

July 2018

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

Lu Lu

College of William and Mary

Follow this and additional works at: <https://engagedscholarship.csuohio.edu/cltmt>



Part of the [Chinese Studies Commons](#)

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Lu, Lu (2018) "An Analysis of Peer-Assessment in Chinese as a Second Language Classroom Presentation," *Chinese Language Teaching Methodology and Technology*. Vol. 1 : Iss. 3 , Article 3. Available at: <https://engagedscholarship.csuohio.edu/cltmt/vol1/iss3/3>

This Article is brought to you for free and open access by the Confucius Institute at EngagedScholarship@CSU. It has been accepted for inclusion in Chinese Language Teaching Methodology and Technology by an authorized editor of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

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

学生互评在中文二语课堂口语报告分析及应用

陆璐

在课堂口语报告测评时，除了由教师提供回馈和测评意见外，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及实践性。根据数据表明，学生在发音准确度，语言流利度，语言清晰度以及语言准确度与教师评分都有着较大误差，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. Langan 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,

2000; Matsuno, 2017; Peng, 2010), few studies have been done in the Chinese as a second language teaching (CSL). Also, among PA studies related with the foreign languages, most of the studies focused on the writing practice, studies focused on classroom oral presentations and foreign language proficiency are very limited (Cheng & Warren, 2005; Zamorano & Montanero, 2018). Therefore, PA in CSL is an underexplored area, and needs more attention.

Purpose of the study

In CSL context, there has not been a study that combines the following elements: a) investigating the PA of classroom oral presentations through one semester in a digital platform; b) comparing peer and teacher assessment of the same piece of work; and c) understanding the students' attitude toward the process of PA. In general, this paper tries to fill in the gap by analyzing the differences between students and teachers' assessment in CSL classroom oral presentations.

To better understand the validity of the PA in the classroom presentation, this research is designed to answer the following two research questions:

- 1) Is there any statistically significant difference between students' peer assessment and teacher's assessment? If yes, in which category, and if not, what is the level of agreement between peer and teacher's assessment?
- 2) What is the students' attitude about PA?

Research Method

Participants

Twelve university students (6 female and 6 males) who enrolled for Chinese 301 and 302: Upper-intermediate Chinese at a medium-sized public university in the United States were involved in the study. Eleven of the students were native speakers of English and one native speaker of Korean. Two of them were freshman students, one student was sophomore, two students were juniors, and the rest were seniors. They had various majors, such as international relationship, economics, marketing, and some undeclared ones. Students' language proficiency level was considered between intermediate-high and advanced-low according to ACTFL Proficiency Guidelines 2015. This group of students was analyzed for one successive academic year, including Fall semester and Spring semester.

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 I). 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.

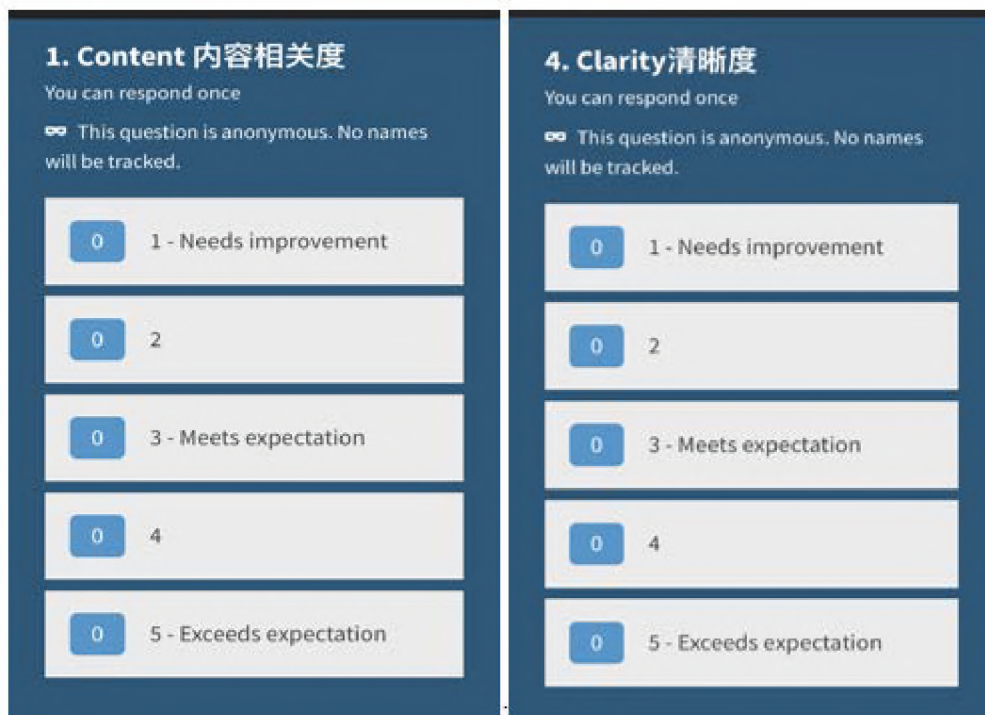


Figure 1. Screenshot of Poll Everywhere

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)*

	Student		Teacher		<i>N</i>	<i>t</i>	<i>P</i>	<i>r</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Content	4.69	0.18	4.56	0.30	12	1.60	.14	.456
Pronunciation	4.66	0.25	4.27	0.49	12	2.78	.02	.269
Fluency	4.72	0.20	4.46	0.35	12	3.14	.01	.287
Clarity	4.82	0.16	4.23	0.41	12	5.27	.00	.295
Accuracy	4.85	0.14	4.17	0.54	12	4.39	.00	.134
Delivery	4.77	0.18	4.48	0.33	12	2.06	.06	.512
Overall								.232

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.

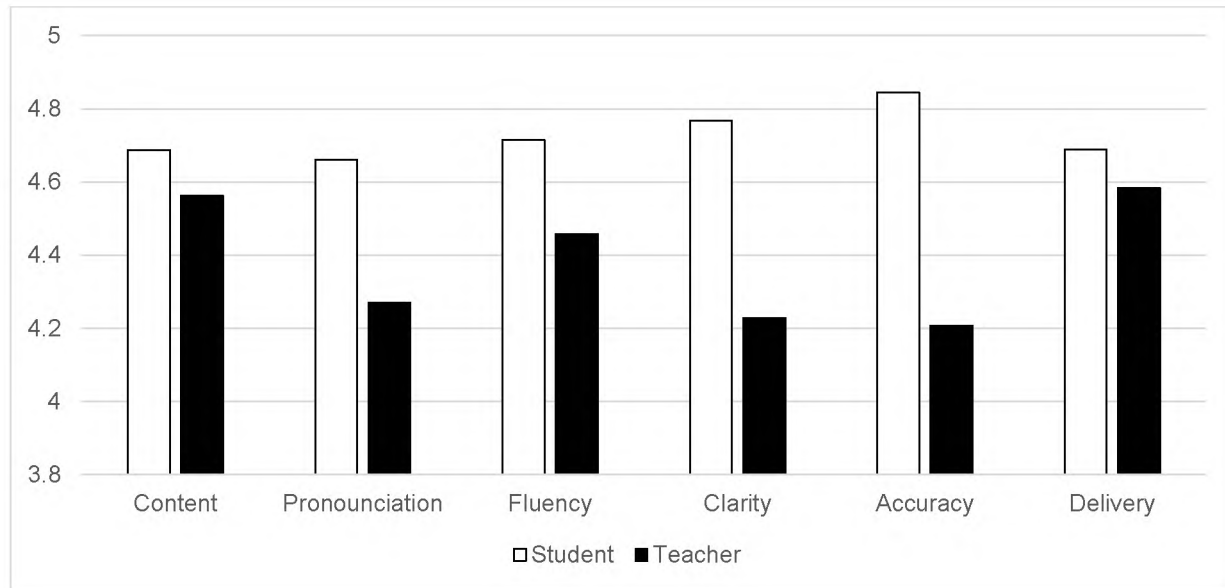


Figure 2. Comparisons of peer and teacher's evaluation of six rubric categories

2. What is the students' attitude about PA?

Peers' average score of PA reflects a positive opinion on PA method ($M1=4.21$, $M2=4.38$, 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 ($t=-0.4$, $p=.34$). 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)

	1 st survey		2 nd survey		<i>t</i>	<i>p</i> (2-tailed)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Learning	4.05	1.24	4.27	0.79	1.65	.23
Interaction	4.10	1.22	4.34	1.02	1.43	.37
Motivation	3.67	1.24	4.01	0.56	3.83	.40
Eligibility	5	0.00	4.91	0.29	-0.40	.34

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.

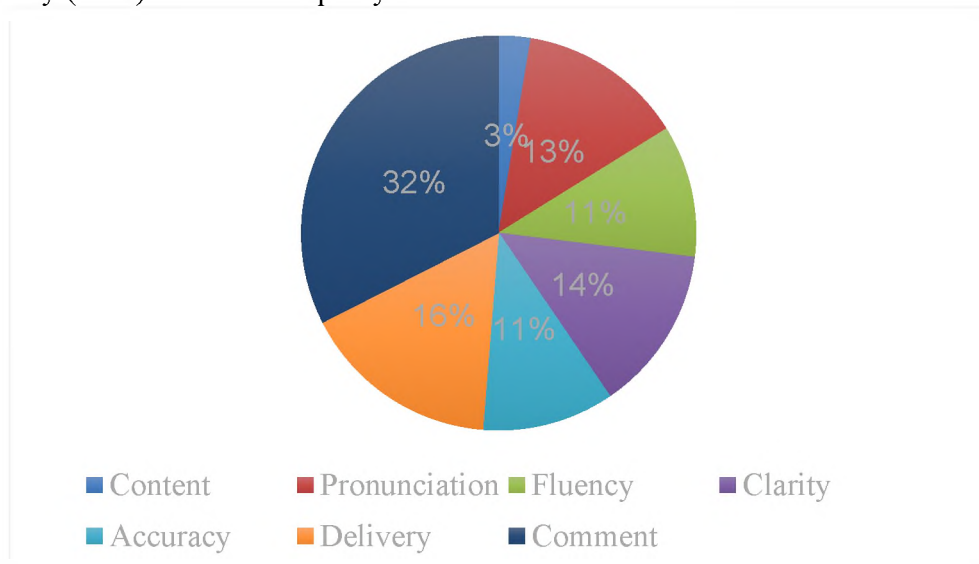


Figure 3. Distribution for the most useful feedback

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 a tonal language 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

- Bloxham, S., & A. West. (2004). Understanding the rules of the game: Making peer assessment as a medium for developing students' conceptions of assessment. *Assessment & Evaluation in Higher Education*, 29 (6), 721–733.
- Bourhis, J., & Allen, M. (1998). The role of videotaped feedback in the instruction of public speaking: A quantitative synthesis of published empirical research. *Communication Research Reports*, 15(3), 256-261.
- Boyle, J. T., & D. J. Nicol. (2003). Using classroom communication systems to support interaction and discussion in large class settings. *Association for Learning Technology Journal*, 11(3), 43–57.
- Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22, 233-239.
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93-121.
- Cho K, Schunn C, & Wilson W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Studies*, 95(4), 891–901.
- Davies, P. (2006). Peer assessment: Judging the quality of students' work by comments rather than marks. *Innovations in Education and Teaching International*, 43(1), 69–82.
- De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer-assessment of oral presentation skills compared with teachers' assessments?. *Active Learning in Higher Education*, 13(2), 129-142.
- Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350.
- Elfering, A., Grebner, S., & Wehr, S. (2012). Loss of feedback information given during oral presentations. *Psychology Learning & Teaching*, 11(1), 66-76.
- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32(2), 175–187.
- Falchikov, N. (2005) *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. New York: Routledge Falmer.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research*, 70(3), 287-322.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education*, 20(3), 289-300.
- Han, C. (2018). A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course. *Studies in Educational Evaluation*. Retrieved from <https://doi.org/10.1016/j.stueduc.2018.01.001>
- Harris, J. (2011). Peer assessment in large undergraduate classes: An evaluation of a procedure for marking laboratory reports and a review of related practices. *Advances in Physiology Education*, 35(2), 178–187.
- Kovach, R. A., D. S. Resch, & S. J. Verhulst. (2009). Peer assessment of professionalism: A five-year experience in medical clerkship. *Journal of General Internal Medicine*, 24(6), 742–746.
- Langan, A. M., Wheater, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., & Preziosi, R. F. (2005). Peer assessment of oral presentations: Effects of student gender, university

- affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education*, 30(1), 21-34.
- Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheeler, C. P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education*, 33(2), 179-190.
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S., & K. Suen, H. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264.
- Lin, S. S. J., E. X. F. Liu, & S. M. Yuan. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420–432.
- Lindblom-Ylänne, S., H. Pihlajamäki, & T. Kotkas. (2006). Self-, peer- and teacher assessment of student essays. *Active Learning in Higher Education*, 7(1), 51–62.
- Matsuno, S. (2017). Adoptability of peer assessment in ESL Classroom. *Creative Education*, 8(08), 1292.
- Murillo-Zamorano, L. R., & Montanero, M. (2018). Oral presentations in higher education: A comparison of the impact of peer and teacher feedback. *Assessment & Evaluation in Higher Education*, 43(1), 138-150.
- Peng, J. C. (2010). Peer assessment in an EFL context: Attitudes and correlations. In *Selected Proceedings of the 2008 Second Language Research Forum*, ed. T. Prior et al (pp. 89-107).
- Price, M. & O'Donovan B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In C. Bryan and K. Clegg (Eds.), *Innovative Assessment in Higher Education* (pp. 100–109). London: Routledge.
- Sims, G. K. (1989). Student Peer Review in the Classroom: A Teaching and Grading Tool. *Journal of Agronomic Education*, 8(2), 105–108.
- Struyven, K., Dochy, F., & Janssens, S. (2003). *Optimising new modes of assessment: In search of qualities and standards*. Springer Netherlands.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. (2003). *Optimising new modes of assessment: In search of qualities and standards*. Springer, Dordrecht.
- Topping, K. (2009). Peer Assessment. *Theory into Practice*, 48(1), 20-27
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & evaluation in higher education*, 25(2), 149-169.
- Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.
- Vickerman, P. (2009). Student perspectives on formative peer-assessment: An attempt to deepen learning?. *Assessment & Evaluation in Higher Education*, 34(2), 221–230.
- Weaver II, R., & Cotrell, H.W. (1986). Peer evaluation: A case study. *Innovative Higher Education*, 11, 25-39.
- Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, 51(1), 27-44.
- Wen, M. L., & C.-C. Tsai. (2008). Online peer assessment in an inservice science and mathematics teacher education course. *Teaching in Higher Education*, 13(1), 55–67.

Appendix I. Individual Oral presentation Grading Rubric

Type	Exceeds expectation (4– 5)	Meets expectation (3)	Needs improvement (1– 2)
Content	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.	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.	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.
Pronunciation	All pronunciation is correct; All intonation sounds natural.	Most pronunciation is correct; Most intonation sounds natural; Pronunciation and intonation do not interfere with comprehensibility.	Some pronunciation is incorrect; Some intonation is inappropriate; Pronunciation and intonation interfere with comprehensibility.
Fluency	Speak fluently all the time; Language flow sounds natural; Pauses are appropriate for speech purpose.	Speak fluently most of the time; Language flow is generally good; Obvious long unnecessary pauses less than 3 times.	Stop frequently; Fluency interferes with language flow.
Clarity	Use quite a few extended vocabularies and/or sentence patterns appropriately; Express ideas or meaning 100% clear.	Vocabularies and sentence patterns show a great variety; Generally, express ideas or meaning clearly.	Vocabularies and sentence patterns do not show variety; Sometimes the ideas or meaning is not clear.
Grammar Accuracy	All vocabularies are used correctly and appropriately; All structures are used correctly and appropriately.	Most vocabularies are used correctly and appropriately; Most structures are used correctly and appropriately; Misusage does not interfere with comprehensibility.	Some vocabularies are used incorrectly or inappropriately; Some structures are incorrect or inappropriate; Misusage interferes with comprehensibility.
Delivery	Appropriate use notes; maintain eye contacts with the audience; Proper body languages.	Read notes very often. Fleeting eye contact with the audience; awkward body languages.	Read note all the time; Practically no eye contact; Almost no body-languages.

Appendix II. Online Peer Assessment (Full Screenshot)

The screenshot displays an online peer assessment interface with a dark blue background and white text. It is organized into eight sections:

- 1. Content 内容相关度**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 2. Pronunciation 发音准确度**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 3. Fluency 流利度**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 4. Clarity 清晰度**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 5. Grammar Accuracy 语言准确度**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 6. Delivery 演讲仪态**
You can respond once
This question is anonymous. No names will be tracked.
Options: 1 - Needs improvement, 2, 3 - Meets expectation, 4, 5 - Exceeds expectation
- 7. Comments 评语**
This question is anonymous. No names will be tracked.
Text input field: Enter a response
- 8. Overall Grade 总体评价 (1-5)**
This question is anonymous. No names will be tracked.
Text input field: Enter a response

At the bottom of the interface, it states: "You have responded to 0 of 8 questions."

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?