

## **Web-based peer assessment: feedback for students with various thinking-styles**

S.S.J. Lin, E.Z.F. Liu & S.M. Yuan

Inst. of Education and Dept of Computer and Information Science, National Chiao Tung University

**Abstract** This study used aptitude treatment interaction design to examine how feedback formats (specific vs. holistic) and executive thinking styles (high vs. low) affect web-based peer assessment. An Internet-based (anonymous) peer-assessment system was developed and used by 58 computer science students who submitted assignments for peer review. The results indicated that while students with high executive thinking styles significantly improved over two rounds of peer assessment, low executive students did not improve through the cycles. In addition, high executive students contributed substantially better feedback than their low executive counterparts. In the second round of peer assessment, thinking style and feedback format interactively affected student learning. Low executive students receiving specific feedback significantly outperformed those receiving holistic feedback. In receiving holistic feedback, high executive thinkers outperformed their low executive counterparts. This study suggests that future web-based peer assessment adopts a specific feedback format for all students.

**Keywords:** Aptitude; Computer science; Discourse analysis; Feedback; Peer assessment; Undergraduate; World-wide web.

### **Introduction**

Peer assessment as an innovative assessment method has been used extensively in higher educational settings in diverse fields such as writing, business, science, engineering, and medicine (Falchikov, 1995; Freeman, 1995; Strachan & Wilcox, 1996; Rada, 1998). While reviewing recent developments in peer assessment, Topping (1998) found that most students benefit from this assessment method. Topping anticipated the increasingly popularity of web-based peer assessment in the new century because the rapid development of Internet technologies has ushered in an increasing interest in online web learning (Yagelski & Powley, 1996; Owston, 1997; Mason & Bacsich, 1998; Steeples & Mayers, 1998; Barrett & Lally, 1999; Fabos & Young, 1999).

In this study, the web-based peer assessment was a two-stage compulsory evaluation that partially substituted for teacher assessment. During the process, a student submitted assignments in HTML format which were then anonymously uploaded through a web-based peer assessment system, named Networked Peer

---

Accepted 20 January 2001

Correspondence: Sunny S.J. Lin, Institute of Education, National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan  
Email: sunnylin@cc.nctu.edu.tw

assessment system (*NetPeas*, Liu *et al.*, 1999). Six anonymous peer reviewers were assigned to mutually assess peers' assignments and provide feedback. The scores and feedback were then sent to the original author who then revised the original assignment based on those peers' feedback. Thus, the web-based peer assessment was formative, anonymous, and asynchronous in nature. Undergraduate students undertake the teachers' role of assessor and feedback provider in addition to their conventional role as learner. These multiple roles require that students exert more effort than in a normal setting.

This assessment scheme is modelled after the authentic journal publication process of an academic society (Rogoff, 1991; Roth, 1997). Each student assumes the role of a researcher who hands in assignments as well as a reviewer who comments on peers' assignments. Thus, students construct and refine knowledge through social interaction in a virtual community linked via the Internet.

Preliminary studies on the learning effect of *NetPeas* (Lin *et al.*, 2001a; Lin *et al.*, 2001b) revealed that most computer science students using this novel learning process performed well. However, some students complained that a holistic peer feedback was often too vague or useless for self-improvement. Therefore, in this study, the effects of holistic and multiple-specific feedback were compared. Students offer specific feedback, if they respond closely to aspects of peers' submissions.

This study also analyses who may benefit more from web-based peer assessment. Individuals with the same ability level may have very different styles in using their abilities. A learning task may be favoured by one, but not by another. A student with a certain learning style is likely to adapt better to the time-consuming and demanding environment of peer assessment. It was assumed that individuals who are more willing to follow instructional rules would perform better than those who urge for independence and creativity. Sternberg (1998) postulated that individuals who tend to follow regulations and solve problems by designated rules have an 'executive' thinking style. Therefore, this study compares how high and low executive thinking style students react to peer assessment.

### Research questions

This study attempted to answer the following research questions.

- Do students achieve similar peer assessment scores during the first round before they offer feedback?
- Is the students' second round achievement significantly higher than that during the first round?
- Do high executive students offer better quality feedback because they are more willing to follow the instructional rules?
- Do high and low executive thinkers achieve differential learning performances under specific vs. holistic feedback formats?
- How do students feel about web-based peer review and *NetPeas*?

### Advantages and weakness of peer assessment

Van Lehn *et al.* (1995) suggested that peer assessment demands cognitive activities such as reviewing, summarising, clarifying, giving feedback, diagnosing errors and identifying missing knowledge or deviations from the ideal. In peer assessment, students have more opportunities to view assignments of peers than in usual teacher

assessment settings. Instead of modelling a teacher's cognitive product or process, students learn through cognitive modelling of peers' work. Moreover, peer assessment emphasises providing and receiving feedback. Previous studies (Crooks, 1988; Kulik & Kulik, 1988; Bangert-Drowns *et al.* 1991) indicate that receiving feedback is correlated with effective learning. Receiving abundant and immediate peer feedback can prevent some errors and provide hints for making progress.

In addition to the positive effects of peer assessment, Lin *et al.* (2001a) observed that some students had negative feelings about this learning strategy. Some students disliked peer assessment because raters were also competitors. In one case, students could change their previous score during a certain period. Upon receiving an unexpectedly low score from peers, students often reduced the previous scores they had given to others as a form of retaliation. Moreover, students often believe that only teachers have the ability and knowledge to evaluate and provide critical feedback (Zhao, 1998). They may suspect peers' ability; in particular, those who receive lower scores regard peer assessment as inaccurate (McDowell, 1995). Furthermore, many educators refuse to adopt peer assessment owing to the possibility of overmarking or undermarking peers' performance.

Web-based peer assessment has some advantages over ordinary peer assessment. First, students evaluate peers' work through the web (not in a face-to-face presentation), thereby ensuring anonymity and facilitating a willingness to critique. Second, web-based peer assessment allows teachers to monitor students' progress at any period of the assessment process. Teachers can always determine how well an assessor or assessee performs and constantly monitors the process whereas this is nearly impossible during ordinary peer assessment when several rounds are involved. Third, web-based peer assessment can decrease photocopying time and expense since assessees do not need to photocopy their assignments for their peer assessors.

Some researchers (Downing & Brown, 1997; Davis & Berrow, 1998; Zhao, 1998) explored the feasibility of Internet supported peer assessment. However these studies mainly relied on general-purpose applications or commercial software, such as email, electronic communication applications or ftp. Using applications not specifically designed for peer assessment may either increase management load or cause difficulty in maintaining anonymity among peers. Kwok & Ma (1999) and Rada (1998) are among the few researchers who have implemented web-based peer assessment. Kwok & Ma (1999) used Group Support Systems (GSS) to support collaborative and peer assessment. Rada (1998) supervised three classes of computer science students to solve exercise problems and submitted solutions for peer review using a Many Using and Creating Hypermedia system (MUCH). In summary, students were found to learn effectively only if they were highly motivated and the grading policy required mandatory peer assessment.

In this study, *NetPeas* was used to support web-based peer assessment and a detailed system description is provided below. A pilot study (Lin *et al.* 2001b) involving 143 computer science majors attempted to verify the functionality and effectiveness of *NetPeas*. Students were asked to summarise readings on Operating Systems and posted them on *NetPeas*. For each of the three cycles of peer assessment, *NetPeas* provided peer assessors with a single box (holistic style) to comment on. Although some students contributed lengthy feedback with quality (explanative or suggestive, Chi, 1996) information, most only sent simple right/wrong feedback. Many students regarded the holistic feedback towards their

assignment as too vague for further modification. Therefore, students suggested future peer assessment should offer specific (multiple) feedback closely parallel with certain aspects of the assignment. Peer assessment research has seldom studied various effects of feedback formats. However, it is hypothesised in this study that higher order thinking processes are demanded for students providing and receiving specific feedback than when giving holistic feedback.

### Students' learning styles in web-based peer assessment

Although some individual differences may influence student's performance in peer assessment, this issue has seldom been addressed. Among rare examples, Carson & Nelson (1996) compared how Chinese and Latino Americans acted in the peer assessment process, indicating that Chinese-speaking students preferred group harmony and avoided harshly critiquing peer assignments.

Although previous research reports on the overall effectiveness of peer assessment, it is very likely that a student with a certain learning style adapts better to the demanding process of peer assessment. Learning style refers to the manner in which individuals concentrate on, process and retain new information (Dunn *et al.* 1994; Dunn & Dunn, 1999). Biological and sociological uniqueness determines how an individual learns differently from another. Identical instructional environments, strategies, and resources are effective for some learners, but not for others. A meta-analysis of 42 experimental studies conducted at 13 different universities revealed that matching the students' learning styles with instructional methods and resources led to a higher degree of achievement (Dunn *et al.* 1995).

Sternberg (1998) used a metaphor of mental self-government to describe how individuals govern or organise their thinking. Individuals appear to have a mental government to direct their thinking processes when they allocate cognitive resources or establish priorities during decision making in order to respond effectively to the changing world. When dealing with working or thinking rules, individuals tend to display three forms of mental governance: legislative, executive and judicial. Executive people tend to follow regulations and solve problems by designated rules. Legislative people tend to create working rules and avoid solving problems that must follow pre-established rules. In contrast, judicial people prefer to evaluate rules and solve problems that require comparing and analysing multiple concepts. Sternberg (1994) contended that schools tend to cultivate executive thinking style but inhibit legislative thinking style.

Thinking styles do not refer to an individual's abilities or motivation, rather, to traits regarding the tendency of how an individual uses his or her abilities. Restated, highly motivated students may choose to follow, create, or critique rules in an instructional situation based on their preferred thinking style.

The experimental setting of this study was a Taiwanese research university. Subjects were computer science majors who ranked in at least the upper one-tenth of a highly competitive nation-wide university entrance examination. Therefore, they were assumed to be capable of executing higher order thinking to give feedback. From previous experience (Lin *et al.* 2001a; Lin *et al.* 2001b), peer assessment increases the normal course load. Following Sternberg's concept, those who have a higher executive thinking style are more willing to follow instructional rules than those who emphasise independence and creativity. Because they are more willing to follow the demanding rules of peer assessment, they could achieve better than low

executive thinkers. Moreover, the motivation of all experimental groups was examined and kept constant. Throughout this examination, any experimental effect could hopefully not be attributed to the differential learning motivation of experimental groups.

## Methodology

### *Subjects and tasks*

Subjects were 58 computer science majors (43 juniors and 15 seniors, 18 females and 40 males.) enrolled in a mandatory 'Operating Systems' course at a Taiwanese university. Eight topics were covered over an 18-week period, i.e. Introduction, Processes, Memory management, File Systems, Input and Output, Deadlocks, Case Study 1: UNIX, and Case Study 2: MS-DOS. The peer assessment task required students to select a topic regarding the process and memory management of some operating systems and, then, to write an exploratory reading summary. For example, a student compared distributed techniques, such as ODP, CORBA and DCE, in terms of functionality, application areas and advantage/weakness. Another student discussed why microprocessors cost less and perform better than mainframes.

### *Experimental design and procedures*

This study used an aptitude-treatment interaction design with feedback formats as the experimental factor and thinking styles as the aptitude factor. The median (mean = 3.683, *s.d.* = 0.67) of executive thinking style (Lin & Chao, 1999) was used to categorise subjects into high ( $n = 30$ ) and low ( $n = 28$ ) executive thinking style groups. They were then randomly assigned to two experimental groups, specific vs. holistic feedback (Table 1). Six peer assessors were randomly assigned to evaluate an assignment within the same experimental group.

**Table 1.** Experimental groups categorised by high versus low executive thinking style and specific versus global assessment criteria.

Aptitudes Treatments	High executive thinking style	Low executive thinking style
Specific Assessment Criteria	Group A $n = 14$	Group C $n = 13$
Holistic Assessment Criteria	Group B $n = 16$	Group D $n = 15$

During the pre-test, subjects answered the Thinking Style Inventory and were then assigned to four experimental groups. They also took the Motivational Strategies for Learning Questionnaire that was treated as a controlled variable.

Students in the specific feedback group had to provide feedback following six specific criteria, prompted by an assessment homepage. The six specific feedback criteria were as follows: (1) relevance of the project to the course contents (2) thoroughness of the assignment (3) sufficiency of the references (4) perspective or theoretical clarity (5) clarity of discussion, and (6) significance of the conclusion. Students in the holistic feedback group gave a total score and offered a general feedback for an entire assignment.

Peer assessment and *NetPeas* were introduced during the second week. The assignment was announced in the fourth week and students were asked to upload their assignments during the ninth week. Students assessed and provided feedback during the tenth week. The second round of peer assessment started from the

eleventh week in which students modified and resubmitted their assignments. During the twelfth week, the students assessed peers' resubmitted assignments. In the following week, the teacher and teaching assistant evaluated students' assignments and feedback quality and, finally, the post-test was distributed to students.

### Measurements

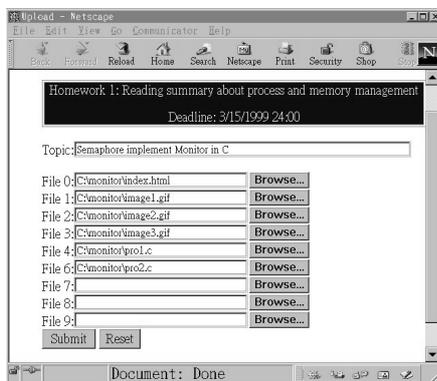
**Thinking style.** A Taiwanese version of Thinking Style Inventory (Lin & Chao, 1999) adapted from the original Thinking Style Inventory by Sternberg and Wagner (Sternberg, 1994) was used. Factor analysis revealed that Taiwanese students' thinking styles paralleled those of Americans'. The internal reliability ( $\alpha$ ) ranged from 0.79 to 0.92 in 13 subscales. 'I like situations in which my role or the way in which I participate is clearly defined' is one of the statements in the executive subscale. A typical statement of the legislative subscale is 'When making decisions, I tend to rely on my own ideas and ways of doing things.' 'When discussing or writing down ideas, I like criticising others' ways of doing things' is typical for the judicial subscale. Only the executive subscale was used in this study.

**Motivation: the controlled variable.** Based on the Motivational Strategies for Learning Questionnaire (MSLQ) of Pintrich *et al.* (1991), a Taiwan version of MSLQ (Wang & Lin, 2000) was used to measure the subjects' motivation. Factor analyses indicated that items were grouped to form six dimensions, intrinsic goal, extrinsic goal, task value, control belief about learning, self-efficacy for learning and performance, and test anxiety, similar to the results obtained from the American college students. The internal reliability of six motivation subscales ranged from 0.69 to 0.78.

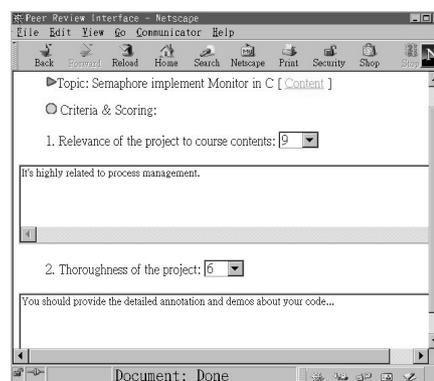
According to results of the pre-tests, subjects in the four groups scored similarly in six motivational subscales, i.e. intrinsic goal, extrinsic goal, task value, control belief about learning, self efficacy for learning and performance, and test anxiety (means = 3.12–3.46, *s.d.* = 1.01 – 1.72, all *ps* > 0.05).

### Web-based peer assessment system: NetPeas

The student interface of *NetPeas* offers four options: assignment uploading, assignment modifying, peer assessment, and complaint filing. Figures 1 and 2 display sample screens of this interface. The teacher interface offers three options: assignment assessment, feedback assessment, and complaint review.



**Fig. 1:** Assignment uploading with homework requirement at the upper middle part of screen



**Fig. 2:** Offers of specific feedback and rating peer assessment score

*Dependent variables*

For each student, four dependent variables were obtained through *NetPeas*. During the first round of peer assessment, each student received (1) an average assignment score assessed by peers (1st-round peer score) and (2) a feedback quality score given by the teachers (1st-round feedback quality). During the second round, students received (3) an average assignment score from peers (2nd-round peer score) and (4) another assignment score from the teachers (2nd-round expert score). In evaluating assignment and feedback quality, both students and the teacher used a 10-point Likert scale.

For each assignment, six peer assessment scores were averaged to increase its stability and reduce bias from each assessor. The teacher and TA (a computer science doctoral student well-trained in educational methodology) separately rated the students' assignments during the second round. Two experts' scores were highly correlated ( $r = 0.672$ , d.f. = 58,  $p < 0.001$ ). The expert assessment score did not correlate with the first-round peer assessment score ( $r = 0.232$ , d.f. = 57, *n. s.*), but it was significantly correlated with the second-round peer assessment score ( $r = 0.396$ , d.f. = 57,  $p < 0.01$ ).

Feedback is of high quality if it offers suggestions for the next step of modifying and explaining the peers' reading summary (Chi, 1996). The feedback quality scores rated by the teacher and the TA were significantly correlated ( $r = 0.703$ , d.f. = 57,  $p < 0.001$ ). Two examples of better quality feedback are as follows.

'You described functionality of some distributed techniques, such as ODP, CORBA, and DCE. Unfortunately, you did not mention the fundamental philosophy of these techniques. As a reader, other than specific functionality of each technique, I hope to know what particular problems can be solved by a technique and any other benefits of using it.'

'The discussion about how microprocessors cost less and perform better than mainframes provides only a single perspective of comparison. Microprocessors may not always surpass mainframes in all tasks and working areas. If you differentiate tasks and then illustrate the relative advantages and weaknesses of microprocessors and mainframes under each task, this essay can thoroughly compare small and large OS systems.'

*Perception towards NetPeas and web-based peer assessment*

Students were asked to fill in a questionnaire about the perceived effectiveness, adequacy, and merits/limitations of web-based peer assessment. Fifteen items were in the 5-point Likert scale, while others were open-ended questions.

**Results and discussion***Subjects in four groups achieved similar peer assessment scores during the first round before they offered feedback.*

The entry levels of subjects' learning capacities were examined. A two-way ANOVA (Table 2) revealed that in the first round, the interaction and main effects of executive thinking styles and feedback formats were not significant. Subjects in Group A, B, C, D achieved similar 1st-round peer scores (Means = 6.632, 6.447, 7.104, and 6.665,  $ps > 0.05$ ). This finding suggests either that all computer science undergraduates entered into peer assessment with similar learning capacities or achievements in all groups were similar before students activated higher order thinking to provide feedback.

**Table 2.** Descriptive statistics and two-way ANOVA summary table for the first round peer assessment score.

Round 1 Peer assessment	Executive thinking style				Total		
	High		Low		mean	s.d.	
	mean	s.d.	mean	s.d.			
Feedback Format	Specific	6.632	0.634	7.104	1.038	7.081	0.961
	Holistic	6.447	1.272	6.665	1.451	6.553	1.343
<b>Total</b>		6.533	1.013	7.083	1.334	6.798	2.232

Source of Variance	SS	Df	MS	F
Thinking Style	4.767	1	4.767	3.579
Feedback	4.232	1	4.232	3.178
Interaction (thinking style x feedback)	1.833	1	1.833	1.376
Error	71.916	54	1.332	
<b>Total</b>	82.163	57		

While the high executive students significantly improved over two rounds of peer assessment, the low executive students did not improve over the cycles.

A paired *t*-test indicates that all subjects' average 1st-round peer score (Mean = 6.798) was roughly the same as 2nd-round peer score (Mean = 6.925,  $t = -0.773$ , *n.s.* in Table 3). However, separate paired *t*-tests for four groups can more accurately illustrate learning growth across two rounds. Peer scores of group A and B significantly improved from the initial to the second round. For Group C, the peer scores of both rounds were approximately the same. In contrast, 1st-round peer score of group D was significantly higher than 2nd-round peer score.

All students' average peer score did not appear to improve from the initial to the final round. A further breakdown of group achievements revealed that, in receiving specific and holistic formats of feedback, the high executive thinkers performed better during the second round. The low executive thinking students in the specific feedback group, which were assumed to have a higher likelihood of activating higher order thinking, did not improve over two rounds but at least maintained the initial level of performance. In receiving holistic feedback, the low executive thinking students' performance declined over the two rounds.

**Table 3.** Mean peer assessment scores in round 1 and round 2 and paired *t* tests for four experimental groups and all students.

Peer Assessment	Round 1		Round 2		Paired <i>t</i> tests	<i>df</i>
	mean	s.d.	mean	s.d.		
All subjects	6.798	1.201	6.925	1.035	-0.773	57
Group A (H Executive/ Specific)	6.632	0.634	7.363	0.802	-6.147 **	13
Group B (H Executive/ Holistic)	6.447	1.272	7.240	0.625	-2.325 *	15
Group C (L Executive/ Specific)	7.104	1.038	7.442	0.700	-0.477	12
Group D (L Executive/ Holistic)	6.665	1.451	5.734	0.924	3.121 **	14

\*  $p < .05$  \*\*  $p < .01$

*The high executive thinking students, who tend to be more willing to follow instructional rules, offered better quality feedback than their low executive thinking counterparts.*

Students' capacities in offering critical feedback significantly differed. The two-way ANOVA (Table 4) indicates that only the main effect of executive thinking style was significant, while the two-way interaction effect and the main effect of feedback were not. This finding suggests that the high executive students (mean = 6.067) contributed substantially better feedback than the low executive thinkers (mean = 3.696,  $F = 23.434$ ,  $p < 0.01$ ). Restated, students who are more willing to follow instructional rules can provide better (suggestive and explanative) feedback when all subjects were assumed to be highly capable of using higher order thinking.

**Table 4.** Descriptive statistics and two-way ANOVA summary table for the first round feedback quality.

Round 1 feedback quality	Executive thinking style				Total	
	High		Low		mean	s.d.
	mean	s.d.	mean	s.d.		
Feedback Specific	6.679	2.207	3.692	1.251	5.241	2.339
Format Holistic	5.531	2.156	3.700	1.709	4.645	2.134
<b>Total</b>	6.067	2.220	3.696	1.487	4.922	2.232

Source of Variance	SS	Df	MS	F
Thinking Style	83.628	1	83.628	23.434 **
Feedback	4.680	1	4.680	1.311
Interaction (thinking style x feedback)	4.807	1	4.807	1.341
Error	192.707	54	3.569	
<b>Total</b>	283.901	57		

\*\*  $p < 0.001$

*During the second round of peer assessment, the high executive thinkers performed well regardless of whether in specific or holistic feedback groups. Although the low executive people maintained quality performance in the specific feedback group, their learning was impeded in the holistic feedback group.*

The interaction of thinking style and feedback was significant during the second round peer assessment ( $F = 15.258$ ,  $p < 0.01$ , see Table 5). The analysis of simple main effects (Table 6) revealed that the low executive students when receiving specific feedback (Group C, mean = 7.442) could modify previous assignments better than when receiving holistic feedback (Group D, mean = 5.734,  $F = 36.47$ ,  $p < 0.01$ ). Receiving holistic feedback, high executive thinkers (Group B, mean = 7.240) outperformed the low executive ones (Group D, mean = 5.734,  $F = 28.52$ ,  $p < 0.01$ ).

Subjects have a relatively lower likelihood of using higher order thinking in offering holistic feedback and have less critical information for further improvement when receiving them. The low executive thinkers in such a group could not achieve as well as the other three groups could.

**Table 5.** Descriptive statistics and two-way ANOVA summary table for the second round peer assessment scores.

Round 2 peer assessment		Executive thinking style				Total	
		High		Low		mean	s.d.
		mean	s.d.	mean	s.d.		
<b>Feedback</b>	<b>Specific</b>	7.363	0.802	7.442	0.700	7.297	0.703
<b>Format</b>	<b>Holistic</b>	7.240	0.625	5.734	0.924	6.527	1.189
<b>Total</b>		7.401	0.741	6.511	1.086	6.925	1.035

Source of variance	SS	Df	MS	F
<b>Interaction (thinking style x feedback)</b>	9.057	1	9.057	15.258 **
<b>Thinking Style</b>	7.333	1	7.333	12.354 **
<b>Feedback</b>	12.073	1	12.073	20.339 **
<b>Error</b>	32.054	54	0.594	
<b>Total</b>	61.072	57		

\*\*  $p < 0.01$ **Table 6.** Summary table of simple main effects on peer assessment in the second round.

Source of Variance	SS	Df	MS	F
<b>Feedback</b>				
within executive (H)	0.04	1	0.04	0.07
within executive (L)	21.65	1	21.65	36.47 **
<b>Executive</b>				
within feedback (S)	0.01	1	0.01	0.02
within feedback (H)	16.93	1	16.93	28.52 **
<b>Within cells</b>	32.054	54	0.59	

\*\*  $p < 0.01$ 

*The post-test questionnaire revealed that students were positive about web-based peer assessment and NetPeas.*

Subjects regarded *NetPeas* as a satisfactory system for uploading assignments (mean = 3.97), displaying assignments (mean = 4.30), giving feedback (mean = 4.42), and receiving feedback (mean = 4.26). In addition, they perceived peer assessment as an effective instructional strategy (mean = 3.98) although they were concerned with the adequacy of peer assessment as a single means of achievement evaluation (mean = 3.03).

The merits of peer review are listed below if the percentage of respondents was greater than 25%: higher likelihood of viewing and modelling peers' works (51%); more opportunities to improve original assignments through formative peer assessment (43%); receiving more feedback for improvement (38%); more opportunities to reconsider the adequacy of ones' own assignment while reviewing peers' assignments (37%) and enormous peer pressure so one must exert more effort (27%).

The limitations of peer assessment, as pointed out by more than 25% of subjects, were as follows:

- peer assessment is time and effort consuming (43%);
- peers may not have adequate knowledge to evaluate others' assignments (40%);
- although peer assessment was anonymous, students still hesitate to criticise (33%);
- peer feedback was often ambiguous or not relevant for further modification (26%);
- some students tended to give extremely low scores to others to keep their achievement at a relatively high level (25%).

### Conclusions

This study examined how feedback formats (specific vs. holistic) and executive thinking styles (high vs. low) affect learning in web-based peer assessment. The experimental design and the control over a possible interference, learning motivation, help clarify the causal relationship. Some important findings are summarised here.

First, when learning motivation was held constant, students in the four experimental groups performed roughly the same during the early stage of web-based peer assessment before offering feedback. Second, all students' average peer assessment score did not appear to improve from the initial to the final round. However, a further breakdown of group achievements revealed that the high executive thinking students significantly improved over two rounds of peer assessment, while the low executive students did not improve over the entire peer assessment.

Third, the high executive thinking students provided better quality feedback than their low executive counterparts. Fourth, in the second round of peer assessment, thinking style and feedback format interactively affected student learning. Low executive students receiving specific feedback significantly outperformed those receiving holistic feedback. In receiving holistic feedback, high executive thinkers outperformed their low executive counterparts. While high executive thinkers could probably overcome the disadvantages of holistic feedback, the low executive thinkers could not.

Fifth, subjects regarded web-based peer review as an effective instructional strategy and *NetPeas* as a satisfactory system. Effective elements of peer review that subjects acknowledged were cognitive modelling, self-improvement, reflective thinking, critical feedback, and peer pressure. Negative elements of peer review were time demands, peers' low evaluation capacity, ineffective feedback, and peer assessment integrity.

This study has provided further insight into individual difference and feedback effect on peer assessment that was not addressed in the review by Topping (1998). High executive students appear to benefit from peer assessment the most. They seem to be more willing to activate higher order thinking or adapt to new instructional strategies. Specific feedback is better than the holistic feedback format since more critical feedback, suggestions for improvement and explanations of original concepts are likely to be provided. The low executive thinkers are generally unwilling to activate higher order thinking, although they benefit from doing so.

### Limitations and suggestions

First, the experimental setting of this work was a Taiwanese research university. Subjects were capable computer science majors. Therefore, the generalisability of the results to less capable university students requires further examination. Second, unfortunately this experiment did not include a control group, thereby leaving some related issues unresolved. For example, subjects of the control group may revise their first draft as in the experimental conditions but do not receive peer feedback. For some students, revision would perhaps be sufficient to improve their assignments. Finally, this study did not gather expert assessments of the assignments for the first round. A direct comparison of expert ratings and peer ratings could broaden the validity of peer assessment.

Future studies should examine several issues more closely, such as the functions of peer assessment as a formative or summative evaluation method as well as reliability and validity in using peer assessment. The quality of various types of peer feedback (simple right/wrong, explanative, and suggestive feedback for example Chi, 1996) and how different types of feedback bring about revision are also critical issues. This deserves a closer examination with qualitative methods, such as content analysis for computer mediated communication (Henri, 1992; Hara *et al.*, 2000).

### Acknowledgements

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract Nos. NSC 89-2520-S-009-001 and NSC 89-2520-S-009-004. Dr Chin-Chung Tsai is also appreciated for his valuable comments.

### References

- Bangert-Drowns, R.L., Kulick, C.L.C., Kulick, J.A. & Morgan, M.T. (1991) The instructional effect of feedback in test-like events. *Review of Educational Research*, **61**, 213–238.
- Barrett, E. & Lally, V. (1999) Gender differences in an on-line learning environment. *Journal of Computer Assisted Learning*, **15**, 1, 48–60.
- Carson, J.G. & Nelson, G.L. (1996) Chinese students' perceptions of ESL peer response and group interaction. *Journal of Second Language Writing*, **5**, 1–19.
- Chi, M.T.H. (1996) Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, **10**, S33–S49.
- Crooks, T.J. (1988) The impact of classroom evaluation practices on students. *Review of Educational Research*, **58**, 45–56.
- Davis, R. & Berrow, T. (1998) An evaluation of the use of computer supported peer review for developing higher-level skills. *Computers and Education*, **30**, 1/2, 111–115.
- Downing, T. & Brown, I. (1997) Learning by cooperative publishing on the World Wide Web. *Active Learning*, **7**, 14–16.
- Dunn, R. & Dunn, K. (1999) *The Complete Guide to the Learning Styles Inservice System*. Allyn and Bacon, Boston.
- Dunn, R., Dunn, K. & Perrin, J. (1994). *Teaching Young Children Through Their Individual Learning Styles*. Allyn and Bacon, Boston.
- Dunn, R., Griggs, S.A., Olson, J., Gorman, B. & Beasley, M. (1995) A meta-analytic validation of the Dunn and Dunn model of learning style preferences. *Journal of Educational Research*, **88**, 6, 353–361.
- Fabos, B. & Young, M.D. (1999) Telecommunication in the classroom: Rhetoric versus reality. *Review of Educational Research*, **69**, 3, 217–259.
- Falchikov, N. (1995) Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, **32**, 175–187.
- Freeman, M. (1995) Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education*, **20**, 289–300.
- Hara, N., Bonk, C.J. & Angeli, C. (2000) Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, **28**, 115–152.
- Henri, F. (1992) Computer conferencing and content analysis. In *Collaborative Learning Through Computer Conferencing: the Najaden Papers*, (ed. A.R. Kaye) pp. 115–136. Springer, New York.
- Kulik, J.A. & Kulik, C.L.C. (1988) Timing of feedback and verbal learning. *Review of Educational Research*, **58**, 79–97.

- Kwok, R.C.W. & Ma, J. (1999) Use of a group support system for collaborative assessment. *Computers and Education*, **32**, 2, 109–125.
- Lin, S.S.J. & Chao, I.-C. (1999) *The Manual for use with Thinking Style Inventory-Taiwan version*. Unpublished manual. Institute of Education, National Chiao Tung University, Hsinchu, Taiwan.
- Lin, S.S.J., Liu, E.Z., Chiu, C.H. & Yuan, S.M. (2001a) Web-Based Peer Assessment: attitude and Achievement. *IEEE Transaction on Education*, **4**, 2, 211.
- Lin, S.S.J., Liu, E.Z. & Yuan, S.M. (2001b) Web peer review: The learner as both adapter and reviewer. *IEEE Transactions on Education*, **44**, 3, 246–251.
- Liu, E.Z.F., Lin, S.S.J., Chiu, C.H., & Yuan, S.-M. (1999) Student participation in computer sciences courses via the Network Peer Assessment System. *Advanced Research in Computers and Communications in Education*, **2**, 744–747.
- Mason, R. & Bacsich, P. (1998) Embedding computer conferencing into University teaching. *Computers and Education*, **30**, 3/4, 249–258.
- McDowell, L. (1995) The impact of innovative assessment on student learning. *Innovations in Education and Training International*, **32**, 302–313.
- Owston, R.D. (1997) The World Wide Web: a technology to enhance teaching and learning? *Educational Researcher*, **26**, 2, 27–33.
- Pintrich, P.R., Smith, D., Garcia, T. & Mckeachie, W.J. (1991) A manual for the use of the Motivated Strategies for learning Questionnaire. *Technical Report # 91-B-004*. School of Education, University of Michigan.
- Rada, R. (1998) Efficiency and effectiveness in computer-supported peer-peer learning. *Computers and Education*, **30**, 3/4, 137–146.
- Roth, W.M. (1997) From everyday science to science education: how science and technology studies inspired curriculum design and classroom research. *Science and Education*, **6**, 373–396.
- Rogoff, B. (1991) Social interaction as apprenticeship in thinking: guidance and participation in spatial planning. In *Perspectives on Socially Shared Cognition*, (eds. L.B. Resnick, J.M. Levine & S.D. Teasley) pp. 349–383. American Psychological Association, Washington, DC.
- Steeple, C. & Mayers, T. (1998) A special section on computer-supported collaborative learning. *Computers and Education*, **30**, 3/4, 219–221.
- Sternberg, R.J. (1994) Thinking styles: Theory and assessment at the interface between intelligence and personality. In *Personality and Intelligence* (eds. R.J. Sterberg & P. Ruzgis) pp. 169–187. Cambridge University Press, New York.
- Sternberg, R.J. (1998) *Thinking Styles*. Cambridge University Press, New York.
- Strachan, I.B. & Wilcox, S. (1996) Peer and self assessment of group work: developing an effective response to increased enrollment in a third year course in microclimatology. *Journal of Geography in Higher Education*, **20**, 3, 343–353.
- Topping, K. (1998) Peer assessment between students in colleges and universities. *Review of Educational Research*, **68**, 249–276.
- Van Lehn, K.A., Chi, M.T.H., Baggett, W. & Murray, R.C. (1995) *Progress report: Towards a theory of learning during tutoring*. Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA.
- Wang, S.-L. & Lin, S.S.J. (2000) *Cross culture validation of Motivational Strategies for Learning Questionnaire*. Paper presented at the Annual Meeting of the American Psychological Association, Washington, D.C.
- Yagelski, R.P. & Powley, S. (1996) Virtual connections and real boundaries: Teaching writing and preparing writing teachers on the Internet. *Computers and Composition*, **13**, 25–36.
- Zhao, Y. (1998) The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunications*, **4**, 4, 311–345.