

## **ALTERNATIVES TO INSTRUCTOR ASSESSMENT: A CASE STUDY OF COMPARING SELF AND PEER ASSESSMENT WITH INSTRUCTOR ASSESSMENT UNDER A NETWORKED INNOVATIVE ASSESSMENT PROCEDURES**

**ERIC Z.F. LIU**

*Ming Hsin University  
of Science and Technology, Taiwan*

**SUNNY S.J. LIN AND S.M. YUAN**

*National Chiao Tung University, Taiwan*

### **ABSTRACT**

Although the effectiveness of self, peer, and instructor assessment in higher education is well studied, the possibility of using networked system to facilitate alternative assessment in undergraduate level has seldom been studied. Quantitative analysis of self, peer and instructor assessment under a networked system is also rare. In addition to outlining procedures of networked innovative assessment, this study also provides a long-term measure of peer-self, peer-instructor and self-instructor correlation coefficients with networked innovative assessment procedures. Analytical results indicated that undergraduate students could evaluate peers adequately but they over-mark themselves usually. Although the findings of this study have agreed with previous research, authors believe that networked innovative assessment procedures should be further improved and extended to other course topics for further evaluation of its effectiveness.

**Keywords:** Peer assessment, self assessment, instructor assessment, networked innovative assessment procedures

### **INTRODUCTION**

Recently, Many innovative assessment strategies have been explored. An example of those strategies is the extensive use of peer-, self-, portfolio- and co-assessment (1, 6, 7, 18, 23, 27). This study examines undergraduate students' peer, self and instructor assessment as innovative pedagogical and learning approaches.

Peer assessment or self assessment, a usual process used by instructors to teach students to responsibly evaluate themselves and their peers from kinder-

garten to university, has been extensively studied in higher education in recent decades (8, 23, 25). Researches in higher education have explored the reliability and practicalities of peer assessment and generally conferred on its acceptability (11, 12, 16, 21, 26). However, the comparison of undergraduate students' peer and self assessment with instructor assessment under networked system has seldom been studied.

While Gopinath (14) showed student self-assessment tended to be higher compared with that of the instructor and that of the peers across all forms of evaluation. Also, Gopinath (14) found that peer assessment is closer to instructor assessment but suffer from low reliability. Researchers (23) pointed out that peer assessment highly correlated with the final grade and staff assessment, and that self assessment highly correlated with the results of peer assessment in some literatures. The extent to which the comparison of peer, self and instructor assessment under a networked system should be investigated.

Researchers (3, 13, 17) designed a Networked Peer assessment system known as NetPeas. Liu (12) adopted NetPeas in peer assessment activities of an undergraduate operating systems course. In this study, authors have added self assessment into NetPeas and is adopted to facilitate undergraduate students and instructor performing peer, self and instructor assessment in operating systems course. The correlation coefficients of peer-self, peer-instructor, self-instructor assessment are also calculated.

Previous studies (12, 17) indicated that using NetPeas to facilitate peer assessment has the following merits:

1. Higher assurance of anonymity than paper-and-pencil peer assessment.
2. Increased freedom of time and location for learners.
3. Ability of students to modify their work more timely.
4. Increased student-instructor and student-student interaction and feedback.
5. Significantly lower transmission and delivery costs than "paper-and-pencil" peer assessment.
6. Fewer limitations on transmission of data than "paper-and-pencil" peer assessment.

## **PERTINENT LITERATURE**

### **Peer assessment**

Falchikov (10) has defined peer assessment as the process whereby groups of individuals grade their peers. Another researcher (25) more accurately defined "peers" as students with similar educational background. Peer assessment has been implemented in several higher education courses such as writing, engineering, sciences, information science and arts. Many other researchers' experiences of peer assessment, as a formative assessment strategy and as part of the learning, can enable students to become more involved in the class activities; they view peer assessment as fair and accurate as well (23).

Dochy and Segers (8) indicated two guidelines for educators who employed

peer assessment in their instruction. We list these two guidelines in the following: 1) Peer assessment criteria should be determined beforehand. Experiences show that it works well these criteria are determined jointly by staff and students and 2) peer assessment criteria should be presented in operational terms with which all students are familiar. Falchikov and Magin (11) have also suggested that reliable and valid outcomes from peer assessment depend on the following: 1) Students must be committed to and fully understand the educational purpose of peer assessment, 2) students must be involved in determining the criteria and agreeing on a grading scale and assessment procedure and 3) Students must receive feedback on peer assessment scores, both in relation to their own performance and to the overall pattern of scores.

### **Self assessment**

Dochy and Segers (8) indicated that self assessment refers to the involvement of learners in making judgements about outcomes or achievements of their learning. Sluijsmans *et al.* (23) and Bond (2) have said that self assessment is not a new assessment strategy, but a good way of increasing the role of students as active participants in the learning process of themselves. Also, self assessment was mostly implemented in formative assessment to foster reflection on one's own learning process and results.

Dochy and Segers (8) also proposed two guidelines for educators who employed self assessment in their instruction. We list these two guidelines in the following: 1) Self assessment spends more time on learning, and sometimes support for students will be necessary during the self assessment and 2) self assessment can be used easily for formative purposes. Students must learn to see this as a tool for learning.

### **Peer and self assessment**

Falchikov (9) and Freeman (12) indicated a fairly high level of agreement between the marks given by peers and those given by the teacher. Cheng and Warren (4) confirmed that evaluating a particular task could improve students' assessment skills when assessing a similar task. Despite its merits, peer assessment is limited by corrections made based on friendship and decibel marking (23, 28). Sluijsmans *et al.* (23) provided several examples to illustrate the reliability of peer and self assessment and concluded that friendship marking should not be taken as too large a problem.

Sluijsmans *et al.* (23) mentioned a literature that studied six cases that included peer and/or self assessment. Students were highly aware of that peer and self assessment helped them to develop important skills, such as problem solving. In the perspective of reliability, Furnham and Stringfield (13) have reported greater reliability in peer assessment than in self assessment. Lennon's (15) research also supported this point of view. Lennon reported that correlation between instructor and peer assessment were modest, correlation between instructor and self assessment were very low (about 0.21), and self assessment

was associated with under-marking and clustering at the median.

### **Networked innovative assessment procedures**

Authors' previous studies have developed networked peer assessment procedures (3, 13, 17) to break up the conventional learning process into smaller categories. Certain arrangements are provided to increase the effective of learning. In this study, authors integrated self assessment into networked peer assessment procedures and would like to call this new combination as networked innovative assessment procedures. Networked innovative assessment procedures comprise of the following steps:

1. Instructor seriously demonstrates the educational objectives of peer and self assessment in the beginning and uses some real samples to prepare students for later assessment activities.
2. Instructor repeats the following steps for each work.
3. Instructor coaches a part of course materials covered in this semester.
4. Students and instructor jointly discuss assignments and the criteria for marking.
5. The assignment completed by the students is uploaded to the NetPeas.
6. NetPeas randomly assigns reviewers (each reviewer marks three assignments).
7. Reviewers assign a score to the homework and comment on it.
8. At the same time, Reviewers also assign a score to their own works and write a narrative to describe the values of their own works.
9. The instructor marks each student's work and observes the comments written by students without revealing the results.
10. NetPeas notifies the students of browsing grades and comments given by peers and comparing peer and self assessment.
11. Based on those comments, each student must make corrections or modifications accordingly.
12. Steps 3 to 11 are repeated once, twice or not at all (researchers can choose the times of repetition based on their needs).

After turning in their own work, students were assigned works to grade, as dispatched by the system. Each student is responsible for assigning a score to peers' and his/her own work. The suggestions and grades evaluated by reviewers are then distributed back to the original author. The authors must reflect on the differences of self and peer assessment and then modify their original work based on the self-evaluated recommendations from peer reviewers.

The above procedures ensure that learning is achieved through three consecutive assessments as well as related improvements. Of priority concern is to help students progress in each round. Therefore, this system allows students to view their grades and comments after each round is completed so as to improve their own work. This study also proposes a method to ensure that each student corrects his (or her) mistakes or eliminates weaknesses during each round. Namely, the same reviewer is assigned to grade an author's homework each round. Doing so allows the same reviewer to examine whether or not the work is modified

according to his (or her) previous comments.

Zhao (29) suggested that anonymity increases critical feedback and makes the participants feel more comfortable in criticizing others. Bradley (3) asked whether or not blind marking could be regarded as truly blind if the material to be assessed contained cues to sex, such as style of handwriting. Falchikov and Magin (11) demonstrated the accuracy of determining the authors' gender through their handwritten assignments. They feared that gender bias would occur during peer assessment if gender could not be concealed. NetPeas ensures anonymity in two ways. First, the instructor instructs students to remove personal information from their electronic files (HTML, Word 97 or Source code). Using electronic files other than handwriting can conceal the author's identity and gender. Second, the system codes the students' identification number into other sequence number randomly. These measures preserve anonymity and reduce the possibility of gender bias.

## **RESEARCH DESIGN**

### **Aims of this study**

- (a) To assess the correlation coefficients between self, peer and instructor assessment in a long-term fashion.
- (b) To compare the mean values of markings between peer-self, instructor-self and peer-instructor assessment in a long-term fashion.

### **Sample**

118 undergraduate students were randomly selected from the operating systems course of the Computer and Information Science Department of a research-oriented university in northern Taiwan.

### **Procedures**

In this study, undergraduate students attended an operating systems course weekly, undertook five open-ended questions selected from a large amount of item banks and a final project. Each student was instructed to do the questions and final project individually. Each question was a writing assignment to verify the understanding of operating systems of students, accounting for 5% of the overall grade. Samples of the question are similar to the following: "What are the differences between a virus and a worm (24)", "The implementation of monitors using semaphores did not use an explicit linked list of blocked processes, whereas the implementation of semaphores using monitors did (24)" and "What advantage is there in having different time-quantum sizes on different levels of a multilevel queuing system ? (22)." The final project was a programming assignment, accounting for 20% of the overall grade. Samples of the final project are similar to the following: "Designing a memory management algorithm for Win CE," "Designing a power saving mechanism for the PDA". The students were

instructed to do the following procedures:

- Instructor demonstrates the educational objectives of peer and self assessment in the beginning and uses some real samples to prepare students for later assessment activities. (This step done in the first of this semester and lasts about 2 hours)
- Instructor coaches a part of course materials covered in this semester. (Each subject was scheduled about 3-4 weeks)
- Students and the instructor collaboratively discuss the assignment and the criteria to make corrections (This step lasts about one day for open-ended questions and about one week for final project).
  1. The instructor discusses the criteria to mark works with students in course time.
  2. Students discuss the final project with the instructor informally.
  3. The only criterion for open-ended questions is "correctness of work". The criterion for final project is also "correctness of execution". The criterion is divided into ten categories from "excellent (10)" to "extremely poor (1)".
- The assignment completed by the students is uploaded to the system (This step lasts about a week).
  1. Students complete the assignment by themselves.
  2. Students are instructed to complete the work and submit it to NetPeas within a week; otherwise they receive no credit on this assignment.
- The system randomly assigns reviewers (each reviewer grades three to four assignments).
  1. The procedure is automatically done by NetPeas after all of the students have uploaded their own assignments to NetPeas.
  2. In this study, three works were assigned to each reviewer.
- Reviewers grade and comment on themselves' and peers' works (This period lasts about a week).
  1. Students can assess each assigned work in one day to alleviate their loading.
- The system notifies the students of their grades and comments.
  1. NetPeas instructs automatically each student to browse the results of peer and self assessment via e-mail after the review process.
- Based on the comments on each student's homework, the authors must make corrections or modifications.
- The above steps are repeated one more time, twice or not all (researchers can select times of repetition based on their needs).
  1. In this study, students were not requested to re-submit their assignments again in the limitation of time.

## Results

Results in this study are divided into two portions. These portions are analyzed quantitatively. The first portion considers the correlation coefficients between self, peer and instructor assessment in a long-term fashion. The second portion compares the mean values of markings between peer-self, instructor-self and peer-instructor

assessment in a long-term fashion. The quantitative data are analyzed through the SPSS 8.0, and all reported values are rounded to the second decimal. Herein, an asterisk represents the significant level of 0.05, two asterisks represent the significant level of 0.01 and three asterisks represent the significant level of 0.001.

### **Correlation coefficients between self, peer and instructor assessment in a long-term fashion**

In the first, authors reported the correlation coefficients of grading of two computer science experts (inter-rater reliability) in these six works ( $r_1=0.72^{**}$ ,  $r_2=0.87^{**}$ ,  $r_3=0.94^{**}$ ,  $r_4=0.92^{**}$ ,  $r_5=0.42^{**}$ ,  $r_6=0.81^{**}$ ). Authors found that inter-rater reliabilities are ranged from modest to significantly high. In the second, the undergraduate students' peer-instructor assessment correlation coefficients of these six works are also ranged from modest to significantly high ( $r_1=0.87^{**}$ ,  $r_2=0.97^{**}$ ,  $r_3=0.98^{**}$ ,  $r_4=0.98^{**}$ ,  $r_5=0.59^{**}$ ,  $r_6=0.96^{**}$ ). In the third, the undergraduate students' peer-self assessment correlation coefficients of these six works are ranged from negative and low to significantly high ( $r_1=0.33^{**}$ ,  $r_2=0.76^{**}$ ,  $r_3=0.83^{**}$ ,  $r_4=0.79^{**}$ ,  $r_5=-0.08$ ,  $r_6=0.83^{**}$ ). In the fourth, the undergraduate students' self-instructor assessment correlation coefficients of these six works are ranged from negative and low to significantly high ( $r_1=0.25^{**}$ ,  $r_2=0.74^{**}$ ,  $r_3=0.83^{**}$ ,  $r_4=0.74^{**}$ ,  $r_5=-0.03$ ,  $r_6=0.83^{**}$ ).

In this study, the peer assessment is more related to (validity) instructor assessment than self assessment, it is particularly clear in the fifth work. This result agreed with previous researchers' finding (13, 15).

### **Comparing the mean values of markings between peer-self, instructor-self and peer-instructor assessment in a long-term fashion**

In the first, authors compared the mean values of peer-self assessment in these six works ( $t_1=-1.04$ ,  $t_2=-2.42^*$ ,  $t_3=-2.68^{**}$ ,  $t_4=-2.26^*$ ,  $t_5=-0.69$ ,  $t_6=-2.37^*$ ). Authors found that peer assessments are significantly lower than self assessments in four works (67%). In the second, authors compared the mean values of instructor-self assessment in these six works ( $t_1=-1.76$ ,  $t_2=-3.56^{**}$ ,  $t_3=-3.29^{**}$ ,  $t_4=-2.74^{**}$ ,  $t_5=-2.51^*$ ,  $t_6=-3.49^{**}$ ). Authors found that instructor assessments are significantly lower than self assessments in five works (83%). In the third, authors compared the mean values of peer-instructor assessment in these six works ( $t_1=1.94$ ,  $t_2=3.51^{**}$ ,  $t_3=1.84$ ,  $t_4=2.57^*$ ,  $t_5=3.07^{**}$ ,  $t_6=2.12^*$ ). Authors found that peer assessments are significantly higher than instructor markings in four works (67%). These results revealed that self assessments are usually significantly higher than peer assessments and peer assessments are often significantly higher than instructor assessments. Gopinath (14) also showed a similar result, but the correlation coefficients of peer-instructor assessment in Gopinath's research are lower than this study.

## LIMITATIONS

This research has been applied to undergraduate students undertaking an operating systems course at a research-oriented university in northern Taiwan. Students are almost consisted of seniors, juniors and some sophomores. Applying the same method to other subjects would not necessarily yield the same results, particularly for freshmen lacking the necessary academic skills and discipline.

Some researchers have suggested authors to rename NetPeas as networked portfolios assessment system (NetPorts). Authors would not agree that statement, because portfolios assessment should include more qualitative evidences analyzing the progress of learners. Although, NetPeas included peer, self and instructor assessment that is strongly correlated with portfolios assessment, Authors would not dare to name NetPeas as NetPorts. Authors are looking forward to approach the portfolios assessment and diligently doing practical researches in the near future.

## CONCLUSIONS

This study has exploited the correlation of peer-instructor, peer-self and self-instructor assessment and comparison the mean values of peer-self, instructor-self and peer-instructor assessment. Analysis results indicated that the peer assessment is more related to (validity) instructor assessment than self assessment. This result agreed with previous researchers' finding (13, 15). And, another results revealed that self assessments are usually significantly higher than peer assessments and peer assessments are often significantly higher than instructor assessments. The second finding agreed with Gopinath's (14) research.

This study suggested that the procedures of networked innovative assessment would be quite well to achieve a higher reliability and validity than before (17). Authors hope to extend it to other courses to further demonstrate its effectiveness. The portfolios assessment would prevail over numerous disciplines, so authors will diligently make more improvements on NetPeas to upgrade it into NetPorts and fulfill the new instructional needs.

Direct Reprint Requests to:

Eric Zhi-Feng Liu

Department of Information Management

Ming Hsin University of Science and Technology

1 Hsin-Hsing Road, Hsin-Fong

Hsinchu, Taiwan 304

Republic of China

E-mail: totem@mis.mhit.edu.tw

## REFERENCES

1. D. Bond, "Assessment and the promotion of academic values," *Studies in Higher Education*, Vol. 15, 101-113, 1990.
2. D. Bond, *Enhancing learning through self-assessment*, London: Kogan Page, 1995.
3. C. Bradley, "Sex bias in student assessment overlooked?," *Assessment & Evaluation in Higher Education*, Vol. 18, 1, 3-8, 1993.
4. W. Cheng, and M. Warren, "Peer and teacher assessment of the oral and written tasks of a group project," *Assessment & Evaluation in Higher Education*, Vol. 24, 3, 301-314, 1999.
5. C. H. Chiu, W. R. Wang, and S. M. Yuan, "Web-based Collaborative Homework Review System," *Proceedings of ICCE'98*, Vol. 2, 474-477, 1998.
6. R. Davies, and T. Berrow, "An evaluation of the use of computer supported peer review for developing higher-level skills," *Computers and Education*, Vol. 30, 1-2, 111-115, 1998.
7. F. J. R. C. Dochy, and L. McDowell, "Assessment as a tool for learning," *Studies in Educational Evaluation*, Vol. 23, 279-298, 1997.
8. F. Dochy, and M. Segers, "The use of self-, peer and co-assessment in higher education: A review," *Studies in Higher Education*, Vol. 23, 3, 331-350, 1999.
9. N. Falchikov, "Group process analysis: self and peer assessment of working together in a group," *Educational and Training Technology International*, Vol. 30, 275-284, 1993.
10. N. Falchikov, "Peer feedback marking: developing peer assessment," *Innovations in Education and Training International*, Vol. 32, 175-187, 1995.
11. N. Falchikov, and D. Magin, "Detecting gender bias in peer marking of students' group process work," *Assessment and Evaluation in Higher Education*, Vol. 22, 4, 385-396, 1997.
12. M. Freeman, "Peer assessment by groups of group work," *Assessment and Evaluation in Higher Education*, Vol. 20, 3, 289-300, 1995.
13. A. Furnham, and P. Stringfield, "Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation," *Journal of Occupational and Organizational Psychology*, Vol. 67, 57-67, 1994.
14. C. Gopinath, "Alternatives to instructor assessment of class participation," *Journal of Education for Business*, Vol. 75, 1, 10-14, 1999.
15. S. Lennon, "Correlations between tutor, peer and self assessments of second year physiotherapy students in movement studies. In S. Griffiths, K. Houston, & A. Lazenblatt (Eds.), *Enhancing student learning through peer tutoring in higher education: Section 3 - Implementing* (Vol 1, 66-71)." Coleraine, Northern Ireland: University of Ulster, 1995.
16. S. S. J. Lin, and E. Z. F. Liu, "The Learning Effects of Web-based Peer Assessment for Various Thinking Styles Students," Paper presented at the annual meeting of the American Education Research Association, New Orleans, Louisiana, 2000.

17. S. S. J. Lin, E. Z. F. Liu, and S. M. Yuan, "Student attitudes toward networked peer assessment: Case studies of undergraduate students and senior high school students," *International Journal of Instructional Media*, Vol. 29, 2, 2002.
18. M. C. Linn, and S. Hsi, "Computers, teachers, peers-science learning partners," Lawrence Erlbaum, Mahwah: NJ, p. 140-148, 2000.
19. E. Z. F. Liu, "Networked peer assessment system: an analysis of student segments," A thesis for the master degree of Computer and Information Science of National Chiao-Tung University, 1999.
20. E. Z. F. Liu, C. H. Chiu, S. S. J. Lin, and S. M. Yuan, "Student participation in computer science courses via the Networked Peer Assessment System (NetPeas)," *Proceedings of the ICCE'99*, Vol. 1, 774-777, 1999.
21. P. Orsmond, S. Merry, and K. Reiling, "The importance of marking criteria in the use of peer assessment," *Assessment and Evaluation in Higher Education*, Vol. 21, 239-249, 1996.
22. A. Silberschatz, and P. B. Galvin, "Operating system concepts forth edition," Addison-Wesley Publishing Company, 1994.
23. D. Sluijsmans, F. Dochy, and G. Moerkerke, "Creating a learning environment by using self-, peer- and co-assessment," *Learning Environments Research*, Vol. 1, 3, 293-319, 1999.
24. A. S. Tanenbaum, "Modern operating systems," Prentice-Hall International, London, 1992.
25. K. Topping, "Peer assessment between students in colleges and universities," *Review of Educational Research*, Vol. 68, 249-276, 1998.
26. K. J. Topping, E. F. Smith, I. Swanson, and A. Elliot, "Formative peer assessment of academic writing between postgraduate students," *Assessment and Evaluation in Higher Education*, Vol. 25, 2, 149-169, 2000.
27. B. Wilcox, and L. Tomei, "Professional Portfolios for Teachers : A guide for learners, experts, and scholars," Christopher-Gordon Publishers Press, 1999.
28. E. Williams, "Student attitudes towards approaches to learning and assessment," *Assessment and Evaluation in Higher Education*, Vol. 17, 1, 45-58, 1992.
29. Y. Zhao, "The effects of anonymity on computer-mediated peer review," *International Journal of Educational Telecommunications*, Vol. 4, 4, 311-345, 1998.

### AUTHORS NOTES

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract Nos. NSC89-2520-S-009-013 and NSC89-2520-S-009-016. The authors also thank to all of research assistants involved in these two projects.