An Analysis of Peer Assessment in Chinese as a Second Language Classroom Presentation

Lu Lu, Ph.D
Department of Modern Languages and Literatures
William and Mary

ABSTRACT

The peer assessment (PA) plays an important role in the classroom presentation. This study aims to 1) investigate the difference in the scorings between PA and teacher assessment in oral presentation, and 2) explore student’s attitude about PA through both quantitative and qualitative approaches. A class of 12 students in an upper-intermediate Chinese language class adopted PA by using Poll Everywhere during an academic year. The results indicate that intermediate-high Chinese learners and instructor interpret the criteria differently in Pronunciation, Fluency, Clarity, and Accuracy, but agree with the scores on Content and Delivery. Most students are positive that PA is a useful method to help them actively participate in class discussions and improve oral presentation ability. Students also reported that written comment is the most useful feedback. Findings and limitations were also discussed.

Keywords: peer assessment, teacher assessment, classroom oral presentation, Chinese as a second language

在课堂口语报告测评时，除了由教师提供反馈和测评意见外，学生的互评也可以带来一定的反馈。本文对北美一高校中高年级汉语课学生随堂个人口语报告进行了近一年的数据收集及追踪，通过 1) 比较学生互评及教师评价的差异性，以及 2) 对学生进行有关互评的问卷调查来了解学生互评在对外汉语课堂中应用的可行性和实践性。根据数据表明，学生在发音准确性，语言流利度，语言清晰度以及语言准确度与教师评分有着较大误差，仅仅在内容和演讲技巧方面保持一致。学生互评普遍高于教师评价，这与学生的语言水平以及对评分标准的应用有较大关系。绝大多数学生认为互评可以有效提高课堂参与度以及自我汉语口语报告能力。书面评语则是最有效的反馈方式。本文也对研究结果、局限以及未来研究方向进行了进一步的讨论。

关键词：学生互评；教师评价；课堂报告；对外汉语
Introduction

Peer assessment (PA) has been increasingly advocated on the grounds of the learning benefits. Topping (1998, 2009) defines PA as “an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners.” PA is a process of a group of individuals grading peers in which may or may not involve agreed criteria among teachers and students, and such process should involve students grading or giving feedback on their peer’s work (Davies, 2006; Falchikov, 1995). It is also a practice that “emphasizes skills, encourages involvement, focuses on learning, establishes a reference, promotes excellence, provides increased feedback, fosters attendance and teaches responsibility” (Weaver & Cotrell, 1986, p. 25).

PA is considered as an effective pedagogical strategy for enhancing learnings. For example, PA can increase students’ engagement (Bloxham & West, 2004), promote critical thinking (Sims, 1989), and increase motivation (Topping, 2009). It can also encourage in-depth learning and help students to develop a better understanding, control and, autonomy of the process of learning (Boyle & Nicole, 2003; Dochy, Segers & Sluijsmans, 1999; Topping, 2009; Wen & Tsai, 2006). However, validity, which refers to the consistency between peer ratings and teacher ratings is always a concern for teachers who are interested in using PA. Previous studies showed inconsistent levels of agreement between students and teacher ratings. Falchikov and Goldfinch (2000) conducted a meta-analysis on previous 48 studies and found that weighted correlation coefficient between teacher and peer rating is 0.69. Similar findings were also found by Cho, Schunn, and Wilson (2006) with an agreement of 0.60 and Harris (2011) with an extremely high correlation of 0.97. However, Kovach, Resch, and Verhulst (2009) found low correlations of 0.29 between peer and teacher’s rating. Lagan et al. (2005) found that students score on average 5% higher than marks given by their instructors. Cheng and Warren (1997, 2005) pointed out that students did not always assess the same elements or criteria as their teachers did, and correlations between teacher and students varied depending on tasks and situations.

In addition to the discussion of benefits and weakness of PA, some studies also focused on students’ perception of PA. Student engagement in assessment could increase their confidence and enhance subject knowledge (Vickerman, 2009). Although students in some studies held a positive attitude toward the evaluating process of PA, students in other researches had a more negative perspective. For example, students stated that PA hindered their relationships with peers and their relationships with classmates might also affect their critical judgment (De Grez & Roozen, 2012; Dochy, Seger, & Sluijsmans, 1999; Falchikov, 1986; Lang et al., 2008; Lindblom-Ylanne et al., 2006; Topping et al., 2000).

There are two modes of PA, which is paper-based and computer-assisted. Computer-assisted PA has the following advantages compared with paper-based PA: a) increase efficiency in classroom for teachers; b) keep anonymity and promote fair assessment without being influenced by “friendship bias” (Li et al., 2016; Lin, Liu, & Yuan, 2001; Wen & Tsai, 2008); and c) perform freely without time and location restriction (Li et al., 2016; Wen & Tsai, 2008). The use of PA on digital platforms is growing.

PA can be used to assess writing, oral presentation, test performances or other skilled behaviors, and its activities can vary in several ways, operating in different subjects (Topping, 2009). Despite the growing popularity of the use of PA in the field of finance and business, law, engineering, biology and ESL (Butcher et al., 1995; Cheng & Warrant, 2005; Fallows & Steven,
of topics they included Chinese idiom story, population issues in the world, my ideal life, the definition of family, difference between Chinese and American education.

**Research Instrument**

Prior to the test, six existing assessments scales to judge the quality of an oral presentation were analyzed by two experienced teachers. Based on the results of the teachers’ discussion, a rubric, consisting of six oral presentation evaluation criteria were developed (See Appendix 1). For each formative assessment, students were asked to evaluate each classmate’s performance on the following six categories, which were 1) Content, 2) Pronunciation, 3) Fluency, 4) Clarity, 5) Grammar Accuracy, and 6) Delivery.

Students were required to present three individual presentations for each semester, and they presented six speeches in total. As a requirement of the course, the instructor decided the topic of each presentation, and students were free to present their ideas related to the topic. The topics included Chinese idiom story, population issues in the world, my ideal life, the definition of family, difference between Chinese and American education.
In the class, the APP, Poll Everywhere, was used to evaluate presenters’ oral performance. The instructor created a presentation classroom on polleverywhere.com, and all the students were asked to join the same presentation room, https://pollev.com/peereval (See Figure 1). Both students and the instructor evaluate peer’s presentations were evaluated. Each category was assessed on a scale from 1 to 5 (1 being the lowest and 5 the highest). An open question for comments was also provided.

By the end of the semester, an online survey about students’ perspectives on PA was sent out. There were eight questions in the survey. For each question, a five-point Likert scale was used. Each question addressed a specific dimension. For instance, “I think PA is a fair evaluation method” addressed the eligibility issue; “The feedback of PA is useful” was related with learning issue; “PA stimulates me to make more effort to prepare for the presentation” addressed the motivation issue; “PA helps me develop a sense of participation” addressed the interaction issue.

![Figure 1. Screenshot of Poll Everywhere](https://engagedscholarship.csuohio.edu/cltmt/vol1/iss3/3)

**Procedure**

Students presented in total six presentations in a successive academic year (3 presentations each semester).

Before each presentation, the assessment criteria were well explained to and discussed with students. The instructor showed a sample video to students and explained the rating scale.

In the class, when one student was presenting, the rest of the class evaluated the presenters’ oral performance anonymously using the APP, Poll Everywhere. The instructor also evaluated students’ presentation. All students’ presentations were videotaped for future reference.

One week after each presentation, students received both instructor’s and peers’ feedbacks. Students had the opportunity to discuss the grade with the instructor about the
assessment in the middle of the semester. After three presentations, students received the online survey. Therefore, the data included peers’ and the instructors’ grading on six presentations and results from two surveys.

Results

1. **Is there any statistically significant difference between students’ PA and teacher’s assessment?**

After calculating the sum score of the six rubric categories, scores of teacher and peer assessors are compared. Figure 2 illustrates the mean scores of peer and teacher’s evaluation of six rubric categories. Overall, it is clear that students intend to rate higher than the instructor through all the categories.

Table 1. **Summary of Peers’ Ratings vs. Teachers’ Ratings (1=Needs improvement, 5=Exceeds expectation)**

<table>
<thead>
<tr>
<th></th>
<th>Student</th>
<th>Teacher</th>
<th>N</th>
<th>t</th>
<th>P</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>4.69</td>
<td>0.18</td>
<td>4.56</td>
<td>0.30</td>
<td>12</td>
<td>1.60</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>4.66</td>
<td>0.25</td>
<td>4.27</td>
<td>0.49</td>
<td>12</td>
<td>2.78</td>
</tr>
<tr>
<td>Fluency</td>
<td>4.72</td>
<td>0.20</td>
<td>4.46</td>
<td>0.35</td>
<td>12</td>
<td>3.14</td>
</tr>
<tr>
<td>Clarity</td>
<td>4.82</td>
<td>0.16</td>
<td>4.23</td>
<td>0.41</td>
<td>12</td>
<td>5.27</td>
</tr>
<tr>
<td>Accuracy</td>
<td>4.85</td>
<td>0.14</td>
<td>4.17</td>
<td>0.54</td>
<td>12</td>
<td>4.39</td>
</tr>
<tr>
<td>Delivery</td>
<td>4.77</td>
<td>0.18</td>
<td>4.48</td>
<td>0.33</td>
<td>12</td>
<td>2.06</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A paired t-test is conducted to compare the means of students’ grading and teacher’s grading. The results indicate that intermediate-high Chinese learners and the instructor agree with the score on Content ($t=1.60, p=.14$), and Delivery ($t=2.06, p=.06$), but there is a significant difference on Pronunciation ($t=2.78, p<.05$), Fluency ($t=3.14, p<.05$), Clarity ($t=5.27, p<.05$), and Accuracy ($t=4.39, p<.05$). The correlation between the PA and the instructor’s assessment is 0.232, indicating that the students’ PA of the oral presentation is not highly consistent with the assessment of the instructor. Also, students tend to score within a more bunched range than the instructor since the standard deviations of students’ scorings are approximately half of the instructor’s.
2. **What is the students’ attitude about PA?**

Peers’ average score of PA reflects a positive opinion on PA method ($M_1=4.21$, $M_2=4.38$, overall $M=4.29$). The results of one-sample t-test indicate that students believe that the feedback from PA is helpful for them to learn Chinese ($t=1.65$, $p=.23$), PA motivates them to prepare better for the presentation ($t=3.83$, $p=.40$), and PA helps them to develop a sense of participation in the class ($t=1.43$, $p=0.37$), and PA is a fair evaluation method ($t=-0.40$, $p=.34$). Although there is no obvious score difference between the first and the second survey regarding all dimensions ($p=.25$), all the means of the Learning, Interaction and Motivation show a slight increase in the second assessment. The only drop happens on students’ views of the eligibility. All the students believe that they are eligible to assess peer’s presentation and their friendship may not affect their evaluation in the first survey ($M=5$). However, one student changed his mind in the second survey because he was afraid his friendship with peers might affect his judgments ($M=4.91$). See Table 2.

Students also provided their comments on PA. Most students hold the positive attitude toward PA. For example, students commented, “I could identify classmates’ strengths and weakness through evaluating their presentations. It somehow helps me to improve”; “I would rehearse more because I want to be better in front of my classmates”; “The PA helps me to concentrate on other’s presentation in class, and it helps me to realize the pronunciation of certain words and learn many new vocabularies.”

Table 2. **Summary of 1st and 2nd After-assessment survey** (1=Strongly Disagree, 5=Strongly agree)

<table>
<thead>
<tr>
<th></th>
<th>1st survey</th>
<th></th>
<th>2nd survey</th>
<th></th>
<th>$t$</th>
<th>$p$(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>4.05</td>
<td>1.24</td>
<td>4.27</td>
<td>0.79</td>
<td>1.65</td>
<td>.23</td>
</tr>
<tr>
<td>Interaction</td>
<td>4.10</td>
<td>1.22</td>
<td>4.34</td>
<td>1.02</td>
<td>1.43</td>
<td>.37</td>
</tr>
<tr>
<td>Motivation</td>
<td>3.67</td>
<td>1.24</td>
<td>4.01</td>
<td>0.56</td>
<td>3.83</td>
<td>.40</td>
</tr>
<tr>
<td>Eligibility</td>
<td>5</td>
<td>0.00</td>
<td>4.91</td>
<td>0.29</td>
<td>-0.40</td>
<td>.34</td>
</tr>
</tbody>
</table>
Figure 3 shows the result of the sixth survey question, “among all the categories, which one provides the most useful feedback.” From the pie chart, it is clear that 32% of the students found the Comment the most useful feedback, and Content the least useful feedback. The other categories, such as Pronunciation (13%), Fluency (11%), Clarity (14%), Accuracy (11%) and Delivery (16%) are almost equally distributed.

![Pie chart showing feedback distribution](image)

**Figure 3.** Distribution for the most useful feedback

**Discussions and Conclusions**

This study analyzes the differences between students and instructor’s assessment in CSL classroom oral presentations and explores students’ attitudes toward PA. Comparison of teacher and PA ratings points at a relatively negative relationship, but also at critical similarities. Among all the grading categories, students grade much higher than the instructor in Pronunciation, Fluency, Clarity, and Grammar accuracy. The low correlation on the grading suggests that peers and the instructor interpret the criteria differently, and the low correlation of 0.232 is relatively similar to Kovach, Resch, and Verhulst (2009)’s finding of 0.29. The findings that peers score higher and a narrower range compared with teachers also agree with the results of other previous studies (Cheng & Warren, 2005; Elfering, Grebner, & Wehr, 2012; Langan et al., 2008). Also, there is a level of agreement when assessing different aspects of the oral presentations. As stated above, students and teachers agree on the ratings of the Content (e.g., choice of topic and structure) and Delivery (e.g., eye contact, note reading). Such aspects of judgment do not require high-level language proficiency. Instead, it involves more soft skills or tacit knowledge (e.g., an experience-based knowledge) of presenting a topic. When regarding language-skills aspects which needs more proficient language ability, students and the instructor have different opinions. The instructor makes more strict and negative assessments. De Grez, Valcke, and Roozen (2012) stated that the difference between peer and teachers’ gradings could be explained by the differences in the width and depth of students and teachers’ experience and proficiency basis. Based on the findings, students’ judgments on tacit knowledge-related are similar to teacher’s perspective. Therefore, it can be estimated that students can improve their assessing skills when they have more proficient language skills.
Another possible explanation is that students still lack the confidence of applying the oral presentation rubric to assess peer’s language proficiency. For example, although the assessment criteria were well explained to students before each presentation, students still commented that “A 5 in pronunciation might mean 'perfect native speaking' for some but 'relatively error-free tones' for others”. Therefore, to improve the validity of PA, a comprehensive clarification of assessing rubric is needed. Cheng and Warrant (2005) believed that language learners could be trained to confidently and reliably assess the language proficiency of their peers, and the finding of this study is in line with their ideas.

Regarding the research question focusing on student perceptions of PA, the results reflect a positive attitude towards the value of the PA. One student commented on the survey that, “The peer evaluation helps me to concentrate on other’s presentation in class, and it helps me to realize the pronunciation of certain words and learn many new vocabularies.” The results indicated that the actual process of conducting the PA affects the perception in a very positive way. It is a promising finding in the light of the impact of perceptions on the outcomes of student learning (Struyven et al., 2003). It is possible to predict that students’ perceptions of PA will improve their participation in class and at the same time, considering the feedback generated by peers and do something with that feedback.

Although all the data are naturally collected from the actual classroom within a relatively long-term period, the study remains limited when it comes to sampling size, learners with different proficiency level, duration of the instructional intervention, scope of skills to be mastered and the complexity level of the competencies. Besides, it is true that there are similarities in assessment principles in second language education, essential differences exist due to linguistic features of the target languages and/or different cultural backgrounds of the instructors and students. For instance, an English native speaker may interpret differently from a Chinese native speaker on how the aspects of tonal languages should be assessed. The research focusing on CSL can not only enrich the understanding about PA in language assessment but also provide implications for teaching and learning tonal languages. Based on the current findings, future studies should involve considering the correlation of each aspect of the presentation. For example, there is a tendency that students score higher in pronunciation when the presenter has a better non-verbal behavior. Also, more investigation needs to work on the effect of peers’ comments and after-presentation group discussion. In addition, Falchikoc (2005) suggests that teachers should develop evaluation criteria in close collaboration with students because it can give students sufficient practice and discussion to develop a shared understanding of explicit and tacit assessment criteria. Future studies can also include students’ opinions when designing the assessing rubrics instead of simply adopting the existing scales.
References


affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education, 30*(1), 21-34.
### Appendix I. Individual Oral Presentation Grading Rubric

<table>
<thead>
<tr>
<th>Type</th>
<th>Exceeds expectation (4–5)</th>
<th>Meets expectation (3)</th>
<th>Needs improvement (1–2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td>The presentation goes beyond the topics of this semester and is very interesting and inspiring. Presenter was well prepared and led a thoughtful and stimulating discussion with peers.</td>
<td>The presentation is interesting, and relevant to the topics of this semester. Presenter seemed to have prepared questions for discussion, but no significant dialogue with peers took place.</td>
<td>The presentation is not very relevant to the topics of this semester. Presenter did not prepare to lead the class discussion. No questions were prepared.</td>
</tr>
<tr>
<td><strong>Pronunciation</strong></td>
<td>All pronunciation is correct; All intonation sounds natural.</td>
<td>Most pronunciation is correct; Most intonation sounds natural; Pronunciation and intonation do not interfere with comprehensibility.</td>
<td>Some pronunciation is incorrect. Some intonation is inappropriate; Pronunciation and intonation interfere with comprehensibility.</td>
</tr>
<tr>
<td><strong>Fluency</strong></td>
<td>Speak fluently all the time; Language flow sounds natural; Pauses are appropriate for speech purpose.</td>
<td>Speak fluently most of the time; Language flow is generally good; Obvious long unnecessary pauses less than 3 times.</td>
<td>Stop frequently; Fluency interferes with language flow.</td>
</tr>
<tr>
<td><strong>Clarity</strong></td>
<td>Use quite a few extended vocabularies and/or sentence patterns appropriately; Express ideas or meaning 100% clear.</td>
<td>Vocabularies and sentence patterns show a great variety; Generally, express ideas or meaning clearly.</td>
<td>Vocabularies and sentence patterns do not show variety; Sometimes the ideas or meaning is not clear.</td>
</tr>
<tr>
<td><strong>Grammar Accuracy</strong></td>
<td>All vocabularies are used correctly and appropriately; All structures are used correctly and appropriately.</td>
<td>Most vocabularies are used correctly and appropriately; Most structures are used correctly and appropriately; Misusage does not interfere with comprehensibility.</td>
<td>Some vocabularies are used incorrectly or inappropriately; Some structures are incorrect or inappropriate; Misusage interferes with comprehensibility.</td>
</tr>
<tr>
<td><strong>Delivery</strong></td>
<td>Appropriate use notes; maintain eye contacts with the audience; Proper body languages.</td>
<td>Read notes very often. Fleeting eye contact with the audience; awkward body languages.</td>
<td>Read note all the time; Practically no eye contact; Almost no body-languages.</td>
</tr>
</tbody>
</table>
Appendix II. Online Peer Assessment (Full Screenshot)
Appendix III. Online Survey

1. I think PA is a fair evaluation method
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
2. The feedback of PA is useful
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
3. PA helps me develop a sense of participation in the class
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
4. PA stimulates me to make more effort to prepare for the presentation
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
5. The use of online evaluation system is efficient.
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
6. My friendship with my classmates may affect my evaluations of them.
   □ Strongly Agree □ Agree □ Neither Agree or Disagree □ Disagree □ Strongly Disagree
7. Among all the categories, which one provides the most useful feedback?
   □ Content
   □ Pronunciation
   □ Fluency
   □ Clarity
   □ Grammar Accuracy
   □ Delivery
   □ Comments
   □ Overall Grade
8. What suggestions do you have for the presentation evaluation?