Title
The Effect of Peer Assessment on Students' English Learning

Author(s)
大津 理香

Citation
茨城大学人文学部紀要 人文コミュニケーション学科論集, 13: 163-172

Issue Date
2012-09

URL
http://hdl.handle.net/10109/3304

Rights
このリポジトリに収録されているコンテンツの著作権は、それぞれの著作権者に帰属します。引用、転載、複製等される場合は、著作権法を遵守してください。
The Effect of Peer Assessment on Students’ English Learning

Rika Otsu

Abstract

In this study, I investigated the effects of peer assessment on students’ English language learning by examining two research questions: 1) whether the students’ understanding of the assessment criteria for a speech presentation would affect their own speech performance, and 2) what kind of reactions the students would have toward peer assessment practice sessions in small groups, followed by a fifth presentation done with the whole class. Each student assessed the other 86 students, and two teachers assessed all 87 students in the fifth presentation. Students then responded to an anonymous questionnaire about the peer assessment activity. The results gained from the fifth presentation revealed that there was no significant relationship between students’ ability to assess their peers and their own teacher-awarded scores on the presentation. At the same time, however, the questionnaire indicated that most students thought they understood the assessment criteria well. Thus, no matter how well students believed they understood the criteria, they still required additional training. Nevertheless, the questionnaire revealed the majority of students held positive attitudes toward peer assessment activities. This paper will conclude with a discussion of the implications of this study for conducting peer assessment activities in class.

I. Purpose

The purpose of this paper is to report the results of an experiment of peer assessment’s effects on students’ English language learning. Many researchers address the benefits of peer assessment; however, the relationship between students’ ability to assess their peer through peer assessment activities and their own performance is rarely reported statistically. This motivated me to research whether the students’ understanding of the assessment criteria built through peer assessment training affects their own performance. In other words, do the students who are considered as proficient assessors perform better than those who are considered as non-proficient assessors? I was also interested in finding what kind of reactions the students would have toward peer assessment practice. For example, do students perceive peer assessment as useful in motivating themselves to
learn the language? Thus, research questions were set as follows:

1) Did the students’ understanding of the criteria affect their performance?
2) What effects did students feel or confirm through peer assessment?

II. Literature Review

Peer assessment is defined as students evaluating the performance of their peers based on the criteria. This includes scoring and giving feedback on each other’s work. Such student-student interaction is considered as a valuable factor in promoting students’ learning of their L2.

Since Vigotsky introduced fundamental ideas on learning through a sociocultural point of view in 1970’s, more research on second language acquisition with an emphasis on interaction has been conducted. Vigotsky’s claim was that peer interaction can support one’s learning and exceed what can be achieved alone. His theory is that more experienced or higher level learners in some environments can contribute to less experienced or lower level learners (Vigotsky, 1978). This idea has been expanded more, for example, Wenger (1999) and Haneda (2006) acknowledged that active involvement in an environment that the students are in enhances their language learning as it gives them opportunities to reflect upon their own practices.

Indeed, many studies confirm that the benefits of peer assessment for motivational reasons derived from student-student interaction (Boud, Cohen, & Simpson, 1999; Cheng & Warren, 2005; Clifford, 1999; Eisenkopf, 2010; Smith et al, 2002; Venables & Summit, 2003). Another benefit some researchers claim is that peer assessment can maximize the students’ responsibility and let students actively participate in the learning process (Falchikov, 2007; Dochy et al, 1999). Through cooperation and social interaction, students are able to find concrete learning goals for themselves, which allows them to be autonomous learners (Beaman, 1998; Luoma, 2004).

There are two important things to consider when conducting peer assessment in the classroom. One is to set relevant criteria and the other is to let students understand the criteria well enough to assess peers confidently and precisely.

To address the first issue, as Luoma (2004) noted, “linguistic criteria may not be suitable, because students are not as adept at language analysis as teachers or raters, whereas task-related criteria may prove more effective”(p.189). One of the studies shows that even though the agreement in assessment between student and teacher was observed, and the peer assessment overall was thought to be beneficial, one third of students indeed felt unqualified and uncomfortable to evaluate their peers’ language proficiency (Cheng & Warren, 2005). According to Okuda & Otsu (2010), when incorporating peer assessment into the teacher’s final grading, the criteria which involved
linguistic rules should be carefully considered. In their study, while a high correlation between teacher scoring and peer scoring in speech assessment was confirmed in non-linguistic areas such as eye-contact ($r=.84$) and content ($r=.81$), linguistic areas such as grammatical accuracy ($r=.31$) was confirmed as weak. Thus, setting the criteria that students can handle would be key to a successful peer assessment.

To address the second issue, several researchers claim the importance of training. According to Miller and Ng (1996), a teacher-student correlation was improved when students’ listening skill was improved and they received training to assess peers. Stefani (1994) acknowledges that systematic instructions and administrating homework assignments related to peer assessment led a high correlation between teacher and student assessment ($r=.89$). Another example that should be noted here is the study conducted by Freeman (1995). He conducted four studies related to peer assessment. The first two of the peer assessment were conducted only with the criteria explanation by the teacher. Then, the correlation between teacher’s scoring and students’ scoring was weak. During the third assessment, the students and the teacher discussed and decided the criteria together before the peer assessment, however, the correlation was not yet improved. During the fourth assessment, they used a video of previous year’s presentation and discussed each speaker according to the criteria set. This training that they actually practiced with the real examples led students to be good assessors comparable to the teacher. As a similar case to Freeman (1995), Patri (2002) utilized a video of students’ presentation to learn how to assess the speaker and she claims that setting the criteria clearly, conducting training with real samples from the video as well as actually conducting peer assessment practice with real student speakers after students give feedback to each other, are all necessary steps for a successful peer assessment.

Thus, not only can peer assessment itself be a great impact on language learners, but the process in incorporating peer assessment seems to be more important. All the benefit of peer assessment written above must be derived from student to student interaction and training. Moreover, the deep understanding of the criteria gained through the student to student interaction and training must play an important role in their own language learning.

Many studies prove the benefits of peer assessment and the optimum way of incorporating it in their classroom, however, the relationship between students’ ability to assess peers and their performance improvement has been rarely reported statistically. Most studies focus on either the quality of students’ performance or producing better student assessors.
III. Methods

1) Population

For this study, 87 freshmen at Ibaraki University participated. They belonged to five different departments: Agriculture, Education, Humanities, Science, and Engineering. English is the university’s required subject and they are in the university’s English program called Integrated English Program (IEP), whose focus is on improving overall English skills. These skills are reading, listening, writing, and speaking. The program is divided into five levels [level one being the beginning level and the level five being the advanced level]. The participants were all placed in level three, which is the intermediate level.

2) Data Collection Procedure

First, the teacher explained the assessment criteria with demonstration of good and bad examples. The assessment criteria contained six categories: Voice volume, pronunciation, eye contact, fluency, grammatical accuracy, and content. Each criterion consists of scales from one to five [one being the lowest and five the highest]. In this study, the category of voice volume was excluded due to mechanical problems.

The participants were asked to make five oral presentations. The topics were taken from the textbook that they used in class, which were familiar topics such as What You Were Nervous About, and Your Future Plan. The first four presentations were done mainly for practicing assessing peers in small groups. The time limit of each speech was two minutes, followed by a one-minute assessing time. After completing these four practice sessions and exchanging their feedback with other peers, the students were asked to make the fifth presentation to the whole class. Then, all students assessed the presenter, but this time students’ peer assessment was not shown to their peers and submitted directly to the class teacher. In this final presentation, two teachers assessed the presenter for the purpose of raising the scoring reliability and of grading, but one teacher assessed by watching the video of students’ presentations that the other taped. After the final presentation, students filled out a questionnaire about the peer assessment (See Appendix 1).

3) Data Analysis

In this study, composite scores that each student assessor scored for his/her 86 peers in the fifth presentation were compared with the averaged scores that two teachers scored for the 87 student presenters by using Pearson r. Then, the students were grouped together according to the degree of the agreement. After that, each group’s averaged score that each group member earned from the teachers for his/her own speech presentation was calculated by Excel.
Other data which indicate students’ attitude toward peer assessment were gathered by an anonymous questionnaire. Students’ comments were sorted according to their positive and negative characteristics to analyze patterns and the result was represented with percentages.

IV. Results

The data responding to my first research question which asked if students’ understanding of criteria affects their performance showed that there was no significant agreement between students’ ability to assess their peers and their own scores earned by two teachers. Table 1 revealed that the students who are considered as proficient assessors \((r=0.701~0.773)\) had the lowest scores (average=16.7) from the two teachers. The students who are considered as the second most proficient assessors \((r=0.606~0.675)\) had the average score 18.2, which is higher than the most proficient assessors. The students who are considered the third most proficient assessors \((r=0.512~0.598)\) had the average score 17.8. Thus, the statistic results were not consistent and could not conclude that the assessor’s ability to assess peers like teachers affected his/her performance directly.

Table 1.

<table>
<thead>
<tr>
<th></th>
<th>CorrelWithT</th>
<th>T Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.701 ~ 0.773</td>
<td>16.7</td>
</tr>
<tr>
<td>10</td>
<td>0.606 ~ 0.675</td>
<td>18.2</td>
</tr>
<tr>
<td>14</td>
<td>0.512 ~ 0.598</td>
<td>17.8</td>
</tr>
<tr>
<td>21</td>
<td>0.404 ~ 0.497</td>
<td>18.1</td>
</tr>
<tr>
<td>13</td>
<td>0.332 ~ 0.394</td>
<td>19.5</td>
</tr>
<tr>
<td>10</td>
<td>0.200 ~ 0.295</td>
<td>18.45</td>
</tr>
<tr>
<td>6</td>
<td>0.142 ~ 0.192</td>
<td>17.25</td>
</tr>
<tr>
<td>2</td>
<td>0.074 ~ 0.098</td>
<td>17.25</td>
</tr>
<tr>
<td>2</td>
<td>-0.123 ~ 0.016</td>
<td>19.75</td>
</tr>
</tbody>
</table>

*Note.* CorrelWithT = correlation between students’ scoring and teachers’ scoring. TMean = Average teachers scores for performance by students of that assessment proficiency level.

However, the results of my second research question which asked what effects students feel or confirm through peer assessment revealed positive results from the questionnaire.

According to the results shown in Figure 1.1, although the degree of their understanding the assessment criteria differs, overall, 85% of students indicated that they understood each criterion. Figure 1.2 also shows a positive response toward assessing peers. Overall, 85% of students thought
1.1. Did you clearly understand each criterion?

1.2. Was assessing others helpful for your own speech?

1.3. Did the assessment you made tend to be lenient because it was shown to your peers?
   (in group presentation)

1.4. Was it easier to assess your peers because your assessment wasn’t shown to them?
   (in class presentation)

1.5. Did you tend to assess more leniently than in the case of group speech assessment?

Figure 1. Analysis of students’ responses to the questionnaire questions
   (1= strongly agree, 6= strongly disagree)
assessing others was helpful for their own speech. To support this data, the comments students made for peer assessment indicates that 78% of the students made positive comments, while 22% of the students made relatively negative comments. All the positive comments have some prominent patterns: students thought that 1) assessing other students gave them the idea of what good speech was like, 2) being assessed by other students allowed them to learn the bad points of their own speech and they felt they were able to improve their final speech, 3) peer assessment training was beneficial in getting students used to making a speech in front of others, and 4) peer assessment was new and simply fun. Another prominent comment was that presentation for the whole class was more dynamic and tense, which inspired the students to concentrate on what they were doing.

Mainly two patterns were found in students’ negative comments: students thought that 1) it is difficult to assess their peers because they are friends, and 2) peer assessment made them feel nervous.

Figure 1.3, 1.4, and 1.5 shows the result of whether students felt comfortable in showing their assessment to their peers. Figure 1.3 revealed that overall, 52% felt no problem with showing their assessment to their peers in their group session.

Contradictorily, Figure 1.4 indicated that overall, 78% of the students felt that it was easier to assess their peers in the final presentation for the whole class because their assessment was not shown to their peers.

Figure 1.5 indicated that overall, 88% of students assessed their peers without being lenient even though they had to show their peers the scores they marked. But the response for the question which asked whether students feel a difference between the first four group peer assessment sessions and the final peer assessment indicated the opposite attitude of the students. 73% of the students commented that they found a difference, while 27% of the students answered that they did not feel the difference between the two. A majority of the comments at the students who found a difference made claims that the final speech assessment was easier to grade because they did not need to show the score to their peers and they scored more strictly than in the group presentation.

V. Discussion

Considering the results revealed through the two research questions, even though students’ understanding of the criteria and ability to assess their peers did not affect their speech performance statistically, the anonymously conducted questionnaire that caught students’ raw voice supported Vigotsky’s as well as Wenger’s and Haneda’s theories. Interesting was that students perceived and remembered especially other peers’ “good points,” and when being assessed, they were glad to see
their bad points so that they could improve. This subtle difference indicates that students in this context tend to view themselves naturally to some degree as immature learners compared to their peers. This seems to agree with the study of Cheng & Warren (2005) in the point that students felt unqualified to judge their peers; however, the difference is that the motivation that the students had was quite high in this study. The environment created by this peer assessment was new, dynamic, and tense as some students commented. Other peers’ good points in their speech may have caused the students to want to improve their bad points. Also, the responsibility as an assessor may have led the students to be an active participant.

Another thing that should be discussed here is that friendship bias among the students definitely influenced their scoring. The result on whether showing the assessment to peers affected their scoring was not consistent at all in Figure 1.3, 1.4, and 1.5, and many students commented that they were more nervous to assess others in group presentations because they had to show their assessment to others.

VI. Conclusion

This study was able to successfully process peer assessment in class using two important steps mentioned earlier: 1) setting the clear assessment criteria, and 2) letting students understand the criteria through training. In fact, Table 1 shows that 62% of the students became proficient or close-to-proficient assessors. Also, peer assessment activities were confirmed as positive classroom activities in supporting students’ learning motivation. But the study also revealed that there were other factors which would make students’ performance better. For example, students might need more time to practice their speech after noticing their weakness. This study was conducted only during a semester and it was challenging to observe students’ speech improvement in such a short period. Also, the limitation of my data analysis might have prevented more accurate and detailed results. This study analyzed only composite scores. Therefore, further investigation of student speech scores in each category is necessary.

Acknowledgements

I would like to express my appreciation to Rieko Okuda, Cheryl Boyd Zimmerman, and Nathan Carr for their comments and support.
References


Appendix 1 Speech Assessment Questionnaire

Please answer the following questions. We will be using the information to improve our future classes.

1. Did you clearly understand each criterion?
Not at all 1 2 3 4 5 6 Yes, definitely

2. Was assessing others helpful for your own speech?
Not at all 1 2 3 4 5 6 Yes, definitely

3. Did the assessment you made tend to be lenient because it was shown to your peers?
   (in group presentation)
Not at all 1 2 3 4 5 6 Yes, definitely

4. Was it easier to assess your peers because your assessment wasn’t shown to them?
   (in class presentation)
Not at all 1 2 3 4 5 6 Yes, definitely

5. Did you tend to assess more leniently than in the case of group speech assessment?
Not at all 1 2 3 4 5 6 Yes, definitely

Please feel free to make comments.

We have done 5 speeches so far. What do you think about the peer speech assessment?

Did you notice any differences between group peer assessment and that of the class assessment?