. Most learners thought that clarity and honesty were essential requirements of a valid and fair assessment approach. Students were optimistic about the impacts of alternative assessment on their learning.

The main advantages of peer assessment are that it develops evaluative and critical abilities, offers opportunities for skill development, integrates knowledge, and motivates student collaboration. Sometimes, there are weaker points in organizational matters, such as the cheating, time constraints and stress. Finally, there could be a misfit between marks without feedback and learning (McDowell, 1995).

Boud (1992, 1995) presented an assessment plan to provide an analytical and comprehensive record of learning in settings where learners have heavy responsibility for their performance. The major instruction is a handout which provides the students with the headings. As effective tools, Assessment schedules are used to enable learners to reflect on their success, to arrange a wide variety of their learning, and to investigate the practical implications for further learning (Boud, 1992; Boud & Knights, 1994). It is suggested that getting to know the criteria of a rubric and observing the peers' practice and work, results in a profound understanding of the quality of one's own performance (Falchikov, 1995a; Freeman, 1995).

Self and peer assessment training might have noticeable influences on the development of the skills concerning the content, if the training is integrated in the current syllabus (Mehrens, Popham, & Ryan, 1998). From this perspective, the assessment skill is not drilled as a skill in isolation, but is closely connected to course content. Learners will acquire how to negotiate about criteria for an acceptable presentation if the training of defining criteria is embedded in a teacher's course on presentation skills. Understanding these criteria paves the way for the students to promote their own presentations, therefore, students' development of their presentation skills will be supported by the assessment training.

Cheng & Warren (2005) have found that students adopt highly favorable attitudes towards peer assessment activities, and that such task can develop self-confidence and intrinsic motivation in the learner (Falchikov & Goldfinch, 2000). Consequently, students actually get pleasure from judging each other's practice, and derive substantial benefit from conducting this task. As a matter of fact, Topping (1998) claims that students assess their peers in a manner which is almost comparable to their instructors, which implies it is helpful for teachers to deeply think about applying peer assessment activities in classrooms. Yet, learners must be provided with the correct instruments to do the task, and need to be trained appropriately before carrying out any formal peer evaluation activities (Cheng & Warren, 2005).

A few studies have been conducted in which learners assessed their peers in EFL contexts:

In a research by Jafarpur (1991), Iranian students evaluated their peers' listening and speaking skills. He found low agreement between peer and teacher- assessment. Moreover, he noted that it was more difficult for less able students to self- or peer-assess when compared to more able ones. He focused on learner training and guidance in interpreting the scoring criteria. He maintained that if the EFL learners are assumed to shoulder the burden of responsibility in study programs, this will require some modifications in testing procedures.

Miller and Ng (1994) led a study that engaged learners evaluating and allocating scores to their peers' oral performance along with expressing their perceptions towards participating in peer assessment tasks. The findings demonstrated that students were able to actually judge their group members' oral language skill under definite conditions, specifically if the group was homogeneous, the student assessors were highly proficient language learners, the students were engaged in the preparation of the test, they had former experience with each other's oral language skill, and the testing environment was friendly (Miller and Ng, 1994). The outcomes also illustrated a comparatively high level of agreement between teacher- and peer assessments. On the subject of perceptions, the researchers observed that their learners, all together, took a hostile attitude towards peer assessment, citing causes such as the eccentricity of the task, the partiality of the task, the inequity of the whole practice, embarrassment in front of classmates, and feelings of being inexpert, unqualified and not proficient enough in English to evaluate their peers' oral production.

Forde's (1996) research investigated the stability with which Chinese speakers of English in Hong Kong, signed up in a training course for IELTS, and graded the oral skills of their peers. The study demonstrated that the majority of the students were able to perfectly judge their peers' oral proficiency, when compared with the evaluations given by the IELTS examiner.

Yamashiro (1999) carried out a detailed exploration of promoting public speaking skill among Japanese university students. He developed a validation scale and the validity of this rating scale based on peer, teacher, and self-rating was confirmed.

Patri (2002) studied the conformity between teacher-, self- and peer assessment of students' oral skills with and without the use of peer feedback. He noted that there was a higher correlation between teacher and peer assessments when the procedure was boosted by peer feedback, but most of proficient students were not able to assess each other's performances in the same way as the teacher had done, irrespective of whether or not peer feedback was integrated into the practice.

Cheng and Warren (2005) examined the reliability and potential advantages of integrating peer assessment into English language programs. They compared the learners' perceptions towards evaluating both the English language proficiency and other features of their peers' performance. In addition, they made a comparison between peer and teacher assessments. The results proposed that learners adopted a less optimistic attitude towards judging their peers' language proficiency, but they did not rate their peers' language proficiency very differently from the other rating criteria.

Saito (2008) investigated the influences of training on peer assessment and comments received concerning oral productions in EFL classrooms in Japan. Analyses of comments indicated that the treatment groups outperformed in both quantity and quality of comments. Accompanied by a meta-analytic summary, it was maintained that peer assessment was a vigorous method wherein training on skill features might be sufficient to attain a definite level of correspondence with the criterion variable (teacher), but training may increase student comments and decrease mismatching ratings.

The above-cited studies reveal that careful planning and good management as well as learner involvement in language programs can produce good results. The question now remains unanswered as to why it is not a more prevalent aspect in all EFL teaching programs. Part of the issue may address teacher indecision as to how learners will respond: will it be regarded as handing over the undertaking or as a positive step in learner engagement? In addition, as it was previously mentioned, it is comparatively easy to accomplish student involvement as a stage of a writing course, but what about speaking and listening?

This study set out to deal with some of the above mentioned areas of concern. In this study, students in a speaking skills proficiency course were given the opportunity to assess each other's oral presentation skills.

2.1. Research Questions and Null Hypotheses

This study focused on how peer assessment could improve the oral performance of the learners in an EFL context. In attempting to answer this general question, this study specifically explored the following questions:

1. Does peer assessment improve the oral presentation skills of EFL learners?
2. Can peer assessment enable students to judge the performance of their peers in a manner comparable to those of the teachers?

To probe the above questions, the following null hypotheses were formulated:

H01: Peer assessment does not improve the oral presentation skills of EFL learners.

H02: Peer assessment cannot enable students to judge the performance of their peers in a manner comparable to those of the teachers.

3. Method

3.1 Participants

The participants in the study were 52 (18 male and 34 female) university students from the State and Azad universities in Eastern Azerbaijan, Tabriz. They were graduates of Science (41), Engineering (5), Accounting (2), and Physical Education (4). 36 participants had Masters Degrees and 16 were PhD students, and their ages ranged from 25 to 38. These students were selected according to their performances on the TOEFL (Sharpe, 2001). Out of a total of 70 students, 52 obtained 400 – 450 on the TOEFL (which roughly corresponds to Intermediate High on ACTFL scale) which are typical scores for English non-major students in Iran. One of the groups was randomly selected as the experimental group and the other one as the control group. An independent t-test on the mean scores of the two groups on the TOEFL confirmed the homogeneity of the groups.

3.2 Instrumentation

The instrumentation used to collect data included:

- (1) In order to sample the subjects, the TOEFL test (Sharpe, 2001) was used for subject selection.
- (2) The topics for presentation were chosen from Barron's TOEFL Handbook (Lougheed, 2004). The treatment was built into the researchers' syllabus.

(3) The instrumentation also included a peer assessment questionnaire (Appendix 1) adopted from Patri (2002). The reliability of the instrument was reported to be .70 (Saito, 2008). Fourteen questions were included in the questionnaire which was split into four categories. Category 1 comprised questions 1–6 that covered organization and content of presentation. Category 2 (questions 7–9) was created on the use of language. Category 3 (questions 10–12) was related to manner. Lastly, Category 4 contained questions 13 and 14 which were covering interaction with the audience.

The rating was done on the basis of a 5-point Likert scale in which the categories were: 1 – poor; 2 – unsatisfactory; 3 – satisfactory; 4 – good; 5 – excellent.

(4) The same questionnaire was used for teacher assessment with slight adjustments; changes were restricted to phrasing in the rubric.

3.3 Procedure

To evaluate the effect of peer assessment, the control and experimental groups were set up, each consisting of 26 students. An independent t-test on the TOEFL (Sharpe, 2001) scores of the two groups confirmed the homogeneity of the groups (Table 1). Participants took part in a 28- hour course with two hours of instruction each week. The experimental group met on Sundays, while the control group's classes were held on Tuesdays. The course curriculum included making oral presentations with a focus on organization and content of the talk, language use, manner and interaction with the audience. The experiment conducted was built into the teachers' syllabus.

In week 1, the teachers held an instruction session for both groups and elaborated on the essential aspects of a good oral presentation. These characteristics (which covered the 14 items of the peer assessment questionnaire) were explained thoroughly in order to clearly establish the criteria set by the researchers. Items 8–14 – which referred to pronunciation, fluency, and speaker's eye contact, confidence, verbal and non-verbal interaction – had to be explained in more detail. Then the researchers described the characteristics of a poor, an average, and a good presentation (Freeman, 1995). Both classes were then divided into halves. Half the class was given topic (a): 'How do movies or television influence people's behavior?' and the other half was given topic (b): 'Television has destroyed communication among friends and family.' to prepare for their presentation in the following week.

In week 2, the week after the training session, participants from both the control and experimental groups made their presentations. Half the class presented topic (a), and the other half topic (b).

3.4 Control group: Weeks 2-14

The participants in the control group presented their talks. After each presentation which took about three minutes, the researchers assessed their performance according to the teacher assessment questionnaire and explained the basis for his comments. Then they drew the students' attention to the major themes of each presentation and asked them to carefully follow the rules of a good presentation. The participants were given two new topics to prepare for their presentations in the following weeks.

3.5 Experimental group: Weeks 2-6

Participants made their oral presentations on the two topics. The participants in the experimental group were divided into groups of five or six depending on the class size. Participants were assigned to different groups every week to encourage maximum interaction and to expose students to different points of view. Each group was asked to assess their peers. Group members were given an appropriate number (equal to the number of students making presentation in the class) of peer assessment questionnaires. Each student from the class made a three minute presentation on the topic assigned by the researchers. The researchers assessed the participants during the presentation using the teacher-assessment form. During the presentation, peers noted their comments and filled in the questionnaires. The researchers spent about five minutes with each group discussing the comments made by the peers. The participants compared their assessments at the end of each performance. When there were differences, the researchers explained the basis for their judgments. For example, 'this student has provided a lot of detail and the detail is appropriate and diverse. That is why I gave him a "5".' In week 2, only 8 students made their presentations and the others assessed their peers. The teacher spent more time with groups to clarify and explain the rating process item by item. In the following weeks, the number of students who managed to make their oral presentation was increasing. As the students got more familiar with the assessment criteria, the teacher spent less time to discuss the process of peer assessment. Hence, more and more students had the opportunity to be assessed by their peers under the teacher's guidance. During 5 weeks, each student did 56 peer assessments and made 3 oral presentations. At the end of each session, the researchers collected the peer assessment questionnaires. The participants were given two new topics to prepare for their presentations in the following week. Weeks 2- 6 provided the experimental group with practice in peer assessment.

3.6 Experimental group: Weeks 7-14

After the session in week 6, the researchers did not make any comments. The whole process of rating was done by the students. The participants were given two new topics to prepare for their presentations in the following week.

At the end of the course, the participants from both groups gave an oral presentation each. Their performance was assessed by the researchers.

3.7 Data analysis

Data obtained from the experimental group in week 6 were used to correlate teacher and peer assessments. Pearson correlation coefficient was calculated between teacher-peer (T-P) assessments for the mean peer rating of students with the teacher's rating for each student. Also a one-sample t-test was run on the mean scores obtained through peer assessment and teacher assessment to examine to what extent peer assessment was efficient to enable students to make sound judgments of their peers' oral performance.

At the end of the treatment, an independent t-test was run on the final scores to explore the difference between the scores of the experimental and control groups.

4. Results

In order to test the homogeneity of both groups, an independent t-test was run on their TOEFL scores. Table 1 shows that there was no significant difference ($t=.07$, $p<.05$) between control and experimental groups with respect to their TOEFL scores.

Table 1. Independent T-Test Displaying the TOEFL Scores

| | t-test for Equality of Means | | | | |
|-------|------------------------------|----|-----------------|-----------------|-----------------------|
| | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| TOEFL | .079 | 50 | .937 | .385 | 4.878 |

The data recorded in week 6 included teacher- and peer assessments (TA and PA) obtained from the experimental group. In Table 2 Pearson correlation coefficients were calculated. The high correlation between the experimental group students' peer assessments and the teacher ratings ($r=.87$) which was significant at level ($p<.01$) shows that peer assessment was in high agreement with the teacher assessment. This suggests that the students were able to make judgments of their peers' oral presentations comparable to those made by the teacher.

Table 2. Correlation between teacher and peer assessment in week 6

| | | TA | PA |
|----|---------------------|------|------|
| TA | Pearson Correlation | 1 | .874 |
| | Sig. (2-tailed) | | .000 |
| | N | 26 | 26 |
| PA | Pearson Correlation | .874 | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 26 | 26 |

A one-sample t-test was also run on the data obtained from the experimental group in week 6 to examine to what extent peer assessment was efficient to enable students to make sound judgments of their peers' oral performance. The researchers randomly selected one of the participants (whose rating on teacher assessment was 3.42) and recorded how other students assessed the same student's oral presentation. Tables 3a and 3b reassure that students are able to assess their peers reasonably.

Table 3a. Descriptive statistics for Experimental Group's Peer Assessments

| | N | Mean | Std. Deviation | Std. Error Mean |
|----|----|--------|----------------|-----------------|
| PA | 25 | 3.6464 | .50155 | .10031 |

As it is observed peer assessment mean is 3.64 which is a little higher than 3.42 assessed by the teacher. The small amount of the SD ($=.5$) shows that the assessments done by the students are more congruent, and they have grasped the main points of peer assessment.

Table 3b. One-Sample T-Test on Teacher and Peer Assessment

| | Test Value = 3.42 | | | | | |
|----|-------------------|----|-----------------|-----------------|--|-------|
| | t | df | Sig. (2-tailed) | Mean Difference | 5% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| PA | 2.257 | 24 | .033 | .22640 | .0194 | .4334 |

Table 3b shows that the difference ($p < .01$) was not significant and thus there was no difference between teacher assessment and peer assessments.

In order to find out whether the treatment had an effect on the oral performance of the experimental group, an independent t-test was carried out on the final scores of both groups. The results indicated that there was a significant difference between the ratings the students obtained at the level of $p < .05$ (see Table 4) suggesting that peer assessment had a significantly positive effect on oral presentation of students receiving the treatment.

Table 4. Independent T-Test Displaying the Results of Posttest

| | t-test for Equality of Means | | | | |
|-------|------------------------------|----|-----------------|-----------------|-----------------------|
| | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| Final | 11.542 | 50 | .000 | 1.15077 | .09970 |

5. Discussion

The results of this research demonstrate that under certain circumstances students are able to convincingly assess each other's oral language ability. As the level of our students correlate to TOFEL scores ranging from 400-450, we can say that they have an intermediate level of language ability. We can call our students 'sophisticated language learners' in that they were actively involved in pursuing their language studies further, and they demonstrated their ability to cope with a complex concept in a second language testing.

To be able to assess their peers' oral skills, students must be able to access some implicit or explicit criteria. The researchers in this study gave the students a set of assessing criteria. We believe that the students, being homogenous, have an internal set of rules leading the use - and wrong use- of English. These 'rules' are used whenever they hear English being spoken by another speaker, or when they themselves use English. Accordingly, the students were able to institute criteria of satisfactory language use. We do not know what the results would be if, say, an Iranian speaker of English were to try to assess a Spanish speaker of English. As Spaventa and Williamson (1989) point out: "...students who come from a Japanese background and students who come from a European background have different abilities in the traditional language skill areas".

In order for the learners to assess oral production fairly and reliably, they must either be aware of the assessee's oral skills (Miller and Ng, 1994) or be trained in evaluating and have experience as assessors (Patri, 2002; Saito, 2008). By conducting the oral assessments half way through the course, we gave the students enough time (6 weeks) to become accustomed to their peers' oral abilities before we asked them to assess each other. In this way, my students were better prepared for the task of judging their peers. The peers were familiar with each other and therefore the assessments were conducted under friendly circumstances. This ensured that the participants were able to perform to the best of their ability. Jones (1985, p.13) explains the importance of an examiner/ examinee relationship and concludes, "... that ease of elicitation in an oral interview is often affected by the social relationship between the examiners and the examinee. The factors can include age, race, social class, education and profession." As all our students had similar social profiles, we feel this was an important factor in their ability to assess each other fairly.

it can be concluded from the results that a) they illustrated an analogous level of assessment to that of the teacher; b) the students were open in participating in this peer assessment practice; c) the procedure did not lead to lowering the standards; and d) the students promoted their understanding of and perception towards assessment by involving in this study.

For most second language learners, speaking itself is a multifaceted (and occasionally distressing) task since it necessitates them to focus concurrently on content, diction, pronunciation, body language, eye contact, and so on. This sophistication is limited to oral tasks compared to writing or listening tasks. With low-achieving learners, this could be even more demanding. It is essential to point out here that the majority of the participants in the present study had obtained 400 - 450 on TOEFL. Previous studies on peer assessments with EFL (English as a Foreign Language) students had found that it was difficult for them to make sound evaluations of their peers' speaking and learning abilities (see Jafarpur, 1991).

The teacher and peer assessments correlations are all together substantially high. This observation lends strong support to Hypothesis 2 that peer assessment enhances learners' ability to make judgments on their peers' oral presentation skills comparable to those of the teacher.

The participants in this study demonstrated that if the criteria rules were established definitely (Patri, 2002; Saito, 2008; Yamashiro, 1999), learners could rate their peers acceptably and improve their judgments so that they could acquire a better understanding of their own skills. This explains why the experimental group outperformed the control group.

6. Conclusions and Implications

Peer assessment recently has been regularly practiced as an alternative assessment technique, predominantly in higher education. In the method of peer assessment, learners are able to witness their peers and arrive at a better understanding of how their peers learn. Sluijsmans et al. (1999) have reviewed a body of research articles concerning peer assessment and have maintained that the process of peer assessment involves students in making judgments. In addition to rating or scoring, peer assessment is also an integral segment of learning experiences. During the course of the assessment, students learn to shoulder high levels of responsibility and to concentrate on learning itself. Peer evaluation also offers the learners a situation in which they can perceive the role of their teachers and appreciate the nature of assessment. It supports students to learn about learning and, consequently, increases students' metacognitive understanding about their own learning.

This research investigated the impact of peer assessment on learners' oral presentation improvement. The results of this study indicated that when assessment criteria are clearly set, peer assessment will enable students to make judgments of their peers comparable to those of the teacher. Considering that both the control and experimental groups consisted of students with similar English proficiency – all students got 400 – 450 on TOEFL, and shared similar background – it can be assumed that peer assessment had assisted in achieving higher correlation with teacher assessments. If this, in fact, is the case, then teacher assessment could be complemented with peer assessment at a lower cost in the context of oral skills. If peers can be engaged in the task of assessment, teachers' time could be exploited more efficiently on subjects associated with promoting their teaching methods.

This research can provide comfort with language teaching practitioners to proceed with the use of peer assessment, which is formerly recommended by many researchers and is prevalent in many language teaching classes principally in the south-east Asian countries. The findings of this study may have implications for SLA theory as well as for material development and syllabus designing.

References

- Boud, D. (1992). The use of self-assessment schedules in negotiated learning. *Studies in Higher Education*, 17, 185–200.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D., & Knights, S. (1994). Designing courses to promote reflective practice. *Research and Development in Higher Education*, 16, 229–234.
- Cheng, W., & Warren, M. (1997). Having second thoughts: student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22, 233–239.
- Falchikov, N. (1995a). Peer feedback marking: development peer assessment. *Innovations in Education and Training International*, 32, 175 – 87.
- Falchikov, N. (1995b). Improving feedback to and from students. In Knight, P. (Ed.). *Assessment for Learning in Higher Education*. London: Kogan Page, 157 – 66.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.
- Forde, K. (1996). The effects of gender and proficiency on oral self- and peer-assessments. *English Language Studies Working Papers, City University of Hong Kong*, 1, 34 – 47.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education*, 20, 289 – 300.
- Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation and Research in Education*, 5, 145 – 57.
- Jones, N. (1995). Business writing, Chinese students and communicative language teaching. *TESOL Journal*, 4, 12 – 15.
- Joughin, G. (1998). Dimensions of oral assessment. *Assessment & Evaluation in Higher Education*, 23, 367-378.
- Lougheed, L. (2004). *Barron's Computer-Based TOEFL Essays*. NY: Barron's Educational Series, Inc.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32, 302–313.
- Mehrens, W.A., Popham, W.J., & Ryan, J.M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice*, 17, 18-22.

- Miller, L. and Ng, R. (1994). Peer assessment of oral language proficiency. *Perspectives: working papers of the department of English, City Polytechnic of Hong Kong*, 6, 41 – 56.
- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19, 109 – 31.
- Purchase, H. C. (2000). Learning about interface design through peer assessment. *Assessment and Evaluation in Higher Education*, 27, 341-352.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28, 147-164.
- Sambell, K., & McDowell, L. (1998). The value of self and peer assessment to the developing lifelong learner. In C. Rust (Ed.), *Improving student learning — improving students as learners*. Oxford, UK: Oxford Centre for Staff and Learning Development.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25, 553.
- Sharpe, P. J. (2001). *Barron's TOEFL Test of English as a Foreign Language (10th ed.)*. NY: Barron's Educational Series, Inc.
- Sluijsmans, D. M. A., Brand-Gruwel, S., vanMerriënboer, J. J. G., & Bastiens, T. J. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, 29, 23-42.
- Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment and Evaluation in Higher Education*, 29, 369-378.
- Spaventa, L. J. & Williamson, J. S. (1989). Participatory placement: A case study. In M. Pennington (Ed.), *Building Better English Language Programs*. NAFSA: Washington, D.C.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249 – 76.
- Underhill, N. (1987). *Testing Spoken Language*. Cambridge: Cambridge University Press.
- Yamashiro, A. (1999). Using structural equation modeling to validate a rating scale. *Presented at the 21st Language Testing Research Colloquium, Tsukuba, Japan*.

Appendix 1

Peer assessment questionnaire

Name:

Topic:

Date:

Rate your classmate by using the scale:

| | Poor | Unsatisfactory | Satisfactory | Good | Excellent |
|--|------|----------------|--------------|------|-----------|
| | 1 | 2 | 3 | 4 | 5 |
| A. Introduction | | | | | |
| 1. Topic sentence – suitable? | | | | 1 | 2 3 4 5 |
| 2. Topic sentence – fascinating? | | | | 1 | 2 3 4 5 |
| 3. The opinion on the issue – well-argued? | | | | 1 | 2 3 4 5 |
| B. Body | | | | | |
| 4. Details supporting the main points – enough? | | | | 1 | 2 3 4 5 |
| 5. Details supporting the main points – related? | | | | 1 | 2 3 4 5 |
| C. Conclusion | | | | | |
| 6. The main points – brief? | | | | 1 | 2 3 4 5 |
| D. Language Use | | | | | |
| 7. Grammar – precise? | | | | 1 | 2 3 4 5 |
| 8. Fluency | | | | 1 | 2 3 4 5 |
| 9. Pronunciation – words noticeably pronounced? | | | | 1 | 2 3 4 5 |
| E. Manner | | | | | |
| 10. Confidence (not nervous) | | | | 1 | 2 3 4 5 |

| | | | | | |
|--|---|---|---|---|---|
| 11. Confidence (looked very little at his/her notes) | 1 | 2 | 3 | 4 | 5 |
| 12. Eye contact | 1 | 2 | 3 | 4 | 5 |
| F. Interaction | | | | | |
| 13. Non-verbal interaction with the audience (facial expressions, gestures) | 1 | 2 | 3 | 4 | 5 |
| 14. Verbal interaction (involving the audience by asking questions and encouraging them to answer) | 1 | 2 | 3 | 4 | 5 |

Appendix 2

Teacher assessment questionnaire

Name:

Topic:

Date:

Rate the student by using the scale:

| Poor | Unsatisfactory | Satisfactory | Good | Excellent |
|------|----------------|--------------|------|-----------|
| 1 | 2 | 3 | 4 | 5 |

A. Introduction

| | | | | | |
|---|---|---|---|---|---|
| 1. Topic sentence – appropriate? | 1 | 2 | 3 | 4 | 5 |
| 2. Topic sentence – interesting? | 1 | 2 | 3 | 4 | 5 |
| 3. The opinion on the issue – clearly stated? | 1 | 2 | 3 | 4 | 5 |

B. Body

| | | | | | |
|---|---|---|---|---|---|
| 4. Details supporting the main points – sufficient? | 1 | 2 | 3 | 4 | 5 |
| 5. Details supporting the main points – relevant? | 1 | 2 | 3 | 4 | 5 |

C. Conclusion

| | | | | | |
|----------------------------------|---|---|---|---|---|
| 6. The main points – summarized? | 1 | 2 | 3 | 4 | 5 |
|----------------------------------|---|---|---|---|---|

D. Language Use

| | | | | | |
|--|---|---|---|---|---|
| 7. Grammar – accurate? | 1 | 2 | 3 | 4 | 5 |
| 8. Fluency | 1 | 2 | 3 | 4 | 5 |
| 9. Pronunciation – words clearly pronounced? | 1 | 2 | 3 | 4 | 5 |

E. Manner

| | | | | | |
|--|---|---|---|---|---|
| 10. Confidence (not nervous) | 1 | 2 | 3 | 4 | 5 |
| 11. Confidence (depended very little on his/her notes) | 1 | 2 | 3 | 4 | 5 |
| 12. Eye contact | 1 | 2 | 3 | 4 | 5 |

F. Interaction

| | | | | | |
|---|---|---|---|---|---|
| 13. Non-verbal interaction with the audience (facial expressions, gestures) | 1 | 2 | 3 | 4 | 5 |
| 14. Verbal interaction (involving the audience during the talk by asking questions and encouraging them to respond) | 1 | 2 | 3 | 4 | 5 |