Peer Assessment of Oral English Proficiency

Pilantana Yinjaroen ¹, Thanyapa Chiramanee ²

1. M.A. in Teaching English as an International Language, Department of Languages and Linguistics, Faculty of Liberal Arts, Prince of Songkla University.
2. Assoc. Prof. Thanyapa Chiramanee Ph.D, Department of Languages and Linguistics, Faculty of Liberal Arts, Prince of Songkla University

Abstract

This study aims to investigate the effectiveness of peer assessment done by students with different levels of language proficiency. Peer assessment data were collected while students were performing their final role play at the end of the first semester, 2010. The subjects were 51 students of Prince of Songkla University faculty of Liberal Arts. The subjects were divided into three groups based on their English scores. The descriptive statistics, ANOVA and Pearson Product Moment were used to analyze the results. The results of the study indicate that students tend to overestimate their peers, and students’ peer assessment by three proficiency groups was comparable to that of their teacher’s only in terms of voice & pronunciation.

Key words: Peer assessment, Teacher assessment
1. Background

English seems to be the most popular language that people around the world use to communicate to each other. English is used as a primary language in international business, economics, science, aviation, technology, and tourism (Kitao, 1996; Al-Lssa, 2006). Many countries use English as an official language and some use English as a second or foreign language. The number of people using English has increased; 329 million people use English as the first language, 422 million people as a second language, and another 100 million people as a foreign language (e.g. in Thailand, China and Japan) (Crystal, 2003 as cited in Narongraj, 2008).

In Thailand, although Thai people do not use English as an official language, it has been taught in every program of studies from kindergarten to university level. English is a compulsory foreign language subject starting from level 1 in primary education (6 years of ages). It is one of the eight compulsory strands that students will have to take in the core and elective courses (Wiriyachitra, 2002).

However, the effectiveness of English teaching is one of the most important issues receiving great attention from Thai educators. In addition to the teaching of English, evaluation is another important issue in language teaching that is widely discussed. In the past two decades, alternative assessments have been developed in the classroom such as self assessment and peer assessment (Brown and Hudson, 1998). Self-assessment refers to the involvement of learners in making judgments about their own learning, especially about their achievements in language learning (Alderson, 1985). Since learners’ involvement in their own learning has attracted educators’ attention, peer assessment is a topic of great interest. Topping (1998) defines peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality of success of the products or outcomes of learning of peer similar status” (p.250).

Peer assessment has been shown by various studies to have many benefits. Alderson, (1985), for example, states that during the process of peer assessment, the students might increase their awareness of the criteria by which themselves are assessed. Peer assessment may help them in developing meta linguistic skills which are related to the ability to self-evaluate and critically monitor themselves. Jafarpur, (1991) indicates that peer assessment can increase responsibility of the learners in EFL study programs. McLanghlin & Simpson (2004) find that peer assessment affects students’ attitudes. Students enjoy doing the task and don’t feel uncomfortable in assessing their peers.
In Thailand, it would be of interest to investigate the effectiveness of peer assessment in language learning of Thai students. Peer assessment always involves productive skills like speaking and writing. In speaking, presentation for monologue and role-play for dialogue are commonly used for peer assessment. For this study, role-play is used as an instrument for data collection.

2. Research Question

Can students with different language proficiency assess their peer oral performance compared to that of their teacher?

3. Technical Terms

3.1 Peer assessment is defined as the process whereby groups of individuals rate their peers (Falchikov, 1995).

3.2 Role Play is defined as a classroom activity that provides students the opportunity to practice language through the actual roles they may need outside the classroom (Freeman, 1986).

3.3 English proficiency refers students’ English O-NET scores from National Institute of Educational Testing Service Public Organization.

4. Literature Review

4.1 Peer assessment

Various studies have been conducted into classroom involvement of language learners, and their involvement in assessment is of great importance and interest. According to Lewknewicz and Moon (1985), criteria used in general classroom evaluation are not usually explained to the learners. Therefore, in many cases those marking scales used by the teachers in evaluating learners’ work are unclear and inconsistent. Moreover, using different standards and scales for grading in different types of work may confuse learners about whether their work is being compared with that of other learners or with their previous work. They do not understand the criteria by which they are judged, and they do not know why the work is marked in such a way.

In the last decade, alternative assessments such as self-, peer, and co-assessment have received much attention and several forms of more authentic assessment have been
introduced (Birenbaum & Dochy, 1996). Peer assessment is one of the alternative assessments believed to enable learners to develop abilities and skills more than the teacher alone assessment (Cheng and Warren, 2005).

The definitions of peer assessment vary. Topping (1998) defines peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of similar status” (p.250). Falchikov (1995) defines peer assessment as the process whereby groups of individuals rate their peers. Peer assessment might entail previous discussion or agreement over criteria. Rating instruments or checklists might be used depending on needs. Sluijsmans et al. (1999) view peer assessment as a grading procedure and a learning process in which skills are developed.

Peer assessment is believed to benefit learners by developing students’ critical thinking (Alderson, 1985; Cheng and Warren, 2005). Moreover, it is believed that if learners can be trained to assess their peer confidently and reliably, they will also be able to evaluate their own language skills to improve themselves (Cheng and Warren, 2005). Patri (2002) proposes that using peer assessment training could help students evaluate their peer performance comparable to their teachers, even with learners of low level of language proficiency. Nakamura (2002) claims that understanding rating criteria in evaluating peers can help students to prepare their own presentation to reach their goal.

In addition, many researchers believe that peer assessment can increase students’ responsibility. Peer assessment provides an opportunity for learners to take responsibility in considering and evaluating both the learning process and the product of their peers (Jafarpur, 1991; Cheng and Warren, 2005).

However, there are certain limitations in peer assessment. One is determining the criteria for peer assessment. Another one is the language proficiency of the students. Cheng and Warren’s study (2005) showed that students felt uncomfortable and uncertain of their ability to assess their peers’ language proficiency compared to other language criteria they were asked to assess because they were not confident in their language proficiency in doing peer assessment task. Miller and Ng (1994) find that students were doubtful about the reliability of their peer assessment. They commented that evaluators should have more language proficiency than themselves. In addition, they preferred to be assessed by their instructor rather than by their peers.
4.2 Related Research

Cheng and Warren (2005) conducted a study investigating 51 students’ attitudes toward peer assessment and the effect of peer assessment training. Students’ proficiency was also examined. Most of students scored an ‘E’ grade which roughly corresponds to a score of 500 on TOFEL. Peer assessment was used as a part of evaluating students’ group project that was composed of seminar, an oral presentation and a writing report. Among important results were that students were uncomfortable and uncertain in their ability in assessing their peer’s performances. It was also found that significant differences between peer and teacher assessments were more prevalent in seminars than in oral presentations for both the language proficiency and other assessment criteria. The researchers proposed that if language learner could be trained to assess their peers’ language proficiency confidently and reliably, they would also be able to evaluate their own language skills and improve themselves. Also, it is believed that peer assessment is beneficial in developing students critical thinking and facilitating a deeper approach to language learning.

4.3 Role-play

Olsen and Christainsen (1966 cited in Krysia, M. Yardley-Matwiejczuk 1997), define role-play as “a range of activities characterized by involving participants in ‘as-if’ or ‘simulated’ actions and circumstances. For example, someone may be asked to ‘imagine’ being in a dentist’s waiting room anxiously awaiting a painful procedure, or to be a victim following a mugging.” Role-play is one of communicative activities which give students an opportunity to practice communicating meaningfully in different contexts and different roles (Freeman, 1986). The major advantage of role-play is that, it is highly motivating and gives students simple, direct and rapid feedback on their actions (Morry, 1999).

5. Framework of the study

This study aims to investigate the effectiveness of peer assessment of first year students of Prince of Songkla University, Faculty of Liberal Arts, in “Fundamental English Listening and Speaking course” when compared to that of their teacher. “Fundamental English Listening and Speaking” is a subject that emphasizes listening and speaking skills. Role Play is used as a part of the evaluation of students’ oral language proficiency. The criteria for students’ role-play evaluation are composed of 6 categories. The first criterion, content, focuses on the task. The second criterion, preparation, concerns students’ group work,
participation and time management. The third, expression & vocabulary, focuses on the use of expression and vocabulary learned in the course. The fourth criterion, fluency, focuses on smooth speech delivery. The fifth, comprehensibility, considers the clarity of communication and the last criterion, voice & pronunciation, evaluates natural voice and pronunciation of the students while role-playing.

6. Research Methodology

6.1 Subjects of the study

The subjects of this study were 51 first year students, Faculty of Liberal Arts of Prince of Songkla University in the first semester of 2010 academic year, selected through purposes random sampling technique. The subjects were attending Fundamental English Listening and Speaking Course (890-101) a compulsory subject for all first year students at Prince of Songkhla University, HadYai Campus. They were of mixed language abilities.

In this study, the students were systematically divided into three groups based on their English O-NET scores (The Ordinary National Educational Test) which is used to assess grade 12 students’ academic proficiency, including English, for university entry. This data was analyzed to establish the correlation between proficiency level and the teacher assessment. In other words, the study investigates whether proficiency level affects the degree of consonance between the peer assessments and the teachers’ assessments.

The subjects were divided into three proficiency groups based on their English O-NET scores as follows:

Table 1 Subjects by proficiency levels

<table>
<thead>
<tr>
<th>English O-NET Scores (Total=100)</th>
<th>Numbers of students</th>
<th>Proficiency level</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-30</td>
<td>7</td>
<td>Low</td>
</tr>
<tr>
<td>31-49</td>
<td>26</td>
<td>Middle</td>
</tr>
<tr>
<td>50-68</td>
<td>18</td>
<td>High</td>
</tr>
</tbody>
</table>

7. Instruments

7.1 Role Play task

Role Play was used to evaluate oral language proficiency of students for final exams. In the final exam, the students were divided into groups of four and required to do a 10 minute role play on a topic given by the teacher. Each group was given a cue card to perform in front of the class and was evaluated by their teacher and peers. The topics for role play were based
on real life communication learned in the course e.g. asking for places in a city; talking about items in a supermarket and indoor exercise.

7.2 Peer assessment form

Peer assessment, in addition to teacher assessment, is normally used in Fundamental English Listening and speaking course (890-101) as an instrument to check students’ performance for role play tasks that are a part of exams. Usually, the teachers assign students to do role play and peer assessment in a group. Namely, the groups evaluate the other groups’ role play performance, while the teacher assesses the performance of every group.

In the study, with the permission from the teachers teaching the subject group, peer assessment was done individually instead of groups. That is, each subject individually evaluated their peers and in turn was evaluated by their peers based on the assessment criteria.

The peer assessment form consists of 6 criteria: content, preparation, expression & vocabulary, fluency, comprehensibility, voice & pronunciation for evaluating students’ performance. Detailed criteria were earlier mentioned under framework of the study.

7.3 Teacher assessment form

In addition to peer assessment, the teachers also evaluated their students’ role play performance by using teacher’s assessment form. The criteria in the teacher assessment form parallel those in the peer assessment form. The teacher assessment was then compared with peer assessment.

8. Data collection procedure

The data were collected at the end of the first semester of the 2010 academic year. The research was conducted in the following stages:

8.1 Orientation

At the beginning of the semester, the researcher spent about half an hour explaining to the subjects the objectives of the research, the process of peer assessment and the criteria of peer assessment.
8.2 Role play

About one week before the final role play exam, each group consisting of 4 students was given a cue card to perform in front of the class to prepare themselves for the role play. Each group had to create their conversations for the role-play using vocabulary and knowledge learned from lessons 6 to 10 of English book “Talk Time 2”, Unit 6: City Square, Unit: 7 at Supermarket, Unit: 8 Shop & Store, Unit: 9 Hobbies and Unit: 10 Travel plan. Each role play was a combination of the mentioned topics.

Each group had to perform a 10 minute role-play in front of the class. While students were performing the role play, their peers individually evaluated their performance using a peer assessment form, and the teacher assessed students’ performance using the teacher’s assessment form. All subjects were evaluated by the same teacher and their peer. Then a comparison between the students’ and the teacher’s assessment was done.

9. Data Analysis

The data of peer assessment from the final role-play and teacher assessments were computed in terms of mean, SD, and correlation. The data analysis procedure was to answer the research question: Can students with different language proficiency assess their peer oral performance compared to that of their teacher?
10. Findings

To investigate peer and teacher assessments, descriptive statistics were used to analyze each of the 6 assessment criteria that both students and teachers assessed.

Table 1 Comparison of students’ peer assessment by different proficiency groups and teacher assessment

<table>
<thead>
<tr>
<th>Proficiency Levels</th>
<th>Students Peer Assessment</th>
<th>Teacher Assessment</th>
<th>t-test</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Middle</td>
<td>High</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>μ</td>
<td>sd.</td>
<td>μ</td>
<td>sd.</td>
</tr>
<tr>
<td>Content</td>
<td>4.36</td>
<td>0.66</td>
<td>4.35</td>
<td>0.59</td>
</tr>
<tr>
<td>Preparation</td>
<td>4.83</td>
<td>0.38</td>
<td>4.62</td>
<td>0.51</td>
</tr>
<tr>
<td>Expression &amp; vocabulary</td>
<td>4.74</td>
<td>0.45</td>
<td>4.76</td>
<td>0.47</td>
</tr>
<tr>
<td>Fluency</td>
<td>4.17</td>
<td>0.44</td>
<td>4.33</td>
<td>0.55</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>4.71</td>
<td>0.46</td>
<td>4.51</td>
<td>0.54</td>
</tr>
<tr>
<td>Voice &amp; pronunciation</td>
<td>4.26</td>
<td>0.59</td>
<td>4.35</td>
<td>0.57</td>
</tr>
</tbody>
</table>

** significant at .01

Table 1 shows the means and standard deviations for the scores given by teachers and students for each criterion. The result shows that there was a significant difference between peer assessment by students with different proficiency levels and the teacher assessment in terms of preparation, expression & vocabulary and voice & pronunciation (p < 0.01). In other word, students evaluated their peers significantly differently from their teacher.

On the whole, the peer assessment mean scores among three proficiency groups were not different. Students tended to give high scores to their peers in all criteria. The total mean scores of peer assessment for each assessment criteria were higher, though not significantly, than those of the teacher assessment. It can be seen that the mean for the content of peer assessment was 4.45 while that of teacher for the content was 4.36. The total mean score for
preparation was 4.69 by peer assessment and 4.14 by the teacher. The total mean score for the criterion of expression & vocabulary was 4.71 for peer assessment and that of the teacher was 3.79. The means for fluency and comprehensibility by peer assessment were 4.36 and 4.54 while those by the teacher were 4.21 and 4.29. The mean for voice & pronunciation assessed by the students was 4.34 and that assessed by teacher was 3.57. In other words, students tended to overrate their peer’s performance. It can be seen that the mean scores of peer assessment were higher than 4 in all categories while the teacher assessed language criteria (expression & vocabulary and voice & pronunciation) less than 4.

This might show that the students’ language proficiency was not sufficient enough to evaluate their peers in terms of language criteria; they did not know how to assess language criteria correctly. In other words, they might not be able to judge the language accuracy of their peers.

It should also be mentioned that friendship may influence peer assessment. This is illustrated in the finding that the subjects tended to overrate their peer’s performance. According to Cheng and Warren’s study (2005), close relationship affected subjects’ scores given to peers. The subjects in their study believed that the scores given to close friends were not entirely fair and responsible.
Furthermore, ANOVA was utilized to see if proficiency levels and teacher assessment were related by assessment criteria. The results are in the table below.

**Table 2:** Peer assessment among three proficiency groups and teacher assessment

<table>
<thead>
<tr>
<th>Assessment criteria</th>
<th>Peer assessment Low</th>
<th>Middle</th>
<th>High</th>
<th>Teacher assessment</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean, sd</td>
<td>Mean, sd</td>
<td>Mean, sd</td>
<td>Mean, sd</td>
<td>Mean, sd</td>
<td>Mean, sd</td>
</tr>
<tr>
<td>Content</td>
<td>4.36, 0.66</td>
<td>4.35, 0.59</td>
<td>4.62, 0.49</td>
<td>4.36, 0.50</td>
<td>5.89</td>
<td>0.01**</td>
</tr>
<tr>
<td>Preparation</td>
<td>4.83, 0.38</td>
<td>4.62, 0.51</td>
<td>4.79, 0.41</td>
<td>4.14, 0.53</td>
<td>11.17</td>
<td>0.01**</td>
</tr>
<tr>
<td>Expression &amp; vocabulary</td>
<td>4.74, 0.45</td>
<td>4.76, 0.47</td>
<td>4.74, 0.44</td>
<td>3.79, 0.70</td>
<td>18.77</td>
<td>0.01**</td>
</tr>
<tr>
<td>Fluency</td>
<td>4.17, 0.44</td>
<td>4.33, 0.55</td>
<td>4.48, 0.55</td>
<td>4.21, 0.58</td>
<td>4.19</td>
<td>0.01**</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>4.71, 0.46</td>
<td>4.51, 0.54</td>
<td>4.55, 0.53</td>
<td>4.29, 0.61</td>
<td>2.79</td>
<td>0.05*</td>
</tr>
<tr>
<td>Voice &amp; pronunciation</td>
<td>4.26, 0.59</td>
<td>4.35, 0.57</td>
<td>4.44, 0.52</td>
<td>3.57, 0.51</td>
<td>10.63</td>
<td>0.01**</td>
</tr>
</tbody>
</table>

*significant at .05  ** significant at .01

As can be seen in Table 2, the results show that there were significant differences among the assessment by different proficiency groups and teacher assessment in terms of content, preparation, expression & vocabulary, fluency, voice & pronunciation (p < 0.01) and comprehensibility (p<0.05).

It can be concluded that students in the three proficiency groups and the teacher assessed students’ performances differently. It also can be noticed that each proficiency group did not assess their peers comparable to other proficiency groups and to that of their teacher in all criteria. It might be possible to conclude that, in this study, proficiency levels had nothing to do with peer assessment.
To confirm if level of proficiency influences peer assessment, the correlations of peer assessment between the three proficiency groups was analyzed.

Table 3: Peer assessment between three proficiency groups.

| Teacher assessment | Students’ Peer Assessment | | | | | |
|---------------------|---------------------------|---|---|---|---|---|---|
| Criteria            | Low group | Middle group | High group | F | Sig. |
|                     | Mean   | sd     | Mean   | sd   | Mean   | sd   | |
| Content             | 4.36   | 0.66   | 4.35   | 0.59 | 4.62   | 0.49 | 8.55 | 0.01** |
| Preparation         | 4.83   | 0.38   | 4.62   | 0.51 | 4.79   | 0.41 | 6.68 | 0.01** |
| Expression & vocabulary | 4.74 | 0.45   | 4.76   | 0.47 | 4.74   | 0.44 | 0.11 | 0.90 |
| Fluency             | 4.17   | 0.44   | 4.33   | 0.55 | 4.48   | 0.55 | 5.81 | 0.01** |
| Comprehensibility   | 4.71   | 0.46   | 4.51   | 0.54 | 4.55   | 0.53 | 2.57 | 0.08 |
| Voice & pronunciation | 4.26  | 0.59   | 4.35   | 0.57 | 4.44   | 0.52 | 1.74 | 0.18 |

Significance: **p<0.01

As can be seen in Table 3, the results indicate that there was a significant difference among peer assessment by the proficiency groups in terms of content, preparation and fluency (p < 0.01), but not in terms of expression & vocabulary, fluency, comprehensibility, and voice & pronunciation.

From the data, no systematic pattern of the effect of proficiency level on peer assessment emerged, suggesting that, at least in this study, proficiency levels did not play an important role in peer assessment. Suffice it to conclude that students with different language proficiency assessed their peers significantly differently from their teacher regardless of proficiency levels.
This conclusion seems to be confirmed by the analysis of the relationship between proficiency groups and teacher assessment which reveals some interesting results.

**Table 4:** The correlation between peer assessment by different proficiency groups and teacher assessment scores awarded for each of the assessment criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Teacher assessment</th>
<th>Peer assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low group</td>
<td>Middle group</td>
</tr>
<tr>
<td>Content</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Preparation</td>
<td>0.42</td>
<td>0.33</td>
</tr>
<tr>
<td>Expression &amp; vocabulary</td>
<td>0.47</td>
<td>0.03</td>
</tr>
<tr>
<td>Fluency</td>
<td>0.31</td>
<td>0.00</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.860**</td>
<td>0.40</td>
</tr>
<tr>
<td>Voice &amp; pronunciation</td>
<td>0.60**</td>
<td>0.67**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level 1–tailed.

The result shows there was no significant correlation between peer- and teacher assessment in terms of content, preparation, expression & vocabulary and fluency in all three proficiency groups.

There were only significantly moderate correlation between the assessment by students’ proficiency groups and that of the teacher in terms of voice & pronunciation and a strong correlation between the assessment by the low group and the teacher assessment. There was no significantly correlation in other criteria. Again, this might help confirm the conclusion reached earlier that proficiency level did not affect peer assessment.

The finding of a significant correlation between peer assessment by the three proficiency groups and the teacher assessment in terms of voice & pronunciation might be explained by the fact that this criterion is clearly noticeable and obvious to be judged and evaluated by the students regardless of proficiency level since it involves sound projection. The students could simply evaluate it by judging whether their peer sounded native-like or not. And this simple nature might help explain the significant correlation between peer- and teacher assessment.
11. Discussion

11.1 Comparison of students peer assessment with different proficiency and teacher assessment

The result of peer assessment indicates that students tended to overestimate their peers. This can be supported by the result of the total mean scores of peer assessment for each assessment criterion that were higher than those of teacher assessment. This shows that students can not evaluate their peers comparable to that of their teacher. This is in line with the finding of Patri’s (2002) who found that students were not able to judge each other’s performances in the same way as their teacher. This is probably because students in the study lacked peer assessment training. Patri (2002) suggested that peer assessment training can help students evaluate their peer performance compared to that of their teachers.

11.2 Peer assessment among three proficiency groups and teacher assessment

The study finds a significant difference between all groups’ peer assessment and the teacher assessment. Peer assessment and teacher assessment were different in all categories. This indicates that level of language proficiency influences peer assessment differently. However, no systemic patterns of proficiency level effect were found.

11.3 The correlation between peer assessment with different proficiency levels and teacher assessment

Although the results show that there was no significant correlation between peer assessment of all three proficiency groups and teacher assessment for the first four criteria content, preparation, expression & vocabulary and fluency, there was a significant correlation of the criterion of voice & pronunciation. In brief, peer assessment of voice & pronunciation was comparable that of their teacher.

This suggests that it might be easier for students in all groups to evaluate voice and pronunciation than other criteria. Students might think that a person with good pronunciation and clear voice is good at English. In fact, oral proficiency, as reflected through voice and pronunciation, is somewhat easier to recognize and does not require the same level of proficiency as the evaluation of other language criteria e.g. content and preparation, etc. This is consistent with Cheng and Warren’s (2005), which also claims that oral proficiency does not require the same linguistic competence involved in evaluating structures and so on.
12. Conclusion

The study found that teacher assessment and peer assessment was not comparable in terms of content, preparation, expression & vocabulary, fluency and comprehensibility, but the assessment of the two groups was related in terms of voice and pronunciation. This might be because the students lacked training in assessing oral proficiency. If students were given training and more opportunities to practice peer assessment, it might be possible that students will evaluate their peers’ performance consistently and comparable to that of their teacher.

The study demonstrates that students with different language proficiency evaluate their peers differently and are not comparable to teacher assessment. This may show that students’ language proficiency does not influence peer assessment. It should be pointed out here, however, that during the process of peer assessment, students learn to take responsibility in considering and evaluating both the learning process and product of their peers. Peer assessment can increase students’ responsibility and awareness of what is important in oral skills (Jafarpur, 1991; Cheng and Warren, 2005).
References


publications.


