Influences of Peer Assessment Types on Performance and Students’ Perspectives on Online Peer Assessment

Ren-Yu Liao
Dept. of Information Management, National Yunlin University of Science & Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
E-mail: kevin4143@gmail.com

Ching-Tao Chang
Dept. of Information Management, National Yunlin University of Science & Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
E-mail: changct@yuntech.edu.tw

Chun-Ying Chen (Corresponding author)
Center for General Education, National Taichung University of Education
140, Minsheng Rd., West Dist., Taichung 40306, Taiwan
E-mail: cychen666@mail.ntcu.edu.tw

Received: July 23, 2020    Accepted: September 9, 2020    Published: September 15, 2020
doi: 10.5296/ire.v8i2.17405    URL: https://doi.org/10.5296/ire.v8i2.17405

Abstract
This paper reports on a study involving the design of online peer assessment (PA) activities to support university students’ small-group project-based learning in an introductory course. The study aimed to investigate the influences of different types of PA in terms of the rubric (quantitative ratings), peer feedback (qualitative comments) and hybrid (a combination of the rubric and peer feedback) on students’ project performance, and to explore further students’ perspectives on online PA. The quantitative findings suggested that (a) students in the hybrid condition likely had better project performance than those in the peer feedback condition did,
and (b) students in the rubric condition could perform equally well as those in both of the hybrid and peer feedback conditions. The qualitative findings suggested that besides types of assessment, other possible confounding variables that might affect performance included perceived learning benefits, professional assessment, acceptance, and the online PA system.

**Keywords:** Online peer assessment, Peer feedback, Rubric, The Workshop module of Moodle

1. **Introduction**

Due to the popularity of online learning, online assessment has received much attention, particularly the evaluation of open-response assignments and a wide variety of produced work, including writing, portfolios, presentations, reports, and artifacts. Those take more time and effort for evaluation than multiple-choice questions do. Assessment of learning becomes increasingly problematic when with large classes or there are many assignments during a course.

Online assessment systems have been developed to manage PA and have demonstrated efficacy in dealing with the mentioned problem. The primary system function refers to the automation of administrative logistics that helps to resolve the above issue related to the problematic aspect of assessment work (for a review, see Rosa, Coutinho, & Flores, 2016). Examples of the administrative logistics include the submission of grading and feedback, anonymizing and random distribution of assessed tasks, making feedback available, and calculation of marks (Mostert & Snowball, 2013).

Accuracy of PA is often a significant concern. There is substantial evidence showing the high validity of PA (Topping, 1998; Falchikov & Goldfinch, 2000; Tseng & Tsai, 2007). Topping (2009) emphasized three things for the reliability and validity of PA. First, PA is plentiful that offers triangulation to improve the overall reliability and validity. Second, student assessors could produce an equally reliable and valid assessment to that of a teacher when they spend more time to do it. Third, peer feedback is available in high volume with immediacy than teacher feedback, which compensates for a quality disadvantage. However, the forms of PA vary that could have different influences on students’ learning.

2. **Types of Peer Assessment and the Impact on Performance**

PA can be expressed in either a quantitative or a qualitative way. Generally, it has three types: quantitative ratings (rubric), qualitative comments (peer feedback), and a combination of both (hybrid). Regarding quantitative scores, the form of a rubric with indicators to clarify evaluation criteria is often used as a scoring guide to rate the quality of open-response assignments or project works. This approach is relatively reliable and valid when subject matter experts establish the judgment criteria (Ashton & Davies, 2015).

During formative PA, judgments often include qualitative comments labeling ‘peer feedback’ (Gielen et al., 2010). Topping et al. (2000) argued that rich and detailed qualitative feedback information on strengths and weaknesses and suggestions for improvement seems to be most helpful for learning, not merely a quantitative mark or grade. To increase the quality of peer
feedback, it is crucial to (a) provide assessing guidelines through carefully structuring the peer feedback template and (b) train on how to do that in advance (Gielen & De Wever, 2015). Otherwise, inadequate peer feedback quality may not help to improve learning. Additionally, students do not prefer peer feedback assessments due to issues such as unfairness, high workload, and reliability (Wilson, Diao & Huang, 2015).

The hybrid approach attempts to maximize the benefits of PA by combining the rubric with peer feedback. In this way, the hybrid assessment not only could gain a rather reliable and valid evaluation from the rubric approach but also obtain informative and constructive feedback from the peer feedback approach. Although the hybrid approach combines the advantages of the rubric and peer feedback, it may also make the already time-consuming process of PA lengthier, leading to students’ resentment or rejection.

Which types of PA are more effective for learning? Xiao and Lucking (2008) compared the effects of the rubric with the hybrid type on students’ academic writing performance with 232 university students. The results showed that students in the hybrid group demonstrated more considerable improvement in their writing than those in the rubric group. Lu and Law (2012) investigated the effects of online PA in the form of peer grading and peer feedback on 180 high school students’ project performance. The results indicated that peer feedback rather than peer grading was a significant predictor of students’ performance. Yu and Wu (2013) examined the predictive effects of the three types of PA on the quality of produced work with 233 5th-grade students. The regression results indicated that all of the three types of PA significantly predicted the quality of the produced work. Meanwhile, peer feedback explained more variance in students’ performance than did the rubric. Although all types of PA were associated with students’ performance, it remains inconclusive that which is more effective for learning.

3. The Present Study and Hypotheses

We conducted this study to (a) observe the impact of the three types of PA in terms of the rubric, peer feedback, and the hybrid on students’ project performance; and (b) understand students’ perspectives on online PA. Two questions helped to guide this research:

1) Which types of PA in terms of the rubric, peer feedback, and the hybrid were more effective on students’ project performance?

2) How did online PA influence the project performance from students’ perspectives?

This study combined both quantitative and qualitative approaches to collect, analyze, and report findings. We conducted an empirical experiment to answer research question one. The hybrid type that combines the advantages of both the rubric and peer feedback could provide objective rating scores along with critical qualitative feedback to benefit students’ learning. Additionally, previous research showed that peer feedback rather than peer grading was a significant predictor of students’ performance (Lu & Law, 2012). Accordingly, this study tested the following hypotheses:

**Hypothesis 1.** Students in the peer feedback condition would have better project performance
than those in the rubric condition.

**Hypothesis 2.** Students in the hybrid condition would have better project performance than those in the rubric condition.

**Hypothesis 3.** Students in the hybrid condition would have better project performance than those in the peer feedback condition.

To answer research question two, we conducted semi-structured interviews to investigate students’ perspectives on online PA as well as to confirm and explain the above quantitative findings.

## 4. Method

### 4.1 Participants and the Course

The participants were 120 university students in Taiwan. They were taking the course Information and Life, which was a compulsory subject from the General Education Program. Students from all majors could take this course, except for those who were majoring in computer-related fields. This course was an introductory course regarding digital literacy for non-majors, and further explored the possible positive and negative effects of information and communication technologies on human life. The learning activities included three parts: (a) small-group discussions and online quizzes accounted for 40% of the total grade, (b) midterm exam accounted for 30% of the total grade, and (c) small-group final project accounted for 30% of the total grade.

At the beginning of the semester, students were divided into small groups of four to six because they needed to work collaboratively for small-group discussions before midterm, and then for the small-group final project near the end of the semester. This course took place in the computer lab, where the participants normally attended computer-based classes. It was taught in face-to-face mode, but all leaning activities were performed online via the learning management system Moodle.

We adopted the Workshop module of Moodle for online PA. The module allows several functions to make PA a convenient job, such as handling assessments anonymously, submission and random distribution of assignments, allowing self-assessments, grading and providing feedback using an assessment grid, and making assessed work available to the assesees along with any peer feedback.

### 4.2 Project for Peer Assessment

The task for PA was the small-group final project that accounted for 30% of the total grade. The instructor notified students that the rubric (Appendix A) was the evaluation criteria for the quality of their final works. This project included the following sub-tasks requiring students to

1) choose a research topic and identify its relevant representative technologies,

2) collect and analyze relevant information regarding the representative technologies,
3) explore and criticize possible positive and negative effects of those technologies on human life,

4) present and produce the project content in the form of a multimedia presentation,

5) upload the presentation video to the YouTube platform,

6) conduct the first round of online PA,

7) revise and modify the produced work based on the assessments,

8) upload the revised work to the YouTube platform again, and

9) perform the second round of online PA.

4.3 Procedure

When every group has finished the final project near the end of the semester, we conducted two rounds of online PA. All participants were randomly divided into the rubric, peer feedback, and hybrid conditions, with approximately 40 participants (eight groups) in each condition. Then, participants in each condition were assigned randomly three projects for PA by the Workshop module. Self-assessment was prohibited to avoid the problem that students might tend to overrate their own performance. All participants were both assessors and assessees.

There was an orientation before conducting the online PA. During this phase, the instructor (i.e., the third author) demonstrated how to perform PA using the Workshop module. Then, participants were given about 5 min to give a try with the assistance of the instructor and the teaching assistant (i.e., the first author).

Participants were allotted 30 min to complete the first round of PA. After that, each group had an hour to revise and modify the project based on the assessments. They then uploaded the revised work to the YouTube platform again for the second round of PA. Participants were allotted 30 min to complete the second round of PA. In this round, participants in all conditions adopted only the rubric approach to assessing their peers’ revised work. The grade received at this stage was also the final grade of the project, calculated as the average of several peer assessments. Meanwhile, the instructor (the subject expert) manually assessed each group project using the rubric. After the second round of PA, the first author conducted six face-to-face interviews with six volunteers (two from each condition) individually.

4.4 Instruments

All instruments were presented in traditional Chinese, which is the official language of Taiwan. The rubric condition used the instrument as shown in Appendix A. The instructor designed and developed the rubric to meet the learning goal of the final project. The assessment form used by the peer feedback condition asked the assessees to provide feedback in terms of strength, weakness, and suggestions for further improvement to their peers’ projects. The assessment approach for the hybrid condition combined the rubric and peer feedback types of PA.
We prepared an interview protocol (Table 1) to guide the semi-structured interviews. The protocol contains interview questions along with several probing questions to guide the interviewees to talk more about their experiences and perspectives on the online PA activities.

Table 1. Interview protocol

<table>
<thead>
<tr>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tell me about your learning experiences in online PA.</td>
</tr>
<tr>
<td>Probe:</td>
</tr>
<tr>
<td>(a) Describe how you process online PA activities.</td>
</tr>
<tr>
<td>(b) Do you ever encounter difficulties during the assessment process? How do those difficulties influence your assessment task?</td>
</tr>
<tr>
<td>2. Please comments on the online PA activities.</td>
</tr>
<tr>
<td>Probe:</td>
</tr>
<tr>
<td>(a) What types of PA in terms of the rubric, peer feedback, and hybrid do you prefer, and why?</td>
</tr>
<tr>
<td>(b) Provide feedback in terms of strength, weakness, and suggestions for further improvement to the online PA activities.</td>
</tr>
<tr>
<td>3. What are your expectations about the online PA activities?</td>
</tr>
<tr>
<td>Probe:</td>
</tr>
<tr>
<td>(a) Are you satisfied with the online PA activities? If not, why?</td>
</tr>
<tr>
<td>(b) How did PA help to improve your project work?</td>
</tr>
</tbody>
</table>

5. Results

To test the hypotheses, we conducted one-way ANOVAs to compare performance among the three conditions to determine whether there were any significant differences. The dependent variable was the project performance. When there was a significant $F$ test, we further conducted Fisher’s least significant difference (LSD) test as a follow-up analysis. The alpha was set at 0.05 for all statistical tests.

The follow-up interviews were recorded and then transcribed. All interview transcript analysis was consistent with the constant comparative method (Lincoln & Guba, 1985). The analysis took the form of successive iterations involving the procedures based on Lincoln and Guba’s techniques of unitisation and categorization. The iterations were repeated until no new patterns emerged. To ensure the credibility of the findings, we conducted extensive
member checking with the interviewees.

5.1 Project Performance

Table 2 presents the descriptive statistics of the three conditions on project performance assessed by students in the second round of PA and by the instructor, respectively. The ranking of the final grade of project performance evaluated by the students and by the instructor was the same: the hybrid condition was the best, the rubric condition was the second, and the peer feedback condition is the third. Regrading the project performance assessed by students, ANOVA found no significant differences among the three conditions, $F(2, 21) = 1.999, p = .160$. On the other hand, ANOVA found significant differences among the three conditions on the project performance assessed by the instructor, $F(2, 21) = 3.917, p = .036$. LSD showed that there was a significant difference between the peer feedback and the hybrid conditions in project performance ($p = .014$). But no significant difference was found for the comparisons of the peer feedback and the hybrid conditions with the rubric condition ($p = .058$ and $.496$, respectively). Accordingly, the results support Hypothesis 3 but do not support Hypotheses 1 and 2. That is, students in the hybrid condition likely had better project performance than those in the peer feedback condition did. Students in the rubric condition tended to perform equally well as those in both of the hybrid and peer feedback conditions.

Table 2. Descriptive statistics of the three conditions on project performance

<table>
<thead>
<tr>
<th>Types of Assessment</th>
<th>Peer Project</th>
<th>Project Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students’ second-round peer assessment score</td>
<td>Rubric</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Peer Feedback</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24</td>
</tr>
<tr>
<td>Instructor’s final grading score</td>
<td>Rubric</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Peer Feedback</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24</td>
</tr>
</tbody>
</table>
5.2 Students’ Perspectives on Online Peer Assessment Activities

We explored how online PA influenced the project performance from students’ perspectives through the follow-up interviews. The emerging data-driven variables were perceived benefits for learning, professional assessment, acceptance, the online PA system, and types of PA. Those possible confounding variables might affect students’ performance.

5.2.1 Perceived Benefits for Learning

Students acknowledged the benefits of PA for learning included learning by observation, developing critical thinking, and obtaining multiple perspectives from peers besides the instructor. Those benefits could help them improve project performance. Additionally, students’ perceived benefits of PA for learning likely increased their acceptance of and willingness to support and participate in the online PA activities actively.

We received a lot of feedback. I feel that in addition to what I see, I can also hear different opinions and responses. In this way, I have learned a lot. Because everyone is from various majors and has different perspectives, so I think I got many helpful opinions to improve my learning.

5.2.2 Professional Assessment

Professional (or accurate) assessment, which is highly valid and reliable, referred to the assessment made by experts such as professional teachers (Topping, 2009). Regardless of being either assessors or assessees, students were anxious about whether they could make professional assessments. Students identified several variables that might influence professional evaluation, including PA experiences and skills, subject domain knowledge, and being able to make objective judgments.

Generally, teachers rarely apply peer assessment into teaching, so we do not possess PA experiences and skills. We do not know how to do the grading or give comments.

I am from the Department of Art. Maybe I do not know much about information technologies and the Internet, so I cannot make adequate judgments on the presentation content. I can only focus on the proportion of aesthetics, such as presentation layout, the font and typesetting, and so on.

I think there will be problems with students’ assessments. For example, if you are my friend, I will score you higher. If I do not like you, I will give you a relatively low score.

5.2.3 Acceptance

Variables associated with students’ acceptance of the online PA activities included workload, the proportion of the score for the PA activities, perceived benefits for learning, and reliability. Some students indicated that if the PA activities generated excessive workload, their acceptance of PA would become low. Fortunately, they further expressed that the online PA system offered several functions to make the assessment work convenient that compensated for the extra workload brought by PA. Some students also indicated that a relatively high proportion of the score for the PA activities would motivate them to participate actively in the
activities. In this regard, peers would do their best for evaluation. Students’ acceptance of PA also tended to increase when they perceived the benefits of PA for learning, notably when the rich and adequate feedback guided them to improve their performance. Students further indicated that they appreciated the reliability of PA because the assessment was a result of the average of several peer assessments.

If the proportion of the score for the assessment work is relatively low, some students will not take it seriously and will make a careless assessment.

The project should be evaluated by more students; that is, not one or two, maybe ten or eight. Multiple assessments would be more reliable.

5.2.4 The Online Peer Assessment System

Using the Moodle Workshop tool as the online PA system, students considered it convenient and immediate with the benefit of anonymity. Those advantages made the activity process smooth and reduced the occurrence of disputes. This technology-assisted system, therefore, likely increased students’ acceptance of PA. However, some students indicated that although anonymity could make the assessment work fairer, some might take this chance to give negative comments that were very discouraging.

We do not have to write so many words, and it is convenient for everyone. Everyone can receive the evaluation immediately.

In addition to being effective, there is greater anonymity. However, due to anonymity, people may give negative comments, may pick bones in eggs, and say something hurtful.

5.2.5 Types of Peer Assessment

Most students preferred the hybrid approach. Some students indicated that although the rubric offered clear criteria to guide them to do the assessment work, only the grading scales did not allow knowing the specific suggestions for improvement. On the other hand, some students reflected that specific assessment criteria let them know precisely how to do assessments and realize what standard they had achieved. Taken together, the hybrid approach with the advantages of both the rubric and peer feedback met the needs of students.

I prefer the hybrid approach. The assessment is more reliable if there are qualitative comments to support the scores I got.

The scores only allow me to know whether I did right or wrong, but I do not know how to improve my work.

I feel that if there are only comments for the project, it seems something missing. I still hope that there will be a scale to let me know what standard I have reached.

6. Discussion and Conclusion

The study aimed (a) to investigate which types of PA in terms of the rubric, peer feedback, and the hybrid were more effective on students’ project performance, and (b) to explore how online PA influenced the project performance from students’ perspectives. The findings
implied that the hybrid approach was more effective for learners with the inexperience of PA and a basic level of domain knowledge. This particular group of learners indicated that clear evaluation criteria as a scoring guide along with constructive feedback as specific suggestions for performance improvement could best satisfy their learning needs. The findings further suggested that besides types of PA, other possible confounding variables might affect project performance, including perceived learning benefits, professional assessment, acceptance, and the online PA system.

Faculty-student marks comparisons raised the issues of reliability and validity. Generally, the validity (or accuracy) of PA was achieved when there is a correspondence between student peer-generated marks and instructor-generated marks (expert assessments). In this study, the consistent rank ordering of the project performance assessments made by peers and the instructor likely showed the validity of PA. Regarding reliability, a reasonable degree of agreement among assessors could be considered reliable. However, a relatively large standard deviation from peers reflected a large amount of variation among peers’ assessments. Literature indicated that students in advanced courses tended to be the more reliable assessors than those on introductory courses (Falchikov & Goldfinch, 2000; Topping, 2009). The course in this study was an introductory level, and students were from non-majors, so students’ diverse domain knowledge could explain the variation. Besides the level of the course and students’ different domain knowledge, the follow-up interviews further explored other possible confounding variables that influenced their assessments.

Which types of PA were more effective on students’ project performance? ANOVA found no significant differences among the three conditions on the project performance assessed by the students, whereas significant differences were found when evaluated by the instructor. The inconclusive findings suggested that there might be some confounding variables. The qualitative results revealed that besides types of PA, those confounding variables included perceived learning benefits, professional assessment, acceptance, and the online PA system. The findings also revealed their interrelatedness. First, the types of PA were associated with students’ professional assessment. Second, students’ perceived learning benefits of PA, professional assessment, and the online PA system were variables related to their acceptance of PA. Finally, we proposed that professional assessment and acceptance might be two moderators between types of PA and the project performance.

Types of PA were associated with students’ professional assessment, which might be a moderator between types of PA and the project performance. Students indicated that they could not evaluate peers’ work professionally due to the lack of PA experiences and skills and inadequate domain knowledge. In this case, the well-designed assessment criteria as a scaffold to guide them to do the evaluation was crucial. The peer feedback type without specific evaluation criteria was not suitable for the learners in this study. As students indicated that they preferred the hybrid approach because they needed the clear criteria offered by the rubric and the specific suggestions offered by peer feedback. The above observations explain the quantitative finding that students in the hybrid condition likely had better project performance than those in the peer feedback condition did.
Students’ acceptance of PA might be a moderator between types of PA and the project performance. Acceptance of PA could engage them more in the online PA activities that helped to improve performance. Participants identified the perceived learning benefits of PA, professional assessment, and the online PA system related to their acceptance of PA, as described in the sections above. In sum, the two moderators in terms of professional assessment and students’ acceptance of PA might affect the strength of the relationship between types of PA and the project performance. Further research is necessary to confirm the observations.

There were limitations to this study. First, given that our qualitative findings were based on a small sample of student volunteers, the results thus need to be interpreted with caution. However, further research is necessary to ensure the observations. Second, we used the rubric as evaluation criteria to determine the final grade of all projects, and that might disadvantage the peer feedback group, thereby possibly confounding the results. Third, this study focused on a specific group of undergraduate students taking an introductory course from general education, and the produced work for assessments was a multimedia presentation project, which needs to be taken into consideration to make transferability judgments. Last, further investigation should take into account learner characteristics such as PA experiences and skills, subject domain knowledge, and being able to make objective judgments to triangulate possible variables that might influence professional assessment.

References


**Appendix A**

**Assessment criteria (100%)**

1. **Basic requirements.** Does the submitted presentation fulfill the following basic requirements: (a) a presentation cover indicating the project topic, the course name, and group members information; (b) content contains at least eight slides; (c) fair use of all publications and media resources, and (d) contributions of each member? (20%)
   - □ poor (satisfy only two requirements)
   - □ limited (satisfy three requirements)
   - □ good (satisfy all requirements but lack related details)
   - □ excellent (satisfy all requirements with adequate details)

2. **Presentation content organization.** Is the presentation content clearly and well organized into the four main points of (a) introduction, (b) pro arguments, (c) con arguments, and (d) reflections and conclusion? (20%)
   - □ poor (information is not organized, and lacks main points and related details)
3. **Slide communication quality.** Does each slide (a) express only one main idea and contain only a title, (b) simplify and limit the number of words by using key phrases and including only essential information, (c) use visual support to reinforce and complement the message, and (d) avoid overuse of special effects such as sounds and animation? (20%)

- □ poor (most slides express more than one idea and contain too much unorganized information)
- □ limited (few slides express more than one idea and contain too much unorganized information)
- □ good (every slide express only one idea but with unclear or inadequate information)
- □ excellent (every slide express only one idea and with adequate, clear, and concise descriptions and graphics)

4. **Slide information quality.** Does the presentation make the facts insightful and memorable by satisfying the following? (a) messages are simple and clear, and flow in a logical sequence, (b) draw attention or imply importance when necessary using colors and fonts, and (c) offer reflections and criticism, not just the facts. (20%)

- □ poor (satisfy only one requirement)
- □ limited (satisfy only two requirements)
- □ good (satisfy all requirements but with inadequate information regarding reflections and criticism)
- □ excellent (satisfy all requirements with adequate information regarding reflections and criticism)

5. **Presentation content.** Can the presentation content improve the understanding and attitude to the positive and negative impact of modern information technologies and the correct use of them through containing the following? (a) a well-defined research topic, (b) introduction on its relevant representative technologies, (c) exploration and criticism of possible positive and negative effects of those technologies on human life? (20%)

- □ poor (satisfy only one requirement)
- □ limited (satisfy only two requirements)
- □ good (satisfy all requirements but with inadequate details)
- □ excellent (satisfy all requirements with adequate details)
Copyright Disclaimer

Copyright reserved by the authors.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).