An educational research course facilitated by online peer assessment: ...
Chen, Ying-Chih;Tsai, Chin-Chung
Innovations in Education and Teaching International; Feb 2009; 46, 1; ProQuest Central
pg. 105

Routledge
Taylor & Francis Group

An educational research course facilitated by online peer assessment

Ying-Chih Chen[a] and Chin-Chung Tsai[b]*

[a]*Department of Teaching and Learning, University of Iowa, Iowa City, Iowa, USA;*
[b]*Graduate School of Technological and Vocational Education, National Taiwan University of Science and Technology, Taipei, Taiwan*

This study implemented an online peer assessment activity to help 52 Taiwanese in-service science and mathematics teachers to develop research proposals in an educational research course. The participants were divided into 16 groups, and each group was required to submit a proposal via three rounds of online peer-assessment activity. This study found that, through the online peer-assessment activity, the in-service teachers could enhance their proposals at the initial stage of the activity, but there was almost no substantial improvement at the later stage. In addition, the amount of feedback they obtained was positively related to the performance from the first round to the second round; however, few significant relationships were found from the second to the third round. This study also showed that the process of peer assessment could be used to clarify assessment criteria and that sufficient support from the instructor can maintain participants' attitudes towards this activity, which were relatively optimistic throughout.

Keywords: peer assessment; online; Internet; in-service teachers; attitude

Introduction

Peer assessment (PA) activity has many useful features when used within educational practice (Falchikov & Goldfinch, 2000). Through a PA learning activity, not only learners may be enhanced in thinking skills, but also instructors can gain a better understanding about learners' conceptual development to make fairer judgements towards their work (Topping, 1998). Researchers and educators have made great efforts to enhance the merit of PA and it has been widely used in various fields, especially in higher education (van den Berg, Admiraal, & Pilot, 2006; Wen & Tsai, 2008).

A growing number of educators have tried to utilise Internet-based systems to facilitate the process of PA. These have been shown to attain successful learning outcomes (Prins, Sluijsmans, Kirschner, & Strijbos, 2005; Tseng & Tsai, 2007). Internet-based (or online PA) overcomes the obstacle of time and space for learners, providing a timely avenue for learners to reflect on their work. Davies (2000) agrees that with the assistance of the Internet, the use of anonymity with respect to participants' marking and identifying work along the process of PA has been a success.

Moreover, online PA has been used both as an assessment tool and as a learning tool. Online PA could produce electronic portfolios for peer learners and teachers, which can aid student learning and teacher assessment (Liu, Lin, & Yuan, 2001). The

*Corresponding author. Email: cctsai@mail.ntust.edu.tw

electronic portfolio includes assessment and suggestions from matched companions' reflection on their own work and their revisions. Educators have also claimed that online PA provides students with active learning environments to promote critical thinking, planning, monitoring and regulation in computer-supported collaborative learning (Liu & Tsai, 2005). An online database has the capacity to systematically collect student development of work, recording each student's progression and ways of completing the learning task (Liu et al., 2001).

The quality of peer feedback plays an important role in student learning involved in the PA process (Davies, 2000, 2006). However, Topping (1998) states that feedback is helpful only when the recipient executes it, and different types of feedback may produce different learning effects by different learners. Tseng and Tsai (2007), by analysing 184 high school students' peer feedback stored in a PA electronic portfolio, found that the peer feedback with the attributes of reinforcement, encouragement, and friendly suggestion was quite effective for subsequent improvement.

However, some concerns related to PA are often identified by educational researchers, such as the validity and fairness of PA techniques and peer feedback, PA criteria, and instructors' as well as participants' attitudes toward PA. For example, the participants may not have a will to accept any responsibility for evaluating their peers, particularly at an initial stage (Falchikov, 1995). Magin (2001) claims that the existence of bias in peer scores comes from the socially interpersonal friendship among peer students. Although this bias may hinder the validity of PA, the use of anonymous online PA can improve this. Also, the clarification of PA criteria is essential (Orsmond, Merry, & Reiling, 1996; Wen & Tsai, 2008). Wen and Tsai (2008), exploring 37 inservice science and mathematics teachers' attitudes with three rounds of online PA, concluded that the participants' attitudes toward online PA decreased significantly after online PA intervention, probably due to the unclear PA criteria. Wen and Tsai (2008) further asserted that more clarified criteria and additional support from instructors should be offered with much more practice before the PA activity.

This study, first, examines the effectiveness of online PA implemented with a group of inservice teachers on a graduate-level course of educational research. Then, how the peer feedback that they had obtained related to their progression derived from the PA process is explored. Also, this study investigates whether providing more practice and clarified criteria about PA might foster the participants' attitudes toward PA.

Method

Participants

This study was conducted at a research university in north Taiwan, including 52 inservice mathematics and science teachers (30 male and 22 female), who came from elementary, junior, and high schools in the vicinity. All of the participants, interested in Internet-based learning, enrolled in an e-learning master's degree programme. These teachers had teaching experiences ranging from two to 20 years. All of them took a required course of about four months, called 'Educational Research Methods'.

Online PA activity in the course

The objective of the course 'Educational Research Methods' was to help the participants acquire knowledge about how to design and conduct educational research.
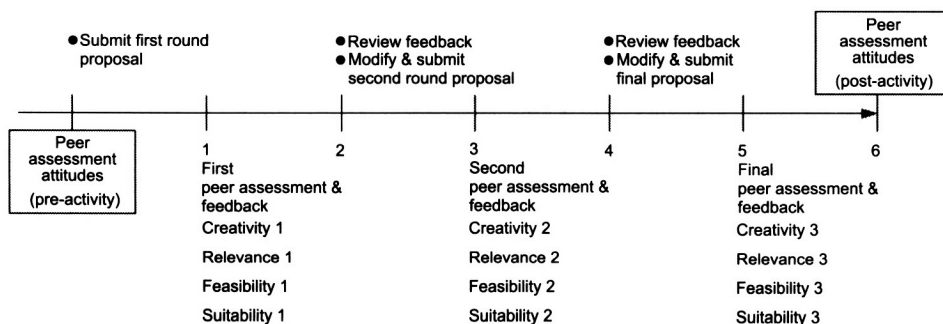
Figure 1.   Three-round online PA procedure across six weeks.

These science and mathematics teachers, three or four as a group, were required to present a complete educational proposal at the end of the course, which was expected to be gradually refined through a three-round online PA system. The required work would be similar to a master's thesis proposal. For example, some teachers were interested in using a game-based e-learning environment to enhance children's mathematical skills, while others wanted to investigate the influence of computer-based instruction on students' alternative conceptions in science.

These 52 teachers were divided into 16 groups by their interest, and each group contained three to four participants. Each group first submitted the proposal after two months of instruction on the course. All proposals experienced three-round online PA treatment by anonymity, shown in Figure 1. Each proposal needs to be refined and submitted three times. Each participant also should review the proposals assigned by the online PA. The participant reviewed the same proposals assigned across different rounds of PA. The online PA took about six weeks. The participants submitted their projects on a group basis, while they reviewed peers' work on an individual basis, exactly the way implemented by Wen and Tsai (2008).

Peer scores and expert scores

For each round of PA, every group's proposal was quantitatively rated on the following four dimensions: creativity, relevance, feasibility, and suitability. The score given by the participants ranged from 1 to 7 (1 = lowest score, 7 = highest score):

(1) Creativity: the extent to which the proposal shows the participants' originality about educational research.
(2) Relevance: the extent to which the proposal is related to significant educational issues.
(3) Feasibility: the extent to which the educational research proposal could be actually conducted.
(4) Suitability: the extent to which the proposal uses appropriate research methods.

Every proposal was rated by peer reviewers in these four dimensions for three rounds, shown in Figure 1, with the outcome variables from 'Creativity 1' to 'Suitability 3'.

In the same way, the course instructor gave evaluations for each group's proposal, served as expert scores. In addition to the quantitative evaluations of the four dimensions, peers were asked to provide qualitatively detailed comments or feedback to each research project assigned for review. Therefore, each group could modify their proposals by referring to the peer feedback.

As stated earlier, this study attempted to provide more clarified evaluation criteria for the participants when they were involved in the PA process. Therefore, before the actual implementation of PA, the instructor explained the assessment dimensions clearly and in great detail. Also, sample research projects and peer comments were illustrated as an aid to clarify the criteria. In addition, over the six weeks of PA, the instructor would spare some time in each class to discuss and solve related problems which the participants experienced in the online PA. By offering more clarified criteria and adequate support from the instructor, the online PA conducted in this study aimed to facilitate participants' favourable attitudes toward PA.

Content analysis of peer feedback

In order to deeply explore the role of peer feedback, this study conducted a series of content analyses for categorising peer (qualitative) feedback. The content analysis of peer feedback was based on Tsai and Liang's (in press) framework, classifying each feedback into one of seven categories listed in Table 1. The definition of feedback for each category with examples is illustrated in Table 1. The feedback of 'Supporting' and 'Emotional Response' could be viewed as more affective-oriented feedback, the categories of 'Direct Correction', 'Personal Advice' and 'Guiding' were perceived as cognitive comments, while the final two categories, that is, 'Evaluating and Planning' and 'Regulating and Reflecting', were conceptualised as metacognitive-aligned comments by peers.

During the PA activity, there were three rounds of feedback. The authors categorised all of the feedback and summed up the frequencies (number) in each feedback category of each round for each group proposal. In this way, the study explored the correlation between the feedback each group obtained from their peers and its subsequent performance in the research proposal.

Instrument probing students' attitude towards PA

To examine the in-service teachers' attitude towards PA, the study used the instrument developed by Wen and Tsai (2006), a five-point Likert scale questionnaire (1 = strongly disagree, 5 = strongly agree). Two subscales in this instrument were used in this study: (1) Positive Attitude Subscale (PAS), and (2) Online Attitude Subscale (OAS). The PAS enquired the participants' perceptions of the helpfulness of PA *in general*, such as the item that 'PA activities increase the interaction between my classmates and me'. The OAS explored the participants' agreement toward the benefits of using *online* PA in particular, such as the item that 'Online PA activities can be time-saving'. The reliability coefficients of the PAS and OAS subscales were around 0.80 based on the responses of this study, displaying acceptable internal consistency for assessing the participants' attitudes. This questionnaire was administered to each participant both before and after the PA activity.

Table 1.   The framework of categorising peer feedback.

Category	Definition	Example
Supporting	Statements referring to clear support and praise (i.e. showing praise, and supporting one's ideas).	• The research purpose makes a clear statement. • The direction of this proposal is good.
Emotional Response	Simply express one's feeling about the work.	• I don't like this proposal.
Direct Correction	Focus on the correctness of work.	• The format of reference does not conform to the APA Style Manual. • The questionnaire contains some typing errors.
Personal Advice	Offering general advice or personal perspectives without providing solid evidence or concrete directions.	• I can't understand the difference between yours and prior research. • I think you can enrich the literature review.
Guiding	Questioning for direction, or introducing concepts, contents or frameworks.	• The number of participants in your study should increase to 500. • The number of items in the questionnaire is 84 in total. It is too many for primary school students.
Evaluating and Planning	Evaluating: assessment, appraisal or verification of one's knowledge and skills, and the efficacy of a chosen strategy. Planning: selecting, predicting, and ordering an action or strategy necessary to the accomplishment of an action.	• You can see if the items in the questionnaire are too many for sixth grade students. I suggest you can cut down some of the items. • I would suggest you can evaluate the validity and reliability of the Likert-scale instrument you modified – otherwise, it would be not feasible.
Regulating and Reflecting	Regulating: setting up, maintenance and supervision of the overall cognitive task. Reflecting: challenging learners to think deeply and stimulating constructive reflection by learners as they conjecture, probe and contemplate.	• Do you really think that thinking styles correlate closely with the spatial ability? Besides, it is too difficult to analyse the relationship due to the complex factors existing in thinking style. • Maybe you can add some open-ended questions to inquire about students' opinions and the experimental procedures with which they deal. That would see more clearly whether the students have creativity.

Table 2.   The correlation between expert and peer scores on every outcome variable ($n = 16$).

	Creativity	Relevance	Feasibility	Suitability
First round	0.50*	0.73**	0.68**	0.75**
Second round	0.37	0.57*	0.45	0.52*
Third round	0.17	0.44	0.47	0.48

Note: *$p < 0.05$, **$p < 0.01$.

Results

The correlation between the expert and peer scores

For PA study, one fundamental issue to be investigated is the validity of peer evaluations. Table 2 displays the Pearson correlation coefficients between the scores marked by an expert (the course instructor) and those by peers across three rounds. The results indicated that the scores marked by peers were highly correlated with the expert scores on the first round ($r = 0.50$–0.75, $p < 0.05$), but not significant at the 0.05 level on each dimension of the third round (perhaps due to the small sample size, 16 groups). For the outcome variable 'creativity' of the third round, the correlation coefficient was especially low ($r = 0.17$). It is possible that the marking of creativity may involve much more subjective opinions. However, the correlation coefficients, in most cases, had almost approached moderate to high correlation on every outcome variable. These findings suggested that the peers' scores, in general, were consistent with the experts' evaluations, showing a certain degree of their validity.

Effect of PA on the development of proposals

The participants' proposals were marked by the range from 1 to 7 points on four dimensions of creativity, relevance, feasibility, and suitability. The three-round scores were collected and compared in Table 3 for peer scores and Table 4 for expert scores. According to Table 3, the average scores in the first round of online PA as evaluated

Table 3.   The four-dimension scores on three-round PA activity, rated by peers.

Variable	Mean	SD	Range	Paired t-test
Creativity 1	4.75	0.57	4.00–5.80	–
Creativity 2	4.97	0.48	4.30–5.73	–
Creativity 3	5.08	0.42	4.50–6.00	3>1 (−2.81**)
Relevance 1	5.29	0.66	4.00–6.11	–
Relevance 2	5.34	0.42	4.60–5.88	–
Relevance 3	5.36	0.44	4.57–6.00	–
Feasibility 1	5.08	0.70	3.67–6.00	–
Feasibility 2	5.25	0.51	4.50–6.00	–
Feasibility 3	5.12	0.49	4.40–5.80	–
Suitability 1	4.74	0.70	3.20–5.88	–
Suitability 2	5.00	0.52	4.30–6.00	–
Suitability 3	4.97	0.40	4.25–5.80	–

Note: **$p < 0.01$.

Table 4. The four-dimension scores on three-round PA activity, rated by expert.

Variable	Mean	SD	Range	Paired *t*-test
Creativity 1	4.56	0.73	3.00–5.50	2>1 (−2.44*)
Creativity 2	4.88	0.59	3.50–6.00	–
Creativity 3	5.00	0.52	4.00–6.00	3>1 (−3.22**)
Relevance 1	5.22	0.60	4.00–6.00	–
Relevance 2	5.38	0.47	4.50–6.00	–
Relevance 3	5.43	0.44	5.00–6.50	–
Feasibility 1	4.91	0.78	3.00–6.00	2>1 (−2.18*)
Feasibility 2	5.19	0.48	4.50–6.00	3>2 (−3.42**)
Feasibility 3	5.41	0.46	4.50–6.50	3>1 (−3.16*)
Suitability 1	4.53	0.78	3.00–6.00	2>1 (−4.14**)
Suitability 2	5.09	0.49	4.50–6.00	–
Suitability 3	5.28	0.55	4.50–6.50	3>1 (−4.74**)

Note: *$p < 0.05$, **$p < 0.01$.

by their peers were found to be 4.75, 5.29, 5.08, and 4.74 on the four dimensions, respectively. The average scores in the third round as evaluated by their peers were 5.08, 5.36, 5.12, and 4.97 on these dimensions. Also, the average scores evaluated by the expert had a similar increasing trend across different rounds, shown in Table 4.

A series of paired *t*-tests was used to analyse these participants' score changes across the three rounds of PA, shown in Tables 3 and 4. Although Table 3 indicated that the scores evaluated by peers only significantly increased in the 'creativity' dimension between the first and the third round, the scores marked by the expert, presented in Table 4, indeed revealed the participants' apparent progression along the PA activity. According to Table 4, the scores evaluated by the expert on the dimensions of creativity, feasibility, and suitability showed significant increase from the first to the second round ($t = -2.44, -2.18$, and -4.14 on creativity, feasibility, and suitability, $p < 0.05$). From the second to the third round, the score evaluated by the expert progressed significantly only on the dimension of feasibility ($t = -3.42, p < 0.05$). These results might suggest that the research proposals related to creativity, feasibility and suitability could be improved greatly at the initial stage of PA, but at a later round of PA significant enhancement in terms of the creativity and suitability was not found.

Nevertheless, by comparing the expert scores of the first round and the third round, the dimensions of creativity, feasibility, and suitability made a statistically substantial progression ($t = -3.22, -3.16$, and -4.74, respectively, $p < 0.01$). Also, scores evaluated by the expert on the dimension of relevance had no statistically significant progression probably due to the relatively high score assigned by the expert in the first round (mean = 5.22). The results indicated that these participants benefited from the PA intervention for developing more creative, feasible and suitable research proposals, particularly from the expert's perspective.

The correlation between peer feedback and proposal progression

As described previously, each peer feedback was categorised into one of the seven categories listed in Table 1. Table 5 shows the frequency of feedback each group

Table 5.   Content analysis of peer feedback by categories.

	Mean (SD)		
	First round	Second round	Third round
Supporting	9.38 (5.16)	14.12 (5.83)	11.5 (5.19)
Emotional Response	0.19 (0.40)	0.25 (0.58)	0.06 (0.25)
Direct Correction	10.06 (7.73)	10.94 (7.08)	8.06 (3.68)
Personal Advice	6.44 (3.85)	4.81 (4.05)	3.19 (1.87)
Guiding	16.06 (10.16)	11.94 (5.56)	9.31 (7.25)
Evaluating and Planning	19.94 (10.10)	16.44 (7.81)	15.31 (7.47)
Regulating and Reflecting	8.69 (5.77)	8.00 (2.56)	9.69 (7.9)
Total	70.75 (23.49)	66.50 (12.38)	57.12 (16.35)

obtained on the seven categories across three rounds of PA. These results indicated that the highest frequency of the peer feedback of the three rounds provided by learning peers was the type of 'Evaluating and Planning', implying that the peers might have highly metacognitive reflective thinking for giving comments to their peers. Nevertheless, the total frequency of the peer feedback gradually decreased from the first round to the third round during the PA activity (from 70.75 for the first round to 57.12 for the last round).

Furthermore, it is plausible to assume that the feedback provided by peers under the first round of PA might influence the performance progression from the first round to the second round. Similarly, the peer feedback from the second round might be related to the progression from the second to the third round. Table 6 shows the correlation between the frequency of each feedback type under the first PA round each group obtained and its score change from the first round to the second round. The results showed that the total frequency of the peer feedback in the first round was positively correlated with subsequent enhancement of the proposal ($r = 0.66$, 0.55 by peer score for the dimensions of feasibility and suitability; $r = 0.55$, 0.67, 0.62, and 0.57 by expert score for the four dimensions, $p < 0.05$). These findings suggest that the more peer feedback given to the group, the more likely the group could improve the research proposal.

Table 6 also reveals that the feedback type of 'Evaluating and Planning' was highly positively correlated with the participants' score progression on some dimensions ($r = 0.52$, 0.62, 0.57 by peer score for the dimensions of relevance, feasibility, and suitability, $r = 0.50$ by expert score for the dimension of suitability, $p < 0.05$), and the positive role of 'Regulating and Reflecting' feedback was also observed on the participants' performance progression ($r = 0.54$ by peer score of feasibility, $r = 0.62$, 0.53 by expert score for creativity and relevance, $p < 0.05$). Also, 'Direct Correction' feedback was positively related to the progression in feasibility and suitability in light of expert perspective ($r = 0.62$ and 0.52, $p < 0.05$), and a similar finding was found on the 'Guiding' feedback on the dimensions of relevance and feasibility ($r = 0.56$ and 0.65 for expert score, $p < 0.05$). These results, therefore, displayed that metacognitive-oriented comments (such as 'Evaluating and Planning' and 'Regulating and Reflecting') played an essential role in the performance progression of many dimensions; for the role of cognitive comments, only the feedback of 'Direct Correction' and perhaps 'Guiding' were helpful in assisting the participants'

Table 6. The relationship between the peer feedback obtained from the first round and score progression from the first round to the second round.

Round 1 feedback classification	Score progression from round 1 to round 2							
	Peer				Expert			
	Creativity 21	Relevance 21	Feasibility 21	Suitability 21	Creativity 21	Relevance 21	Feasibility 21	Suitability 21
Total feedback	0.48	0.40	0.66*	0.55*	0.55*	0.67*	0.62*	0.57*
Supporting	0.10	−0.15	0.02	−0.25	−0.14	−0.23	−0.33	−0.39
Emotional Response	0.33	−0.19	0.14	−0.05	−0.14	0.25	0.05	0.13
Direct Correction	0.26	0.03	0.46	0.38	0.43	0.46	0.62*	0.52*
Personal Advice	0.36	0.10	−0.10	−0.01	−0.06	−0.07	−0.12	0.35
Guiding	0.24	0.23	0.45	0.35	0.40	0.56*	0.65*	0.37
Evaluating and Planning	0.43	0.52*	0.62*	0.57*	0.11	0.47	0.22	0.50*
Regulating and Reflecting	0.37	0.43	0.54*	0.47	0.62*	0.53*	0.36	0.21

Note: *$p < 0.05$.

development of better proposals ($p < 0.05$) at an early stage of PA activity (i.e. from the first round to the second round).

Table 7 presents the relationships between the types of feedback which the participants received from the second round and the score progression from the second to the third round. As a result, the total frequency of feedback under the second round was not highly correlated to the development of proposals from the second round to the third round. Similarly, the results did not reveal a statistically significant correlation between the types of feedback and the progression in each dimension, with only a few exceptions (e.g. the feedback of 'Personal Advice' and expert score change in 'creativity'). These results might imply that the different types of feedback at a later stage (the second round to the third round) did not play a significant role in the progression of the participants' proposals.

Attitude towards (online) PA

This study explored the attitudes toward PA before and after the treatment activity and compared the difference, presented in Table 8. The mean scores of two subscales on PAS and OAS are 3.87 and 3.72, respectively, before this online PA intervention, and the mean scores of the two subscales are 3.81 and 3.80 after it. Participants' attitudes toward online PA were generally positive prior to and after this learning activity in the five-point Likert scale. Besides, no statistically significant differences on the two subscales were found prior to and after the treatment activity. Although this study did not show that the participants expressed significantly more favourable attitudes toward PA as a result of the PA intervention, their positive attitudes were, at least, statistically maintained.

Discussion and conclusion

This study implemented a three-round online PA activity to help in-service teachers develop educational research proposals. It was found that through the online PA activity, the in-service teachers could enhance the quality of their educational proposals in many aspects, especially in the light of expert's evaluations.

Moreover, this study evaluated the participants' work in four dimensions, including the creativity, relevance, feasibility and suitability. The analyses from the dimensions showed some interesting differences. For example, the performance in the dimensions of creativity and suitability had great progression at the initial stage (from the first round to the second round), but no substantial improvement at the later stage (from the second round to the third round). On the other hand, in the dimension of feasibility, the participants could have significant gains from one round to a later round of PA. For the dimension of relevance, the PA might not be quite effective to improve it, probably due to its relatively superior performance initially. These findings likely suggested that for the enhancement of the feasibility of work, more rounds of PA may be helpful, but for the other dimensions, additional rounds may not be quite so useful.

This study further examined the role of peer feedback obtained by the participants in their subsequent performance. The results, in general, revealed that the quantity of feedback of the first round obtained by the participants, particularly that of metacognitive-oriented feedback, was positively correlated with their improvement from the first round to the second round. However, the role of peer feedback for the progression

Table 7. The relationship between the peer feedback obtained from the second round and score progression from the second round to the final round.

| | Score progression from round 2 to round 3 | | | | | | | |
| | Peer | | | | Expert | | | |
Round 2 feedback classification	Creativity 32	Relevance 32	Feasibility 32	Suitability 32	Creativity 32	Relevance 32	Feasibility 32	Suitability 32
Total feedback	-0.12	-0.15	-0.38	-0.16	0.31	0.35	-0.26	0.02
Supporting	-0.24	-0.25	-0.38	-0.28	-0.16	0.18	0.23	0.23
Emotional Response	-0.33	0.38	-0.48	-0.15	0.16	0.38	-0.10	0.28
Direct Correction	-0.33	0.38	-0.48	-0.15	0.16	0.38	-0.11	0.28
Personal Advice	0.41	0.18	0.18	0.18	0.69^{**}	0.12	0.22	0.20
Guiding	0.29	-0.04	0.43	-0.10	0.53^{*}	-0.08	0.10	0.18
Evaluating and Planning	-0.05	0.26	-0.03	0.15	-0.33	0.00	-0.40	-0.16
Regulating and Reflecting	-0.05	0.06	0.04	-0.10	-0.04	0.21	0.66^{**}	0.29

Note: $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 8.   The participants' attitudes toward PA before and after the treatment ($n = 52$).

| Subscale | Mean (SD) | | t value |
	Pre-treatment	Post-treatment	
Positive Attitude Subscale (PAS)	3.87 (0.38)	3.81 (0.38)	1.15 (ns)
Online Attitude Subscale (OAS)	3.72 (0.49)	3.80 (0.44)	−0.17 (ns)

Note: ns, not significant.

between the second round and the final round was not apparent. In the light of these findings, in the early stage of PA, instructors should not only encourage students to provide more feedback to their peers, but also encourage students to provide metacognitive-oriented feedback. For the later stage, more research should be conducted to carefully explore other factors that may contribute to participants' progression.

This study also examined the in-service teachers' attitudes toward (online) PA, and found their attitudes were not statistically significantly different prior to and after the intervention. Wen and Tsai (2008) also conducted a similar study on an in-service teacher education programme using a similar online PA research setting and module. But they found that the in-service teachers' attitudes dropped significantly after the PA treatment. They suspected that the criteria for marking were not explained clearly before the online PA, thus leading to the participants' unfavourable attitudes toward PA after treatment. The importance of marking criteria in PA is also highlighted by Orsmond et al. (1996). Based upon the suggestions of Wen and Tsai (2008) and Orsmond et al. (1996), this study tried to improve the utilisation of online PA. This study showed that clarified criteria and sufficient support from the instructor, can maintain, though not enhance, participants' attitudes toward PA, which were relatively optimistic throughout. More attempts at using online peer assessment are encouraged to examine fully its usefulness as well as limitations.

Acknowledgement

Notes on contributors

Ying-Chih Chen is a doctoral student of Science Education at the University of Iowa in the USA. His current research interests include how language may be used as a learning tool to improve students' understanding of science, and how science teachers develop their pedagogical content knowledge.

Chin-Chung Tsai is currently a Chair Professor at the National Taiwan University of Science and Technology, Taiwan. His research interests deal largely with constructivism, epistemological beliefs, Internet-based instruction and human behaviour in Internet environments.

References

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International, 37*(4), 346–354.
Davies, P. (2006). Peer assessment: Judging the quality of students' work by comments rather than marks. *Innovations on Education and Teaching International, 43*(1), 69–82.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*(2), 175–187.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in high education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Liu, C.-C., & Tsai, C-M. (2005). Peer assessment through web-based knowledge acquisition: Tools to support conceptual awareness. *Innovations in Education and Teaching International, 42*(1), 43–59.

Liu, E.Z.F., Lin, S.S.J., & Yuan, S.M. (2001). Design of a networked portfolio system. *British Journal of Educational Technology, 32*(4), 492–494.

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26*(1), 53–63.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*(3), 239–250.

Prins, F.J., Sluijsmans, D.M.A., Kirschner, P.A., & Strijbos, J.W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education, 30*(4), 417–444.

Topping, K.J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Tsai, C.-C., & Liang, J.-C. (in press). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science.*

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*(4), 1161–1174.

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.

Wen, L.M.C., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*, 27–44.

Wen, L.M.C., & Tsai, C.-C. (2008). Online peer assessment in an in-service science and mathematics teacher education course. *Teaching in Higher Education, 13*, 55–67.