TILLOTSON LI
Tung Wah College, Hong Kong

A CASE STUDY OF PEER-RATER DIFFERENCES IN HONG KONG

Abstract:

Assessing student performance in higher education has never been easy. Pedagogy to engage students in learning and make them own their learning has become increasingly popular. The paradigm of teaching has switched from teacher-centered to student-centered. Likewise, students' involvement in assessments (self- and peer-assessments) is becoming more common.

Self- and peer-assessment have widely been researched and evidenced in enhancing and motivating student learning. Differences in assessment results between peers and instructor have been found insignificant. However, differences among peers have not been studied much.

The purpose of this paper is to investigate differences among peer raters. Do friends rate more leniently and not-so-friendly peers rate more stringently?

Data were collected from a business communication course delivered at a Hong Kong private higher education institute for fall 2013. The course had 108 students in 3 sections, and 28 groups were formed to work on a case study. Students were required to orally present their case studies. Presentation was graded by both peer-rater groups and the instructor.

For inter-rater differences, 22 of the 28 groups were rated significantly differently ($P < 0.05$) by their peers. However, the Cronbach's alpha for all groups ($n = 28$) is 0.952 indicating high inter-rater reliability. When the 28 groups were analyzed separately, a total of 19 (67.86%) groups had Cronbach's alpha below the benchmark 0.7 reliability threshold, but only three (10.71%) of them were significant ($p < 0.05$). Among the 94 anonymous questionnaires collected from students after the process, 18 students (19.15%) admitted they had awarded more marks to friendly groups; and 10 (10.64%) had awarded less marks to unfriendly groups. Six students have both awarded more and less marks to peers due to personal relationship. Discriminatory rating by individual students existed, but it was uncertain if final grades were affected.

Seventy-one (75.53%) students agreed that observing other students' presentation helped, and sixty-seven (71.28%) students agreed that the exercise trained their critical thinking skills.

Peer-assessment not only enhances and motivates students' learning, but also makes evaluation of student performance more transparent. This may reduce grade dispute between teacher and students. To reduce impact on final grade from favoritism and/or hostile grading among peers, instructors may consider removing the highest/lowest marks from peer-rater groups and reducing the weight of peer-assessment on the final grade. What is more important is what students learn from the process.

Keywords:

Peer-assessment, discriminatory grading, peer raters.

Introduction

An important but difficult task for teachers is assessing student performance. Assessment carries two functions. First, assessment for learning refers to assessments that enhance student learning in different learning activities (Xiao & Lucking, 2008). Continuous assessment provides useful information to instructors regarding student progress and feedback to students to enhance their learning. Therefore, most continuous assessments are formative; they promote and improve student learning (Li, Liu, & Steckelberg, 2010). Written assignments, term papers, presentation, and term tests are typical examples of formative assessment. Second, assessment of learning refers to assessment that concludes student performance at the final stage of a learning process. Final examinations, final papers or capstone projects are examples of summative assessment.

Assessment involves assessor and assessee. Assessor and assesse have traditionally been the instructor and the student, respectively. Pedagogy to engage students in learning and make them own their learning has become increasingly popular in higher education. What is important is what students learn instead of what teachers teach. The paradigm of teaching has switched from teacher-centered to student-centered. Likewise, for assessment, it is more important for students to learn from the process rather than what their final grades are. Self- and peer-assessment have been widely researched and evidenced as effective tools that enhance and motivate student learning. As such, they are becoming more popular.

While praising merits of peer-assessment in teaching and learning, instructors have had worries about sentimental bias of raters. Sentimental bias can be two-fold. Favoritism occurs when students are rated leniently by their friendly peers (Dollisso & Koundinya, 2011). This paper calls it the “Buddy Effect”. Hostility rating occurs when students are rated stringently by their unfriendly peers. This paper calls it the “Rival Effect”.

The purposes of this study are to investigate if rater differences are significant in peer-assessment, and what students learn from it. The research questions are:

1. Is the rating difference between students and the instructor significant?
2. Is the rating difference among peer-rater groups significant?
3. Do “Buddy Effect” and/or “Rival Effect” exist?
4. What do students learn from the peer-assessment process?

The next section reviews previous literatures on the topic, followed by the methodology adopted in the study. Results of the study will then be discussed, and the paper will end with a concluding remark.

Peer-Assessment

Peer-assessment refers to the form of assessment in which students' performances or achievements are rated by their peers with similar academic background and standard (Li, Liu, & Zhou, 2012; Tillema, Leenknecht, & Segers, 2011; Topping, Smith, Swanson, & Elliot, 2000; Xiao & Lucking, 2008).

Although peer-assessment is a relatively new concept in higher education in Hong Kong, it has been widely used elsewhere for over 50 years (Xiao & Lucking, 2008). Its benefits to both students and teachers are well documented (Berg, Admiraal, & Pilot, 2006; Davies, 2006; Mok, 2010; Tillema, Leenknecht, & Segers, 2011; Xiao & Lucking, 2008). Peer-assessment sometimes can be more powerful and influential than teachers' assessment. It engages students in the teaching and learning activities so that students' motivation is increased (Chang, Tseng, & Lou, 2012; Tillema, Leenknecht, & Segers, 2011)

Benefits to students include: first, students being assessed obtain more ideas and inspirations from their peers; being the assessors, students observe and learn from their peers (Bouzidi & Jaillet, 2009; Chang, Tseng, & Lou, 2012; Chen, 2010). Second, students expecting their works to be assessed by their peers have the motivation and commitment to work harder (Brew, Riley, & Walta, 2009). Third, students actually learn more of the course materials throughout the process (Guilford, 2001). Li, Liu and Steckelberg (2010) find that student assessors who are capable of providing useful feedback tend to perform better in their own work. Li, Liu and Zhou (2012) summarize from other literatures stating student benefits include enhancing students' learning; improving higher order thinking skills; greater responsibility, autonomy, and motivation in learning; and benefiting future careers. Students are also trained to make judgment critically (Dollisso & Koundinya, 2011; Pombo & Talaia, 2012).

While benefits to students have been well documented, students' concern about themselves being the assessors have always been their ability to rate their peers fairly. Some student assessors feel that they are simply not qualified to rate their peers (Mok, 2010). Students query the values of feedback and marks given by their peer assessors because they are in doubt the capability of their peer-assessors (Li, Liu, & Steckelbert,

2010).

Some students find peer assessment time consuming and socially uncomfortable although it challenges them intellectually and improves their learning (Topping, Smith, Swanson, & Elliot, 2000). Some students even admitted dishonestly in doing face-to-face peer assessment (MacLeod, 1999). Some fear their marks exposed to their peers (Tillema, Leenknecht, & Segers, 2011), and some question the fairness of grades assessed by peers (McConlogue, 2012)

On the other hand, teachers concern about rating bias among student assessors. First, they worry about favoritism (Carson & Nelson, 1996; Dollisso & Koundinya, 2011; Kao, 2013); student assessors rate their friends leniently (the Buddy Effect). Second, student assessors may become hostile to and under-grade (the Rival Effect) members of unfriendly groups (Carson & Nelson, 1996). Third, students simply rate their peers randomly (Dollisso & Koundinya, 2011; Kao, 2013). Group pressure can also exist among student assessors, and they are reluctant to disclose their disagreement (Kao, 2013). Students are simply reluctant to criticize their peers (Brew, Riley, & Walta, 2009).

In addition to the above concerns, teachers also worry that they have not been trained enough to conduct peer assessment (Mok, 2010), and students' may not have the capability to assess their peers (Brew, Riley, & Walta, 2009).

Rater differences

When peer-assessment is adopted to evaluate student performance, there are three issues. The first issue is validity. Are the scores of student assessors valid? Second, are differences in scores between student assessors and teacher significant? Third, are differences among student assessors significant?

For the validity of scores rated by peers, Xiao and Lucking (2008) allege that multiple-raters actually produce high validity and reliability of final grades.

When comparing peer-assessment with teacher-assessment, there are contradicting results. Studies have found that the scores of the two types of assessors are consistent (Sadler & Good, 2006; Sung, Chang, Chiou, & Hou, 2005); while other studies have found them inconsistent (Chang, Tseng, & Lou, 2012; Chen, 2010). The phenomenon can be due to various education levels of students and their

understanding of the assessment criteria. The difference between peer-raters and teacher-rater becomes less as the education level gets higher (Chang, Tseng, & Lou, 2012).

For differences among student assessors, hardly any previous research was found. This study intends to fill the knowledge gap by investigating whether such differences are significant.

Data Collection and Methodology

Data were collected from the 3 sections of a business communication course offered by a local college in Fall 2013 involving a total of 108 Bachelor of Business Administration students. Students were divided into 28 groups. The course requires students to work on a communication-related case study in groups of four students. Some groups had three or five students. Each group had to orally present its case followed by a written report. In the oral presentation, each group was assessed by peer-rater groups and the instructor. The average score of peer-rater groups and the instructor's score contributed equally to the total 10 oral presentation marks. The written report carried 20 marks. The total mark of the group case study project weighted 30 marks (percent) towards the final grade.

To ensure students' understanding of the assessment criteria, the instructor has provided and explained the assessment rubrics to students prior to oral presentation. Mock oral presentations by groups were conducted so that the instructor was able to elaborate how marks would be awarded to different levels of presentations aligned with the assessment rubrics (Appendix 1).

To assess oral presentation, peer-rater groups were anonymous and given 10 minutes after each presentation to discuss the performance of the presenting group and come up with a score for each criterion. The instructor collected the marks immediately before the next group presented. The mean score of all peer-rater groups given to the presenting group was then calculated.

A questionnaire for student comments of the peer-assessment exercise was given to and collected from students in class after the whole process. A total of 94 (87%) anonymous questionnaires were collected among the 108 students registered for the course.

For the first research question concerning differences between the instructor and peer-rater groups, the mean score of peer-rater groups and the score of the instructor for each of the 28 group presentations were paired, and paired t-test unequal variances was run. Correlation of the scores awarded to each group between the instructor and peer-rater groups was found. F-test was run to check for difference in variances between the instructor and peer-rater groups.

For the second research question, marks awarded by peer-rater groups for the 28 presentations were analyzed using ANOVA. The overall and individual group's Cronbach's alphas (Multon, 2010) were checked for inter-rater reliability. The benchmark of inter-rater reliability is a 0.7 Cronbach's alpha.

For the third research question concerning Buddy and Rival Effect, results were based on confession of students. There were two questions in the questionnaire: "I have awarded more mark to my friends' groups" and "I have awarded less mark to the groups whom I do not like". Students who have checked "Strongly Agree" and "Agree" were taken as Buddy/Rival Effect. Further quantifying the results was impossible due to the anonymous nature of the peer-rater forms and questionnaire. Analysis was limited to counting the number and finding the ratio of students who have exercised discriminatory rating based on personal relationship in the sample.

To answer the fourth research question, content analysis was conducted for the written comments from the questionnaire. Comments were categorized and analyzed. It has provided solid information to what students have learned and benefited from the exercise.

Results

For the difference between the instructor and peer-rater groups, the overall mean scores of the instructor and the peer-rater groups were 6.4974 and 6.4377 respectively; the former was slightly higher than the latter. The difference was found insignificant ($n = 28$, $t = 0.3140$, $p = 0.7550$). Paired samples correlation was found strong and significant ($r = 0.78$, $p = 0.000$).

Although the mean scores of the instructor and the peer-rater groups were so close, and the difference was found insignificant, F-test for the variances (instructor, 0.7810; peer-raters, 0.2316) was found significant ($F = 3.3728$, $p = 0.0012$).

For differences among peer-rater groups in their ratings of the 28 group presentations, F-test ANOVA for comparing means of peer ratings was found significant ($p < 0.05$) in 22 (78.57%) groups. After the highest and lowest marks were removed from each group and F-test ANOVA was run again for the 28 groups, difference in means in 6 (21.43%) groups was found significant.

For inter-rater consistency, the Cronbach's Alpha for all groups ($n = 28$) is 0.952 indicating high inter-rater reliability. When the 28 groups were run separately, a total of 19 (67.86%) groups had Cronbach's Alpha below the benchmark 0.7 reliability threshold. However, only three (10.71%) of them are significant ($p < 0.05$). For the 9 groups that have Cronbach's Alpha above 0.7, all of them are significant ($p < 0.05$).

For discriminatory rating, among the 94 anonymous questionnaires returned, 18 students (19.15%) admitted they had awarded more marks to friendly groups; and 10 (10.64%) admitted they had awarded less marks to unfriendly groups. Among these students, six of them admitted that they had both. The actual number of students who had discriminately rated their peers was 22 (23.40%). Peer and Rival Effects existed.

What students have learned from the process? A total of 80 students (85.11%) checked "Strongly Agree" or "Agree" on "I understand the assessment rubrics given to me"; 76 students (80.85%) checked "It makes me understand the assessment criteria of oral presentation"; 71 (75.53%) checked "It helps me improve my oral presentation skills through observing peer's presentation"; 67 (71.27%) checked "It develops my critical thinking skills".

There are three open-end questions in the questionnaire. The first question asked for students' comments to improve the peer-assessment. A total of 22 students (23.40%) wrote suggestions. Among them, 14 suggested to increase different criteria and have more elaboration on how marks should be awarded to each criterion. The other eight suggested different things including using electronic version to be more environmentally friendly; leaving space for peer-rater groups to write comments; returning the peer-rater grading to the presenting group; having each student to rate the presentation of each group.

The second open-end question asked if they had concern about the effect of peer-assessment on their final grades. A total of 29 (30.85%) students wrote their concerns. Among them, seven believed it had no effect on their final grades because they believed that their peers would grade them fairly. For the seven concerned about discriminatory grading among peers, four of them questioned if unfriendly peers had

given lower marks, and three concerned about if friends had awarded more marks to their peers. Eleven worried about the ability of their peers in rating their oral presentations. Two students believed their final grades were improved, and two said the weights of peer-assessment could be lower.

The third open-end question asked the cost and benefits of the process. A total of 14 students (14.89%) wrote about the cost. Among them, an overwhelming 12 said time was the cost. One student said it had increased his/her stress without further explanation. Another student said the cost is “unfair to all groups” if discriminatory rating existed.

For the benefits to students and answer to the fourth research question, a total of 42 students (44.68%) wrote their comments. Among these, 11 said they had paid more attention to other groups’ presentation because they had to grade their peers. Ten said they could improve their presentation skills through grading others. Nine said they had learned from others’ presentation. Four believed their marks had been improved through the process. Three said the process had made them understand the course materials better. Other comments included exchanging ideas among students, becoming aware of the assessments, and improving critical thinking. A student wrote:

> *“This is a good experience because this helps us to judge ourselves. We suddenly become aware of our mistakes when we assess others.”*

Discussion:

The findings of this case study are consistent with previous studies in the field (Sadler & Good, 2006; Sung, Chang, Chiou, & Hou, 2005) showing the difference between the instructor and peer-raters is insignificant. The overall mean score awarded by the instructor was slightly higher than peer-raters. This could be due to the following reasons. First, peer-raters tended to be stricter than the instructor. Second, peer-raters were unable to clearly differentiate marks or grades for different levels of attainment. Third, there were more Buddy Effect than Rival Effect, and the former outweighed the latter.

Correlation between the scores awarded by the instructor and peer-rater groups was high and significant. This has added to the reliability and creditability of peer-assessment. It should reduce worries of students concerning possible negative

impact on their final grades because of peer-assessment.

The study has however found significant difference in the variances of the instructor and the peer-rater groups. This can be explained by the fact that the instructor is more capable to differentiate different levels of performance. Students might be unsure the different thresholds, although in the questionnaire 85% (80/94) of students said they had understood the assessment rubrics. It could be true that they understood the rubrics, but they were unable to differentiate the thresholds for each mark/grade. They might not have the capability (Brew, Riley, & Walta, 2009; Li, Liu, & Steckelbert, 2010; Mok, 2010). To play safe, students conservatively gave some kind of average marks to the presentation that they were unable to differentiate the performance. This behavior is shown by the low variance of students relative to the variance of the instructor; the latter is 3.37 times of the former. It is also possible for students to rate their peers randomly (Dollisso & Koundinya, 2011; Kao, 2013) around the mean.

The overall inter-rater reliability was high (Xiao & Lucking, 2008). The low reliability when peer-rater groups' marks were analyzed separately might imply some discriminatorily biased rating (Carson & Nelson, 1996; Dollisso & Koundinya, 2011; Kao, 2013). A high percentage of students (22/94 or 23.40%) admitted that they had awarded marks attributed to their personal relationship with the presenting groups.

A high overall reliability coupled with low individual reliability is not uncommon. Likewise, for multiple regression analysis, it is not unusual that the overall regressed model is significant with some insignificant variables. Among the low reliability groups, only three were significant at 0.05. Unfortunately this cannot be further investigated because all raters were anonymous. Source of significance was unidentifiable. Different marks awarded to different presentations could be due to different performances, emotional factor, and different interpretations of the assessment rubrics. Results of this case study have found that peer-raters overall were consistent in rating different groups, but some peer-rater groups were inconsistent when it came to assess individual groups. Discriminatory rating was suspected based on quantitative data.

For discriminatory or emotional rating, in addition to a relatively high percentage of students who had taken personal factor into rating, a relatively high ratio (7/29 or 24.13%) of students' written comments concerned about peer-assessment's impact on their final grades. This reinforces the views of some literature in the field (Carson & Nelson, 1996; Dollisso & Koundinya, 2011; Kao, 2013) regarding students' concerns of their final grades being affected by sentimental peer-raters.

Some have awarded more marks to their friends, the Buddy Effect; some have awarded fewer marks to the groups whom they did not like, the Rival Effect. However, due to the anonymous nature of the process, it is unable to investigate further if there is a direct relationship between these 22 students and the 3 low reliability peer-rater groups. Furthermore, these 22 students were in different peer-rater groups. Their discriminatory rating intention and/or behavior could have been minimized, if not removed, through the consensus marks awarded by their respective peer-rater groups.

Buddy and Rival Effects are emotional rating. There are two sides of the effects. It is discriminatory, but it also reflects the interpersonal relationship of students. The course is a communication course. Students who were rated more marks from their friends have more friends possibly due to their communication skills. On the contrary, students who were not well-liked by their peers were awarded less marks from their peers. This could also be due to their relatively poor communication skills.

These results suggest that instructors using peer-assessment will have to; first, find ways to minimize impact of discriminatory rating on the final grade such as deleting the highest and lowest marks; second, educate students to rate their peers fairly without taking personal relationship into rating; third, ensure students' understanding of the assessment rubrics; and fourth, ease students' concerns about peer-assessment.

As for learning most students found observing others' had helped their presentation and critical thinking skills. This is consistent with previous studies (Bouzidi & Jaillet, 2009; Chang, Tseng, & Lou, 2012; Chen, 2010). Furthermore, students were more attentive to others' presentations because they felt the obligation to rate the latter. If students were not required to rate, students would not be paying similar kind of attention to others' presentations. This is also reflected by their written comments that an overwhelming proportion of students' response concerning the cost of peer-assessment process was time. All in all, being more attentive to other's presentation has helped their presentation skills, to realize their own mistakes, to understand more of the course materials and assessment criteria, and to train critical thinking skills while rating others. This helps students earning more marks and better grades of their own work. As they understand more about the marking criteria, they can work targeting specifics in the criteria.

Conclusion

Results of the study will provide useful information to instructors if they continue to or plan to adopt peer assessment for the courses they teach. To reduce the possible

effect of Buddy and/or Rival Effect on students' final grades, instructors may do the following. First, instructors will explain to students marking ethics. Marks and grades should reflect the true standards of peer performance independent from personal relationship. Second, instructors may delete the highest and lowest scores for the calculation of final grade. Third, the weight of peer assessment scores will be reduced. Fourth, instructors may moderate final grades for marginal cases to minimize possible effects of emotional grading.

Instructors should provide more demonstration and elaboration of how marks are awarded through students' mock presentations in order to strengthen peer-raters' ability to differentiate different levels of performances. To investigate the grading behavior of each rater (or rater-group) in the future, each rater should be anonymous but identifiable with a unique symbol. This opens the door for future research on Buddy and Rival Effects.

It is unquestionable that peer-assessment enhances and motivates student learning, the remaining issues are how to train students with the skills to rate their peers and rate them fairly without taking personal relationship into grading. To make peer-assessment a joyful teaching and learning process, students' concerns about possible effect of emotional rating should be addressed, and such effect must be minimized, if not completely removed.

References

Berg, I.V.D., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. Assessment and Evaluation in Higher Education, 31(1), 19-36.

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? Educational Technology and Society, 12(4), 257-268.

Brew, C., Riley, P., & Walta, C. (2009). Education students and their teachers: comparing views on participative assessment practices. Assessment & Evaluation in Higher Education, 34(6), 641-657.

Carson, G. J., & Nelson, L. G. (1996). Chinese Students' Perceptions of ESL Peer Response Group Interaction. Journal of Second Language Writing, 5(1), 1-19.

Chang, C., Tseng, K., & Lou, S. (2012). A comparative analysis of the consistency and

difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. Computers & Education, 303-320.

Chen, H. C. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. Computers and Education, 55(1), 229-236.

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.

Devito, J.A. (2009). Human Communication. Pearson International Edition, 11th Edition, P. 378.

Dollisso, A., & Koundinya, V. (December 2011). An Integrated Framework for Assessing Oral Presentations Using Peer, Self, and Instructor Assessment Strategies. North American Colleges and Teachers of Agriculture Journal, 39-44.

Guilford, H. W. (2001). Teaching peer review and the process of scientific writing. Innovations and Ideas, 25(3), 167-175.

Kao, G. (2013). Enhancing the quality of peer review by reducing student "free riding": Peer assessment with positive interdependence. British Journal of Educational Technology, 44(1), 112-124.

Karnilowicz, W. (2012). A comparison of self-assessment and tutor assessment of undergraduate psychology students. Social Behavior and Personality, 40(4), 591-604.

Li, L., Liu, X., & Steckelbert, L. A. (2010). Assessor or Assessee: How student learning improves by giving and receiving peer feedback. British Journal of Educational Technology, 41(3), 525-536.

Li, L., Liu, X., & Zhou Y. (2012). Give and take: A re-analysis of assessor and assessee's roles in technology-facilitated peer assessment. British Journal of Educational Technology, V43(3), 376-384.

MacLeod, L. (1999). Computer-aided peer review of writing. Business Communication

Quarterly, 62(3), 87-94.

McConlogue, T. (2012). But is it fair? Developing students' understanding of grading complex written work through peer assessment. Assessment & Evaluation in Higher Education, 37(1), 113-123.

Mok, J. (2010). A case study of students' perceptions of peer assessment in Hong Kong. English Language Teachers Journal, V(65), 230-239.

Pombo, L., & Talaia, M. (2012). Evaluation of innovative teaching and learning strategies in science education: Collaborative work and peer assessment. Problems of Education in the 21st Century, 43, 86-94.

Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. Educational Assessment, 11(1), 1-31.

Sung, T., Chang, K., Chiou, S., & Hou, H. (2005). The design and application of a web-based self and peer-assessment system. Computers & Education, 45. 187-202.

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning - A review of research studies. Studies in Educational Evaluation, 25-34.

Topping, J. K., Smith, F.,E., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. Assessment and Evaulation in Higher Education, 25, 2, 149-169.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. Internet and Higher Education, 186-193.

Appendix 1

Group Case Study Presentation (Peer Assessment)

Criteria	Possible	10, 9	8, 7	6, 5	4, 3	2, 1
Clarity, Pronunciation	10	All words are spoken clearly, understand everything without difficulty	Most of the words are spoken clearly and understandable	Some words are not spoken clearly, a little difficulty to understand all	A lot of the words are not spoken clearly, difficult to understand most of the words	Most words are not spoken clearly, most of the words are not understandable
Pace/speed (Suggested speed an average of 150 words per minutes (Devito, 2009)). Break and pause.	10	All words are spoken at the right pace and easily understandable, break and pause are used appropriately	Most words are spoken at the right pace and easily understandable, break and pause are mostly used appropriately.	Some words are not spoken at the right pace, either too fast or too slow, making it a little difficult to understand all words. Seldom use break and pause.	A lot of the words are not spoken at the right pace, either too fast or too slow, making it difficult to understand most of the words. No break and pause are used.	Most of the words are not spoken at the right pace, have great difficulty to understand most of the words. Cannot identify break and pause used.
Talk to audience, eye contact	10	All the time talking to the audience. Appropriately use of eye contact to engage audience, have at least a few seconds eye contact to all audience.	Most of the time talking to the audience. Appropriately use of eye contact to engage audience, have at least a few seconds eye contact to most audience.	About half of the time talking to the audience, and half of the time reading own notes. Appropriately use of eye contact to engage some audience.	Most of the time reading own notes, talk to the audience a little. Use eye contact to engage a few audience.	All the time reading own notes. Have not even looked at the audience. Very few or no eye contact with audience.
Body language, gesture, posture	10	Use appropriate body movement, gesture to emphasize main points, appropriate	Appropriate body language, gesture, and posture in line with presentation, at times a bit	Lack of any of the body movement, hand gesture, or posture to emphasize strong points, or posture to	Very little body language, gesture, or posture throughout the oral presentation.	Stand still most of the time, reading own notes. Almost no non-verbal observed.

		posture to demonstrate friendliness.	exaggerated.	demonstrate friendliness.		
Friendliness, smile	10	Appropriately use of smile to demonstrate friendliness all the time.	Appropriately use of smile to demonstrate friendliness most of the time, without laughing for no reason.	Appropriately use of smile to demonstrate friendliness some of the time.	Appropriately use of smile to demonstrate friendliness a little of the time.	No smile.
Organization of materials	10	Materials are systematically organized and presented – Introduction, body, conclusion.	Materials are systematically organized and presented, but missing out an organized introduction or conclusion.	Materials are somewhat organized and presented without an introduction and conclusion.	Materials are not quite organized, but some kind of structure is observed. Introduction-body-conclusion is sort of observed.	No organization of materials is observed.
Handling questions	10	Attentive to and answer questions smoothly and confidently.	Attentive to and answer questions smoothly, without showing much confidence.	Attentive to questions, but unable to provide answers relevant to questions.	Misunderstand questions, but correct answer to the misunderstood question.	Misunderstand questions, and provide unrelated answers.
Total	70					