Int J Syst Assur Eng Manag (August 2020) 11(4):780–784
https://doi.org/10.1007/s13198-019-00837-2

ORIGINAL ARTICLE

Analysis and reflection on peer assessment results based on short play of game theory

Lifang Qiao[1] · Zhengyou Wang[1] · Sanqing Bao[1] · Yanhui Xia[2]

Received: 5 April 2019 / Revised: 11 July 2019 / Published online: 29 July 2019

Abstract Peer assessment is an important part to evaluate college students' learning effect and ability in blended learning. Peer assessment guided by teachers encourages students to engage in classroom and leads to the learner-centered classroom. Peer assessment studies in higher education mainly focus on the students' attitude to peer assessment and the effectiveness of peer assessment. However, little exploration has been made on how to deal with and prevent the bad results of peer assessment, which has an important impact on improving the quality of peer assessment. By Game Principles Course, this research puts forward the idea of building a peer assessment loop to improve the quality of peer assessment and prevent the bad results of peer assessment.

Keywords Peer assessment · Reliability analysis · Validity analysis

✉ Zhengyou Wang
zhengyouwang@stdu.edu.cn

✉ Yanhui Xia
825172525@qq.com

Lifang Qiao
18712762311@163.com

Sanqing Bao
baosanqing@163.com

[1] School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, China

[2] School of Economy and Management, Shijiazhuang Tiedao University, Shijiazhuang, China

1 Introduction

Peer assessment has become a part of students' evaluation in higher education. Peer assessment is a learning activity based on the theory of social constructivism. It means the learners evaluate one another's learning tasks, and peer assessment can be used in various learning activities, such as written work, oral report, role performance. With the support of technology, various kinds of teaching combined with peer assessment have become a common practice (Wang 2017).

There are more studies on peer assessment in foreign countries, and peer assessment has been widely used in the subjects of natural science, social science business, medicine, two language acquisition and engineering technology (Reinholz 2016). When evaluating others' work, students can create useful feedback to the assessed and may give some insights on how to improve their work (Reinholz 2016). Previous research shows that teachers and students' conceptions about the purpose of assessment largely influence its implementation (Rotsaert et al. 2017). The quality of peer assessment is also a response educational value and learner acceptance.

There are three kinds of research on peer assessment by foreign scholars: conceptual research, quantitative research, and qualitative research. The concept study explores the peer assessment on students' reflective ability and meta-cognition ability. For qualitative research, the effect of peer assessment in the specific environment and managing peer assessment were emphasized. And quantitative study discusses the reliability and validity of peer assessment and the influence of different factors (Luo et al. 2017). Among them, there are many quantitative studies on peer assessment.

In China, the research on peer assessment mainly focuses on primary and secondary school classes, including peer assessment in writing or reading (Chen 2017; Zhou 2018), mathematics teaching (Gong 2016), chemical experiment and English teaching (He 2016).

At present, peer assessment in higher education in China mainly focuses on foreign language and medical teaching. Besides, most studies focus on the implementation and effect of peer assessment. For example, the application of peer assessment method in opening nursing laboratory found the learning effect of peer assessment is better than the traditional practice, and it can effectively improve the students' initiative and enthusiasm to increase their self-confidence (Chen et al. 2017).

In teaching computer experiment, peer assessment is introduced, which can not only stimulate students' interest in learning but also provide a reference for teachers to evaluate the results of the experiment. At the same time, the peer assessment system of students' experimental homework in the computer experiment class is also built (Yuan 2017). There is little research on the results given by students from the perspective of peer assessment.

As a teaching assistant, the author found the results of peer assessment were different. There existed unreasonable scores and friend mark. To further understand the causes for the deviation of the results and put forward feasible solutions, the author conducts empirical research and reflection based on observing the students' classroom performance and analyzing the results of peer assessment.

2 Research questions

The purpose of this study is to explore the reasons for the poor quality of peer assessment in short play. This can be studied through the following research questions:

(1) What affect peer assessment quality?
(2) How to improve the quality of peer assessment in teaching?

3 Methodology

3.1 Participants

The subjects were sophomores, who were familiar with peer assessment and had evaluation ability. Thirty five students took part in peer assessment, including 33-course learners, 21 boys and 12 girls. To ensure the accuracy and effectiveness of peer assessment, teachers arranged 2 graduate students to join in peer assessment.

3.2 Procedure

At the beginning of the course, teachers complete the grouping of learners according to students' knowledge and ability self-assessment. Peer assessment is conducted at the end of the semester. Before peer assessment, learners have completed learning knowledge and skills related to game theory courses. For the first time, learners formally conduct peer assessment, teachers give a brief explanation, and set up scoring standards and scoring tables to support peer review and promote students' deep learning in the proximal development zone. Through collaborative learning and social interaction, learners understand the criteria and significance of peer assessment. Thus, learners can score as reasonably as possible. Each group member freely chooses the theme related to the game and divides the work reasonably. A 5–10 min script is designed and performed using game theory knowledge. During the performance, under guiding of scoring rules, members of each group scored the overall performance and individual performance of other groups by using the peer assessment form.

3.3 Material

The evaluation table used in this paper is the index structure education evaluation table compiled by the instructor according to the teaching experience and teaching arrangement and uses multiple comparison method to decide the weight of each index. Scoring items include theme related, short play creativity, content design, number of scenes, script writing, short play effect, team collaboration, and actor performance. When grading the actors, the raters need to score each member's performance of the group. The final score of the participant is to evaluate all the scores of each item in the table.

4 Data analysis and results

After peer assessment, the author performed a summary of the score results. It was found that individual students did not score properly according to the criteria during the scoring. We found that scores of every item are not high. There are also the same scores for different students. Besides, some scores are not based on the effect of short-drama performance. For students of different levels, graduate students and course learners have large deviations in scores. Graduate students' scores are contrary to those given by classroom participants. To understand the causes affecting the quality of peer assessment, the author analyzed the independent T test, reliability, and validity of the statistical data.

Although the average score of course learners is higher than that of postgraduates, the independent sample t test of SPSS shows that the t value is 1.1394. Because t value is less than 2.0, there is no significant difference between the two groups. Therefore, the subjective factors that appear in peer assessment may be the main reason for the difference in performance (Carvalho 2013). Viewing from designing peer assessment, the reason for the difference may be that the peer assessment process is not detailed enough and the peer assessment materials need to be improved. In addition, learners' evaluation ability may also be the cause of differences in peer assessment.

4.1 Reliability analysis

First, the reliability of peer assessment results was analyzed by Cronbach's alpha of SPSS17.0 statistical software. This process involves group collaborative learning under blended learning, and the scoring items include group collaboration and group members' performance. We want to understand the consistency of scoring items in measuring individual dimensions. Therefore, the author analyzed the reliability coefficient of peer assessment according to the group. The greater the coefficient of reliability, the more reliable the evaluation table is. From DeVillis work (1991), we know the Cronbach's α coefficient needs to be at least higher than 0.6. The reliability analysis of the eight sets of peer assessment results is shown in Table 1.

From Table 1, it can be seen that the consistency of the items measured by the eight groups of peer assessment is above 0.76. The results showed that the items of the scoring table had high consistency and the scoring table was credible. When the results of the postgraduate students were included in the overall scores, the reliability of the evaluation form was found to be slightly improved. Therefore, it is found the evaluation level of the evaluator may affect the reliability of the evaluation form. To further confirm the differences between graduate students and their classmates in peer assessment, we analyzed the skewness coefficient of the scoring data. The results show that the skewness of peer evaluation was 0.843 and the skewness of postgraduate evaluation was 0.443. Moreover, the graduate students' scores are more in line with normal distribution. According to the results of data analysis and teaching practice, it may be due to the higher evaluation ability of graduate students. In order to solve this problem, it is necessary to cultivate students' peer evaluation ability. Students should join the curriculum evaluation system. Because peer assessment and students' learning benefits are related and students can feel the educational value of peer assessment.

The total item statistics in the Cronbach's Alpha analysis data show the evaluators have great differences on topics related, the number of scenarios, the writing of the script, the team collaboration and the individual performance. Possible reasons include that these scoring rules are not clear enough, and the evaluators have doubts about how to accurately evaluate them. To improve the reliability of the evaluation form, teachers can modify and improve the evaluation form according to the analysis results.

4.2 Validity analysis

Next, we use factor analysis for further analysis. KMO (Kaiser–Meyer–Olkin) test and Bartley ball test were performed before factor analysis. Factor analysis can only be carried out when the KMO test coefficient is more than 0.5 and the P value is less than 0.05. The range of the KMO test is 0–1. The closer the KMO value is to 1, the more suitable the data is for factor analysis. The results of each group's validity analysis are shown in Table 2, and factor analysis can be carried out.

Completing factor analysis, the linear regression and the ordered regression method were used to analyze the important causes affecting the quality and the variables of the peer assessment results. Linear regression is used to

Table 1 The reliability analysis of the eight sets of peer assessment results

Group	Peer assessment of Cronbach's alpha values	Cronbach's alpha values of postgraduate	Number of items
Group 1	0.781	0.779	8
Group 2	0.775	0.764	8
Group 3	0.835	0.827	8
Group 4	0.765	0.780	8
Group 5	0.829	0.827	8
Group 6	0.746	0.754	8
Group 7	0.781	0.795	8
Group 8	0.858	0.864	8
Average	0.796	0.802	

Table 2 The results of each group's validity analysis

Group	Peer assessment of Kaiser–Meyer–Olkin	Kaiser–Meyer–Olkin of postgraduate	Sig
Group 1	0.605	0.621	0.000
Group 2	0.523	0.551	0.000
Group 3	0.704	0.714	0.000
Group 4	0.620	0.652	0.000
Group 5	0.782	0.791	0.000
Group 6	0.556	0.66	0.000
Group 7	0.455	0.464	0.000
Group 8	0.793	0.817	0.000

discover the quantitative relationship of the interdependence between two or more variables. Multiple linear regression is used in the study. Through linear regression analysis, it is found that the number of scenes and the creativity of short plays affect the theme creation and script creation. Next, team collaboration had the greatest influence on the number of scenes and the creativity of the short play. But content design and short sentence effect had no significant impact on the results. Therefore, the most important factor affecting the score is teamwork. The content design of the evaluation table and the score of the short play effect are not effective.

In each group, the Ordinal regression analysis found the coefficient was significant only FAC2_1, that is, the most significant factors affecting the results were the number of scenes and the creativity of the short play.

Through analysis, we know the quality of peer assessment needs to be improved. When assessing others' work, the evaluators did not carefully evaluate the content design of the script and did not objectively score each group' performance according to the scoring rules. Besides, scorers are not objective and impartial enough in scoring. Therefore, the evaluation table can't reflect the impact of content design and short play effect through peer assessment results. This also suggests the scoring rules should be clearer and more specific.

5 Conclusion and pedagogical implications

The purpose of this paper is to discuss various factors affecting the results of peer assessment in the course of Game Principles. Based on the results of data analysis and teaching practice, we will put forward countermeasures and ideas to improve the quality of peer assessment from the perspective of students. According to the results of data analysis, the evaluation table is credible, but its validity needs to be improved, and there are great differences between undergraduate and postgraduate scores. Due to the limitation of the number of courses, it is necessary to keep track of the impact of peer assessment, or to increase the number of participants.

Through analysis, it is possible to establish an ecosystem of peer assessment in higher education. In addition, it can also provide students with the right to choose evaluation indicators, establish evaluation tables and allocate weight. To further improve the quality of peer assessment, we should establish peer assessment accountability system, give learners the right to appeal, and enhance learners' responsibility to participate in peer assessment. The details are as follows:

(1) Before the assessment, students join in screening indicators for peer assessment. Teachers can give course learning objectives and ability needs through online group chat before class. By understanding the learning goals, the students give the most important scoring items for a task. Priority ranking of measurement items is collected through network communication. Then a preliminary peer rating table was formed. Teachers asked questions on the problem of peer assessment and guided students to improve the score list, and finally formed peer assessment lists. In this process, students can not only have a clear understanding of the curriculum objectives, but also improve their evaluation ability. In addition, this measure can reduce the negative effects of friend tags and inherent tags.
(2) Through mobile phones, we can get and browse peer assessment data in real time. Also, the students can also check through the mobile phone. Learners can ask questions based on unreasonable data. For teachers, they can check the objectivity and rationality of grading, and send effective feedback to students through the platform.
(3) After the evaluation, the learners can promote deep learning by checking the electronic mutual evaluation files and the mutual evaluation results given by different tasks.
(4) To set up an ecological system of peer assessment, the low frequency peer review can promote students'

learning and teachers' teaching in a cyclic way. Then from the macro view, it guides students and teachers to start from the teaching details and the policy research of the whole teaching team.

To sum up, there are some common problems and differences in peer assessment. For small sample, how to further improve the quality of peer assessment still needs to be explored and studied.

To further understand how to improve the quality of peer assessment in higher education, the author will further expand the sample and constantly improve the peer assessment form. Considering the number of samples and the period of data collection, we will set up a control group to find out the differences in calculating peer assessment data and designing peer assessment tables in different courses. On this basis, students, teachers, colleges and schools are integrated into the peer-to-peer evaluation system and promote each other. How to put forward a complete process or framework of peer review and apply it in different courses is also a research problem that needs our thinking. In addition, we developed an AR application inspired by the game theory course. A number of kindergartens have introduced this AR application, which provides a new opportunity for peer assessment. Next, we can develop peer assessment for kindergarten art teaching and carry out further research.

Acknowledgements This work is supported in part by Higher Education Reform and Research Projects in Hebei Province (No. 2016GJJG117), the project of Hebei Postgraduate Demonstration Course (No. KCJSX2018068), and the projects of Hebei Social Science Fund (Nos. HB17JY069, HB19JY017).

References

Carvalho A (2013) Students' perceptions of fairness in peer assessment: evidence from a problem-based learning course. Teach High Educ 18(5):491–505

Chen T (2017) Implementing students' mutual evaluation in Chinese teaching. Basic Educ Res (18):18–18

Chen D, Wang J, Wei Y, Li X (2017) Application of student assessment in the opening of nursing laboratory. High Med Educ China 12:62–63

DeVillis RF (1991) Scale development: theory and applications. Sage, Newbury Park, CA

Gong X (2016) The practice of student assessment in the rural junior high school mathematics classroom. Ludong University

He H (2016) Student feedback mechanism based on teacher feedback in English writing. Engl Plaza 04:88–89

Luo H, Zuo M, Anthony R (2017) An empirical study on the effect of large-scale open online learning on student interaction assessment. Open Educ Res 23(01):75–83

Reinholz D (2016) The assessment cycle: a model for learning through peer assessment. Assess Eval High Educ 41(2):301–315

Rotsaert T, Panadero E, Estrada E, Schellens T (2017) How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. Stud Educ Eval 53:29–40

Wang F (2017) An empirical study on the effect of student interaction assessment. J High Educ Res 40(04):86–90

Yuan P (2017) Research on the teaching mode of student assessment system in computer experimental classroom. Technol Inf 15(35):119–120

Zhou B (2018) A review of peer assessment in english learning in China [J]. English Teachers 18(17):6–9