

## The design and application of a web-based self- and peer-assessment system

Yao-Ting Sung <sup>a,\*</sup>, Kuo-En Chang <sup>b</sup>, Shen-Kuan Chiou <sup>b</sup>,  
Huei-Tse Hou <sup>b</sup>

<sup>a</sup> *Department of Educational Psychology and Counseling, National Taiwan Normal University,  
162, Heping E. Rd., Taipei, Taiwan*

<sup>b</sup> *Department of Computer and Information Education, National Taiwan Normal University,  
162, Heping E. Rd., Taipei, Taiwan*

---

### Abstract

This study describes the web-based self- and peer-assessments system, or the Web-SPA, which has been shown to provide teachers with a flexible interface with which to arrange various self- and peer-assessment procedures. Secondly, this study examines the effects of the application of the progressively focused self- and peer-assessment (PFSPA) procedures that are made available with the Web-SPA. In an evaluative study with 76 third-graders (14 or 15 years old) of junior school, it has been found that the upon completion of the PFSPA activities, the students demonstrate greater objectivity in their self-assessment scores; it has also been shown that significant consistency is found between the students' self- and peer-assessments results and that the assessment results of teachers. Furthermore, the quality of the students' works improved after the assessment activities.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Authoring tools and methods; Architecture of educational technology system; Self-assessment; Peer-assessment

---

---

\* Corresponding author.

## 1. Introduction

### 1.1. Self- and peer-assessment

Self- and peer-assessment refer to those activities of learners in which they judge and evaluate their own products of work and those of their peers with similar learning backgrounds. Both types of assessment emphasize students' active participation in the evaluation process. The two assessment methods are fast becoming more widespread and influential (Cizek, 1997; Paris & Paris, 2001; Shepard, 2000) through the influence of the learning culture, which emphasizes on active learning, participation in real-world practice, social interaction, self-monitoring and regulation (Greeno, Collins, & Resnick, 1996; McCaslin & Hickey, 2001).

In conventional classroom assessments teachers usually play a major role. In contrast, self- and peer-assessments require that students execute the assessment of their learning results and those of their counterparts. From a learning perspective, the two forms of assessment are special in that not only do they allow students to actively participate in the assessment, but also they provide them with opportunities to observe their own as well as their peers' works, thereby encouraging them to reflect on both the amount of effort they put into their own work and to judge the appropriateness of the standards they set for themselves, two processes which are reportedly helpful in nurturing self-monitoring and regulation among students. Though different researchers may use different procedures when conducting self- or peer-assessment, generally speaking, there are four popular processes: proposing and reflecting on the criteria for judging products; observing products; giving comments and suggestions; and students' (commentator) exchanging what they have learned in commenting. These processes have been demonstrated to produce beneficial effects in the cognitive, meta-cognitive and affective aspects of students' learning (Falchikov & Boud, 1989; Topping, 1998).

### 1.2. Enhancing the benefits of implementing self- and peer-assessment through computers

Traditionally, when teachers have conducted self- and peer-assessment activities, they have usually done so in a classroom situation using pen and paper. However, because implementing self- and peer-assessment involves complicated processes, using pen-and-paper-based methods can have some constraints. The possible constraints to pen-and-paper-based assessments may be compensated by technology. The first constraint is that there is a relative inconvenience of displaying multimedia works using pens and paper. However, in a computerized environment, text, sound and graphics are just a few of the types of input that can be presented in the web-based environment, which is multi-faceted and which can be expanded to great breadths. Secondly, computers can facilitate the efficiency in recording and compiling the results of scoring/commentary. In scoring or providing commentary of the works using pen and paper, limitations become obvious when multiple objects of assessment are dealt with (e.g., Billington, 1997; Hughes & Large, 1993; both studies involved assessing the works of all members in a class). This is further aggravated when multiple assessment standards are used. Printing a large quantity of forms for records is often required. Further, as it is difficult to make real-time compilations and summaries of the scoring results, students may have difficulty in getting immediate feedback on their personal scoring results or in having the general picture right after the scoring activity is completed. Existing database

technology, when incorporated with an appropriate user interface, can provide teachers and students with efficient tools for recording. Coupled with appropriate computation functions, it can even do speedy calculations as well as make quick summaries and presentations of the records of assessment so as to provide users immediate feedback. Those functions are valuable for reducing teachers' workload of preparing and implementing and students' performing the self- and peer-assessment activities.

Thirdly, web-technology can provide students with more opportunities of peer interaction beyond the constraints from time and locations. The essence of self- and peer-assessment lies in personal interactions, such as students' stating the strengths and weaknesses of each other's works and justifying the results of their evaluations. However, given the limited time of a class, even finishing the assessment activities itself poses a challenge, little time is left for discussion. The web-based environment is characterized by its accessibility at any time and its provision of synchronous and asynchronous methods of communication; therefore, it is possible to conduct activities either within the classroom or in after-class situations. With these advantages, it's possible to provide students with more opportunities of interaction beyond the limitation of time and locations.

Fourthly, we can increase the diversity of teachers' implementation of self- and peer-assessment. A remarkable variety of forms of implementing self- or peer-assessment are possible. In setting the standards of scoring, for example, there are the options of letting the teacher set the standards (e.g., [Magin & Helmore, 2001](#)) or allowing students to do it themselves (e.g., [Orsmond, Merry, & Reiling, 2000](#); [Stefani, 1994](#)). Regarding the number of evaluators versus assessee (the number of objectives, e.g., person or work, in scoring), the options of one-to-one (e.g., [Stefani, 1994](#)), one-to-many (e.g., [Hughes & Large, 1993](#)) or group-to-group (e.g., [Lopez-Real & Chan, 1999](#)) are all feasible. As for the form of scoring, it can be based on a marking scheme (e.g., [Magin & Helmore, 2001](#)), a measuring scale (e.g., [Rushton, Ramsey, & Rada, 1993](#)) or be in a written commentary format. The assessment activities may be a one-time event, as in most current studies, or repeated many times (e.g., [Sung, Lin, Lee, & Chang, 2003](#)). The works presented and the assessment results may either bear signatures or be anonymous (e.g., [Lejk & Wyvill, 2001](#)). When incorporated with an appropriate user interface, computers can provide teachers and students with various types of assessment and scoring and increase the flexibility of using different forms and procedures of self- and peer-assessment.

It is noteworthy that although computers may have advantages of supporting the implementation of self- and peer-assessment, these advantages can only be exploited under the condition that the obstacles of using computers in classroom context, like machine access ([Norris, Sullivan, Poirot, & Soloway, 2003](#)) or teachers' belief and capabilities of using computers ([Ertmer, 1999](#)), are removed. Designing appropriate tools for teaching and learning is a feasible approach to reducing the barriers encountered ([Marx, Blumenfeld, Krajcik, & Soloway, 1998](#); [Putnam & Borko, 2000](#)). Considering the possibilities of enhancing the effects of implementing self- and peer-assessment through computers, previous researchers have design some systems, like the MUCH designed by [Rushton et al. \(1993\)](#), the NetPeas designed by [Lin, Liu, and Yuan \(2001\)](#) and the SPARK by [Freeman and McKenzie \(2002\)](#). Although their systems do enable such basic functions as the uploading of assignments, scoring/commentary and the presentation of results, what they do not permit are the more advanced functions, like a wide array of optional methods of determining standards, methods of scoring and forms of grouping. Neither do those systems allow teachers the flexibility to arrange the workflow of the assessment

activities. For the most part, their systems still place considerable constraints on the arrangement of activities.

To exploit the advantages of computer technologies and help teachers and students implement self- and peer-assessment more easily, this study designs a web-based self- and peer-assessment system (Web-SPA). The contribution made by the Web-SPA is that it breaks through the restrictions mentioned above. In addition to the basic functions of uploading assignments, scoring/commentary and the presentation of results, the system provides various standards for determination, methods of scoring/commentary and numerous options for grouping. The dynamic choice mechanism for the workflow of the assessment activities provides teachers with a high degree of freedom to arrange activities based on their individual needs, thereby letting them stay in line with time allocations as well as specific course requirements and assignments. The following sections explain the structure and functions of the Web-SPA system.

## 2. Web-SPA: a web-based self- and peer-assessment system

### 2.1. The structure of the Web-SPA system

The Web-SPA has three main modules. The first module is the user interface for teachers and students; the second is the web server application program module, while the third is the database server. As demonstrated in Fig. 1, the various web server application modules access the data stored in the database server through the database access objects (i.e. operation files) and interact and communicate with users through the http protocol and the browser. Using Microsoft's active

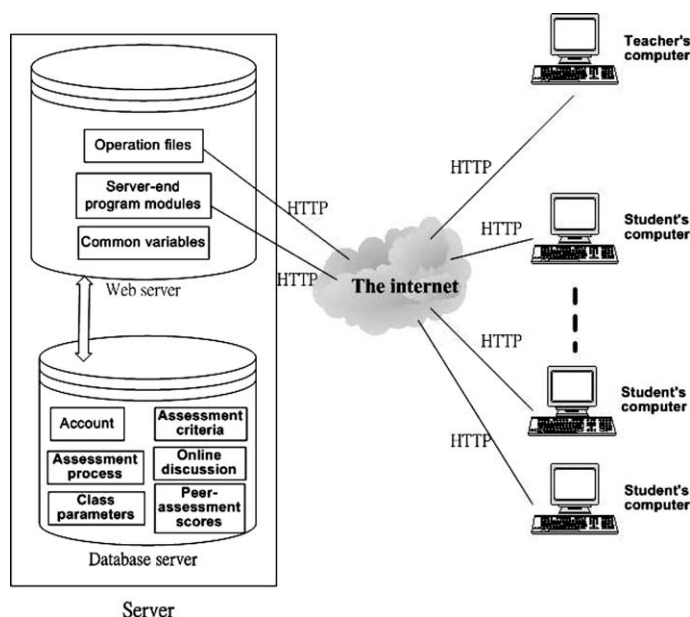


Fig. 1. The framework used by the Web-SPA system.

server pages (ASP) technology, the various server application modules interact and communicate with the forms of the browsers on the user end through the request and response objects; they provide a synchronous communication mechanism with the user end by accessing the memory on the server end using the application and session objects.

## 2.2. The functions of the Web-SPA system

### 2.2.1. The teacher management module

The teacher management module provides teachers with a convenient user interface that allows them to execute various setup and management functions online, such as setting up accounts, setting up class parameters, queries as to students' scoring process and observing various assessment results at any time they like. Web-SPA offers parameters for teachers to configure the options of various types of self- and peer-assessment activities (Figs. 2 and 3). In the section "Type of assignment and settings of the scoring criteria" teachers can choose three types of assignments. In the section "Type of scoring" teachers can choose three types of scoring/ranking methods – the Likert-type six-point scale; percentage; and reviewing with verbal comments (no scoring) – for students to determine the quality of the works. In the section "Assessment activities and procedures" teachers can choose and arrange preferred procedures for self- or peer-assessment. In the section "Random grouping", teachers can choose the customized functions to group students by themselves or let the system conduct random grouping after the teacher has assigned the number of groups. In the section

|   |  |
|---|--|
| Types of assignment   | <input type="radio"/> Group assignment, with the students setting the criteria<br><input type="radio"/> Group assignment, with the teacher setting the criteria<br><input checked="" type="radio"/> Individual assignment, with the teacher setting the criteria <input type="button" value="set criteria"/>   |
| Types of scoring  | <input checked="" type="radio"/> Six-point scale <input type="radio"/> Percentage <input type="radio"/> Observing (no scoring)   |
| Assessment activities and procedures                                    | <input type="checkbox"/> Assess one's own works<br><input checked="" type="checkbox"/> Peer-assessment or re-assessment<br><input checked="" type="checkbox"/> Examine results of in-group assessment results <input type="checkbox"/> Discussion required<br><input checked="" type="checkbox"/> Second-stage assessment<br><input checked="" type="checkbox"/> Second-stage assessment results <input type="checkbox"/> Discussion required<br><input type="checkbox"/> Inter-group discussion |
| Number of groups in the second-stage assessment                         | <input type="text" value="2"/>   |
| Works of higher scores to be assessed in the second stage               | <input checked="" type="checkbox"/>  |
| Works of lower scores to be assessed in the second stage                | <input type="checkbox"/>   |
| Random grouping   | Number of groups: <input type="text" value="5"/> <input type="button" value="OK"/>   |
| Display nicknames   | <input type="checkbox"/>   |
| URL address for downloading works (by system)                           | <input type="text" value="http://203.71.150.57"/>  |
| <input type="button" value="Save"/> <input type="button" value="Exit"/> |  |

Fig. 2. Class parameter setup interface (individual works).

|   |   |
|---|---|
| Types of assignment   | <input checked="" type="radio"/> Group assignment, with the students setting the criteria<br><input type="radio"/> Group assignment, with the teacher setting the criteria<br><input type="radio"/> Individual assignment, with the teacher setting the criteria  |
| Types of scoring  | <input type="radio"/> Six-point scale <input checked="" type="radio"/> Percentage <input type="radio"/> Observing (no scoring)  |
| Assessment activities and procedures                                    | <input type="checkbox"/> Proposing personal criteria<br><input checked="" type="checkbox"/> Proposing group criteria <input checked="" type="checkbox"/> Discussion required<br><input checked="" type="checkbox"/> Self-evaluating the work from one's own group<br><input checked="" type="checkbox"/> Peer-evaluating<br><input checked="" type="checkbox"/> Viewing the evaluating results of within-group members <input type="checkbox"/> Discussion required<br><input checked="" type="checkbox"/> Viewing the evaluating results of inter-group members <input checked="" type="checkbox"/> Discussion required<br><input type="checkbox"/> Inter-group discussion |
| Random grouping   | Number of groups: <input type="text" value="5"/> <input type="button" value="OK"/>  |
| Display nicknames   | <input type="checkbox"/>  |
| URL address for downloading works (by system)                           | <input type="text" value="http://203.71.150.57"/>   |
| <input type="button" value="Save"/> <input type="button" value="Exit"/> |   |

Fig. 3. Class parameter setup interface (group works).

“Display nickname” teachers can make the participants of the peer-assessment activities appear anonymously. Except for the functions above, teachers can also monitor students’ assessment activities via the interface of Web-SPA. For example, teachers can check the within-group and between-group evaluating results (Figs. 4 and 5).

### 2.2.2. The self- and peer-assessment module

The procedure for the students’ self- and peer-assessment activities in group collaboration are as follows (the example uses group works as targets with the teachers setting the assessment criteria, see Fig. 6): (1) Each and every student assesses his own work. (2) Peer-assessing or re-assessing. (3) Examining/discussing the within-group peer- assessment results. (4) Examining/discussing the results of peer assessments among groups. (5) Discussion among groups. In conducting individual assignment assessments, In order to reduce the workload and enhance the efficiency of executing the activities, the Web-SPA provides a progressively focused self- and peer-assessment framework (see details in the following section). The activities for assessing group-work are also modified. For example, Step 4 is modified as conducting the second-stage assessments: after the within-group self-assessments and peer-assessments in the first stage, the system selects some of the best and poorest works on the basis of the average scores (the number is determined by the teacher) as the targets for the second-stage assessment. Step 5 is modified as examining/discussing the results of the second-stage assessments: students can browse the scores and commentary of the second-stage assessments and hold discussions or not. Step 6 is modified as discussing among groups: the entire class performs peer-viewing and holds discussions on the best and/or poorest works.



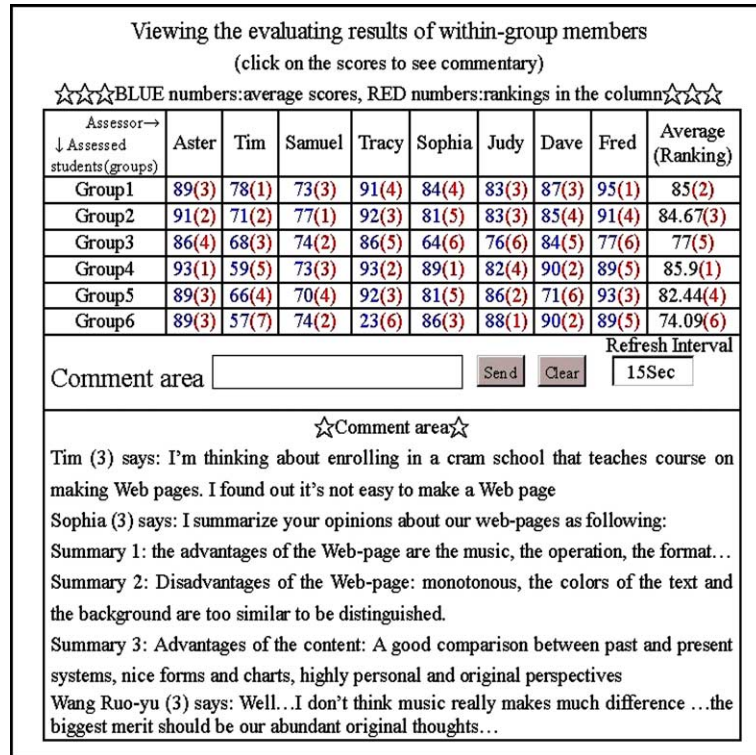


Fig. 4. Sample of the results and discussion of a within-group assessment.

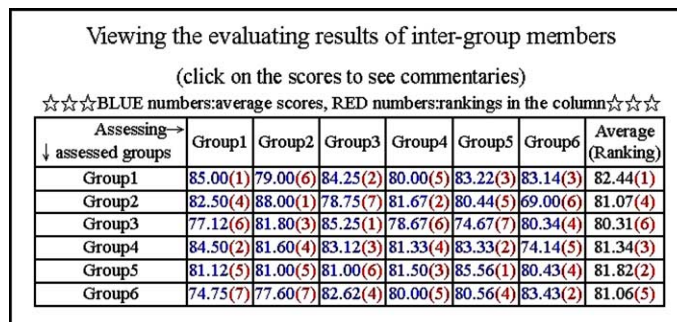


Fig. 5. Sample results of a between-group peer-assessment.

### 3. A study of progressively focused self- and peer-assessment

#### 3.1. Characteristics of the PFSPA

We propose a new approach to implementing self- and peer-assessment, named progressively focused self- and peer-assessment (PFSPA). Three characteristics are basic to the PFSPA

Peer Assessment System - Microsoft Internet Explorer

[Peer-evaluating](#)   [Viewing the evaluating results of within-group members](#)   [Viewing the evaluating results of inter-group members](#)   [Inter-group discussion](#)   [Logout](#)

### Performing peer-assessment or re-assessment

I want to assess group       Unfinished:

| Criterion   | Score                           |
|---|---------------------------------|
| Criterion 1 Attractive layout design (animation or graphics design)                                 | <input type="text" value="5"/>  |
| Criterion 2 Hyperlinks are clear and correct  | <input type="text" value="5"/>  |
| Criterion 3 Framework of the Web page content is clear and organized                                | <input type="text" value="9"/>  |
| Criterion 4 Content matches the title, and both content and title are immediately recognizable      | <input type="text" value="10"/> |
| Criterion 5 Substantive and well-informed reporting   | <input type="text" value="5"/>  |
| Criterion 6 The content of the has been fully digested and absorbed by the author before production | <input type="text" value="6"/>  |
| Criterion 7 The content contains the author's personal perspective                                  | <input type="text" value="5"/>  |
| Merits Clearly presented framework—one can immediately tell what the Web page is about              |                                 |
| Defects The Web page is still work in progress  |                                 |

(Recommendation) Note: the Advantage and Disadvantage fields must be filled out

Fig. 6. Stages in the modulated activities in peer-assessment.

procedures. Firstly, PFSPA emphasizes the integrated and recurring nature of the activities of self-assessment, the observation of works, peer-assessment and peer interactions. In conventional studies, self-assessment and peer-assessment are typically separated from each other and are generally only executed once during the process of the activities. Since the peer-assessment process involves observing the products of others, it is highly possible that these products provide students with reference messages, and maybe helpful for facilitating self-reflection, better monitoring and improved objectivity in scoring on the part of the evaluators. Besides this, seen from a learning perspective, changes in the results of the self- or peer-assessment activities may be an important indicator of the occurrence of learning because any changes in the assessment scores or criteria may well be indicative of changes in the assessor's degree of reflection and thinking in the process of assessment, resulting in new perspectives on both his own performance or on that of others. Therefore, the PFSPA emphasizes the integration of self-assessment, observation, peer-assessment and peer interaction, while encouraging learners to make further reflections on and modifications to the results of self- and peer-assessment induced by their heightened and more insightful perceptions after peer-viewing and interaction.

As the second characteristic, focus is on the increasingly sharp contrast in the quality of work during the process of undertaking the activities as this makes learners' sense of quality and awareness about making judgment increasingly better-tuned. In the PFSPA procedures, learners begin by observing works in a group randomly assigned by the system. The system then compiles the assessment results of within-group members. In this stage, the works for peer-viewing and assessment have not gone through any selection process, so contrasts in quality may be less striking. In the second-stage of activities, every learner is required to view and assess the best and/or poorest



works from other groups, and the differences in quality among the works obviously become more clearly observable. In the third stage, learners observe the best and/or poorest works selected by all in the second-round. In this stage, the works selected fall into the two extremities of the quality spectrum, and the contrast becomes the most distinct. Throughout the three stages of assessment, learners may acquire a more acute sense of the strengths and weaknesses of their own works and those of others as well as a more honed awareness of their own objectivity and fairness in their own scoring criteria in assessment.

The third characteristic of the PFSPA procedures is that it seeks a balance between the thoroughness of evaluating works and the economy of time allocation. Previous researchers have adopted many forms of assessor/assessee correspondence including one-to-one, one-to-many and group-to-group. Falchikov and Goldfinch (2000) have recently discovered that a single assessor's evaluation does not necessarily have less validity than those of multiple assessors in peer-assessment. In fact, when each member of a class is required to present his personal instead of a group work, whether there is a single assessor or there are multiple ones is likely to make a major difference in the efficiency and effects of execution. In the one assessor/assessee situation (e.g., Stenfani, 1993), the smaller demand on time and mental load may make the execution more efficient; however, the benefits may be limited owing to fewer observations and less feedback. In the other extreme case of many assessors/many assessees (e.g. Hughes & Large, 1993 in which each subject is asked to assess the performance of each member of a class of 30), while students may get better effects from peer-viewing and feedback, they are also confronted a greater demand in terms of time and mental energy, which might compromise the efficiency of the execution. The PFSPA takes an eclectic approach. The teacher considers the size of the class and the time available, and then decides the appropriate assessor/assessee correspondence. On the one hand, this increases the opportunities for the assessor to observe and receive feedback and, on the other hand, it requires less time and mental energy.

Take the example of a classroom with forty students, and each of them has one piece of work for evaluation. To save time and reduce workload, we randomly divided the students into groups of five to do self assessment. After that, in the first stage, the within-group members individually observed and evaluated the works of each other. The number of works viewed/assessed in this stage was 7. Then the system nominated one best and one poorest works on the basis of the assessment results from Stage 1. In Stage 2, the system randomly assigned two works (can be less or more, depending on teachers' decisions) from the best works category and two from the poorest category to each student for a total of 4 each (but not works from the student's own group) for peer-viewing/evaluating. Then the system sorted the results of the assessment. The number of works viewed/assessed by each student in this stage was 4. In the third stage, the students observed the best and the poorest works (one for each). The number of works viewed in this stage was 2. In the entire process, students may have been required to conduct repeated assessments on their own works or those of others if necessary. Even though there were 40 pieces of works in the beginning, the PFSPA allows for distribution through three stages. As a result, the number of works decreased, while the discrepancies among them in terms of quality increased, and the students ended up only needing to observe and evaluate 13 pieces of works.

In the following section, we present an empirical investigation about using PFSPA. We had three predictions: (1) The PFSPA should encourage students to reflect on and develop more objective views of their own works; as a result, the discrepancy between their personal self-assessment

scores and the scores from expert assessors should shrink. (2) The quality of students' revised works should be better than their original ones after the feedback from the PFSPA activities. (3) Significant consistency between the best and poorest works chosen by the students through the PFSPA and those chosen by experts is expected.

### 3.2. Method

#### 3.2.1. Participants

The participants in this study were 76 third-graders (14 or 15 years old, mean age was 14.7) in two classes of 37 and 39 students in a junior high school in Taipei County. Of the 76 students in all, 43 were male and 33 female. Most of the students came from districts of middle or high social/economical status. Because most students in the two classes aim to continue their senior high school education, their academic achievements are above average. The students were asked to participate in this study as a partial fulfillment of their "Computer and Information Science" course.

#### 3.2.2. Tools/materials

*Multimedia web pages.* Each of the participants was required to submit his works on creating a web-page design. The subject of the web design was restricted to the introduction to the cultural heritage of Taipei County. The students had to collect the materials needed for the projects on their own, but they were given information about the scoring criteria in advance. The scale for assessing the web-page designs: The teacher set up five criteria for assessing the quality of the students' web pages based on the progress of the course and the students' level of knowledge. The students used the percentage format, with each criterion accounting for 20 points out of 100. The five criteria were: the frames of the web pages are complete and correctly designed; the layout design is aesthetically pleasing and appealing; the hyperlinks are clearly presented and correct; the content of the web pages is substantive and comprehensive; and the marquee tool in the frontpage system is complete and correctly used. Students' works and the scales for assessment were uploaded to the Web-SPA before the assessment activities. The form of the assessment adopted in this study can be described as: using individual works, using the percentage scoring scheme and text commentary, anonymous, repeated progressive assessments.

#### 3.2.3. Procedures

The participants were told to prepare for their work on their web-page designs three weeks prior to the end of the course and were asked not to look at the works of other students because theirs had to be original. The students uploaded their works to a website via the Web-SPA system three days before the start of the assessment. On the day of the evaluation activities, the two classes separately conducted the following activities (lasting for around 100 min.) in eight steps: (1) The teachers informed the students that they were about to conduct self- and peer-assessments of their works and that the objectivity of their assessments was important since it would become one of the basis for the teacher's scoring the web-page design works. (2) Using the Web-SPA system, the teacher randomly divided the students into seven groups for each class, with five or six members in each. Each student began by conducting a self-assessment of his own work using the given criteria. (3) Working in groups, the participants observed and assessed the works by the other members of their group (four or five pieces of work). If the students became dissatisfied with

their previous self-assessment scores after the initial peer-viewing activity, they were allowed to change them here. (4) The system compiled the results from each member's scoring of his own works and those scores by the other members of the group, including the scores on each of the works and the commentaries. The group members studied the information compiled by the system and held discussions on it in their group. (5) The system nominated the best and poorest works in each group, making for a total of 7 best and 7 poorest works from each class. Then, it randomly assigned two "best-in-the-group" works and two "poorest-in-the-group" works to each student (excluding the works from the student's own group) for assessment (due to time constraints, the students only assessed four pieces of works instead of all the best and poorest works). (6) After the activity was completed, the system collected the scores on the best and poorest works given by all the students and calculated the average scores. After sorting the scores, the system presented the best and poorest works of the class to each student. (7) The students did the last self-assessment activity on their own work. (8) The teacher gave a commentary on the process of the students' activities and then asked the students to modify the contents of their own web pages in accordance with the commentary they had collected. The revised works were submitted online a week later. The students were informed of their scores only after the assessment of all of their revised works had been completed by teachers.

Based on the procedure described above, there were a total of 37 pieces of work from one class and 39 from the other. But through the PFSPA procedures, each of the students was only required to undertake peer-assessment on about 12 pieces of works.

### 3.3. Results

This study collected and compiled the participants' data of self-ratings in three stages: before observing others' works, the last self-rating after the first within-group observing/evaluating (having viewed four or five pieces of works), and the last self-rating after the last observing works (having viewed the best and the poorest one piece of work in the class). The results of the three stages of self-ratings are shown in Table 1. The average values were obtained by adding the scores of each participant in the five criteria, and then divided by the number of students in the class. The

Table 1  
Students' self-ratings and deviation scores in the three stages and two experts' ratings of students' works

| Classes       |    | Students' self-ratings |                  |                        |                  |                        |                  | Experts' ratings |                 |
|---------------|----|------------------------|------------------|------------------------|------------------|------------------------|------------------|------------------|-----------------|
|               |    | 1st stage <sup>a</sup> |                  | 2nd stage <sup>b</sup> |                  | 3rd stage <sup>c</sup> |                  | Original version | Revised version |
|               |    | Ratings                | Deviation scores | Ratings                | Deviation scores | Ratings                | Deviation scores |                  |                 |
| A<br>(N = 37) | M  | 68.27                  | 16.91            | 70.68                  | 15.85            | 72.78                  | 12.85            | 79.36            | 82.92           |
|               | SD | 24.73                  | 20.62            | 22.70                  | 16.91            | 22.43                  | 17.93            | 22.39            | 20.18           |
| B<br>(N = 39) | M  | 63.85                  | 14.81            | 67.62                  | 13.68            | 64.90                  | 10.18            | 70.78            | 75.04           |
|               | SD | 24.87                  | 16.19            | 23.17                  | 10.05            | 26.10                  | 10.36            | 22.53            | 16.93           |

<sup>a</sup> Self-ratings before observation.

<sup>b</sup> Self-ratings after within-group observation/evaluation.

<sup>c</sup> Self-ratings after between-group observations.

study also collected students' works on their web-page designs, including both the pre-PFSPA activities version and the post-activity version. Two teachers of the Computer and Information Technology course gave scores on the basis of the five assessment criteria, without being told which was the revised version; the results are shown in Table 1. The Pearson's product-moment Correlation of the two experts' scoring for Class A and Class B are, 0.94 and 0.92, respectively ( $p < 0.01$ ).

### 3.3.1. Comparison of students' self-ratings: pre- and post-PFSPA

To determine whether participants performed more objectively in their self-assessments of their own works? after the PFSPA activities, we compared students' self-ratings in the three rounds with the average of the assessment scores by the two experts, by calculating the differences between the self-ratings of each student and the average of the expert assessment scores (deviation scores, see Table 1). The smaller the difference was, the smaller the deviation was between the student's self-scoring and the experts' scoring, and therefore the more objective the score was. The results of one-way repeated measures analyses of variance show that for Class A, though the deviation scores show a steady decrease in the three rounds, they are not significantly different ( $F(2, 72) = 2.11, p > 0.05$ ). However, the deviation scores for the three rounds of assessment in Class B reach a marginally significant level ( $F(2, 76) = 3.00, p = 0.056$ ). The post hoc comparison using the Tukey HSD method indicates that the deviation scores are not significant between the stages of before-observing and after within-group observing/evaluating ( $p = 0.84$ ). Neither is the difference between the stages of after within-group observing/evaluating and after inter-group observing significant ( $p = 0.19$ ). The difference between the stages of before-observing and after inter-group observing, however, is significant ( $p = 0.056$ ). Table 1 also shows that the deviation value in the third stage is smaller than those values in the first and second stages. In other words, after doing the PFSPA activities, students' self-ratings are closer to the two experts' ratings.

### 3.3.2. Comparisons of students' works

pre- and post-PFSPA: To determine whether the quality of the participants' works change before and after they participate in the PFSPA activities, we conducted repeated measures analyses of variance on the two teachers' averaged scores of the two versions of works (see Table 1). The results show that the scores received for the two works of Class A students are significantly different ( $F(1, 36) = 34.20, p < 0.01$ ); likewise, the scores for those of Class B students are significantly different ( $F(1, 38) = 16.32, p < 0.01$ ). Table 1 also shows that the teachers gave significantly higher scores to the revised works than to the original works. Hence, it is reasonable to state that students' works on the web-page designs demonstrated significant improvement after the self- and peer-assessment activities.

### 3.3.3. Consistency between the results of students' self- and peer-assessments and experts' assessments

In the PFSPA procedures, the repeated self- and peer-assessments ultimately determined the best and poorest works, with the number of these corresponding to the number of groups. In this study, the classes started out evenly divided into seven groups for the within-group self- and peer-assessments; then the groups selected the best and poorest works to participate in a class-wide assessment. The last stage produced the seven best and seven poorest works in the class. The seven

Table 2

Consistency of teachers' and students' categorization of students' quality of web-page works

| Students' categorization | Teachers' categorization |      |      |         |      |      |
|--------------------------|--------------------------|------|------|---------|------|------|
|                          | Class A                  |      |      | Class B |      |      |
|                          | Good                     | Fair | Poor | Good    | Fair | Poor |
| Good                     | 4                        | 3    | 0    | 4       | 3    | 0    |
| Fair                     | 3                        | 19   | 1    | 3       | 18   | 4    |
| Poor                     | 0                        | 1    | 6    | 0       | 3    | 4    |

Students' categorization is based on the results of self- and peer-assessment. The best seven pieces and poorest seven pieces of work selected by students after peer assessment are labeled as good or poor, respectively. Other works are labeled as fair. Teachers' categorization is based on teachers' scores on the works; the top seven pieces of work are labeled as good, the bottom seven pieces of work are labeled as poor, others are labeled as fair.

pieces of works in the “best-among-peers” (good) category, seven in the “poorest-among-peers” (poor) category, and the other ordinary (fair) works (23 pieces in Class A, 25 pieces in Class B) were nominated on the basis of the students' self- and peer-assessments. How consistent were they with the results of expert assessments and selection? To answer this, we calculated the average of the scores given by the two expert teachers, and then sorted the results. The seven pieces of works with the highest scores were found to be in the good category, while the seven with lowest scores were treated as they were in the poor category, and the other were treated as fair. The results of the comparison by category of the student and teacher assessments are shown in Table 2. It is evident that the categorization by the students in Class A has a 78.38% level of consistency with that of the teachers, which compares with 66.67% for Class B. A Kappa consistency analysis of the results in Table 2 shows that the Kappa coefficients for Classes A and B are, 0.60 and 0.38, respectively; both surpass the significance level of 0.01.

#### 4. Discussion

In that teachers are the ones who are actually carrying out formal instruction in schools, they must play a critical role in helping instructional technology exert a substantial influence on changing the forms and contents of teaching. But the inherent shortcomings in the existing applied instructional technology often discourage teachers from using it in their instruction and tend to produce little benefit even when it is applied (Cuban, Kirkpatrick, & Peck, 2001; Ertmer, 1999). This study postulates that one of the major obstacles has been the unavailability of software suited for teachers' needs. Teachers need to have access to appropriate software to help them teach or undertake assessment activities – software that has a user-friendly interface and well-designed on-demand functions to help teachers arrange their teaching activities, while reducing their workload as well as helping students achieve greater learning benefits. If such technological products are available to teachers, it is likely that teachers will be more willing to incorporate information technology into their instruction and that they will be more motivated to make the best use of it to enhance their existing teaching practices.

Based on the assumptions above, this research designs and presents here the Web-SPA and related assessment activities. Through the friendly user interface and versatile functions of the

Web-SPA, teachers may design self- and peer-assessment procedures that can meet their demands and needs in the classroom with a high degree of flexibility. Compared to similar systems, such as the MUCH (Rushton et al., 1993) and NetPeas (Lin et al., 2001), the Web-SPA is more convenient in its application, and in turn, makes the implementation of self- and peer-assessments in class easier and more flexible. During the processes of conducting assessments, we have also found that some shortcomings of previous pen-and-paper-based studies, such as the high costs for preparing paperwork, or receiving few comments from peers owing to the fact that students may distance themselves from their own works (Billington, 1997), may be substantially reduced by the Web-SPA's mechanisms for presenting and evaluating products, providing immediate feedback, displaying nicknames and so on.

In this study, we also find that using the Web-SPA provides more diversity in the forms of assessment than do the traditional pen-and-paper self/peer-assessments. Currently, most studies related to self/peer-assessments have been based on perspective of effectiveness, focusing on exploring the reliability and validity of self- and peer-assessment (e.g., Falchikov & Goldfinch, 2000; Hughes & Large, 1993; Shore, Shore, & Thornton III, 1992). Besides, most self- or peer-evaluations have only been conducted once in their assessment activities. We propose the PFSPA procedures implemented through the Web-SPA system. The empirical evaluation confirms the possible benefits of this integrated and recursive approach. For example, previous research indicated that it is not uncommon that students may over-assess or under-assess their own works (Lejk & Wyvill, 2001; Saavedra & Kwun, 1993; Sung et al., 2003). In the empirical evaluation, we found that after the PFSPA, there is a tendency for the discrepancy between students' assessments of their own works and experts' assessments to decrease. Furthermore, when the number of works increases in the last stage of peer- viewing and assessment, the tendency toward a declining discrepancy becomes even more pronounced. We may assume that in the PFSPA activities which present increasingly stronger contrasts with respect to quality between the works in each round of peer-reviewing, students may find more reference messages, and as a consequence, tend to become more objective in assessing their own works.

Moreover, this study finds that the quality of students' works shows significant improvement after the PFSPA activities. This indicates that the opportunities for mutual learning and feedback during the self- and peer-assessment process not only stimulate reflection and judgment on the quality of their own works, but also help students obtain more information as the basis on which to modify their own works. These findings also support the argument of previous studies that self- and peer-assessments help students learn autonomously and reflectively (Ross, Hogaboam-Gray, & Rolheiser, 2002; Sung et al., 2003; van Kraayenoord & Paris, 1997).

From the perspective of the effectiveness of self- and peer-assessments, this study also finds that there is remarkable consistency between the students' assessment results and teachers' assessment results. There is close association between the best and poorest works selected by the students through the self- and peer-assessment procedures and the selection by the experts. The finding echoes those of previous studies that there is a significant degree of consistency between the results of students' peer-assessments and those of experts' (Falchikov & Goldfinch, 2000; Hughes & Large, 1993; Stefani, 1994). It also supports the assumption that progressive self- and peer-assessment procedures can adequately reflect the correctness of the assessment results.



## 5. Conclusions

Self- and peer-assessments are methods of evaluation highly recommended by researchers, but the traditional pen-and-paper-based self- and peer-assessment method has many restrictions. This study contributes to the field by designing the Web-SPA, and proposes the PFSPA procedures expecting that they will help minimize the difficulties that teachers currently encounter during self- and peer-assessments and increase the possibility that in future, other unique, innovative types of self- and peer-assessment methods can be adopted. The design of the system and the proposal of the assessment procedures presented here are just for demonstration or to serve as examples. We expect that these examples will stimulate more in-depth thinking as well as further development on ways to integrate web technology with instruction/assessment models.

## Acknowledgements

Parts of the evaluation study of the Web-SPA system was supported by Grants from the National Science Council, Taiwan (NSC90-2520-S-003-013) and the Project of Pursuing Academic Excellence, Ministry of Education, Taiwan (89-H-FA07-1-4).

## References

- Billington, H. L. (1997). Poster presentations and peer assessment: Novel forms of evaluation and assessment. *Journal of Biological Education*, 31, 218–220.
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroad. In G. D. Phye (Ed.), *Handbook of classroom assessment*. New York: Academic Press.
- Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal*, 38, 813–834.
- Ertmer, P. A. (1999). Addressing first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47, 47–61.
- Falchikov, N., & Boud, D. (1989). Student self assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395–430.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.
- Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student team work: Benefits of evaluating across different subjects. *British Journal of Educational Technology*, 33, 551–569.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology*. New York: Prentice-Hall.
- Hughes, I. E., & Large, B. J. (1993). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education*, 18(3), 379–385.
- Lejk, M., & Wyvill, M. (2001). The effects of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment and Evaluation in Higher Education*, 26, 551–561.
- Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking styles. *Journal of Computer Assisted Learning*, 17, 420–432.
- Lopez-Real, F., & Chan, Y. R. (1999). Peer assessment of a group project in a primary mathematics education course. *Assessment and Evaluation in Higher Education*, 24, 67–79.

- Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they?. *Studies in Higher Education*, 26, 287–298.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (1998). New technologies for teacher professional development. *Teaching and Teacher Education*, 14, 33–52.
- McCaslin, M., & Hickey, D. (2001). Educational psychology, social constructivism, and educational practice: A case of emergent identity. *Educational Psychologist*, 36, 133–140.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 25, 23–38.
- Norris, C., Sullivan, T., Poirot, J., & Soloway, E. (2003). No access, no use, no impact: Snapshot surveys of educational technology in K-12. *Journal of Research on Technology in Education*, 36, 601–612.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36, 89–101.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29, 4–15.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5–6 mathematics effects on problem-solving achievement. *Educational Assessment*, 8, 43–59.
- Rushton, C., Ramsey, P., & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: A case study. *Journal of Computer-Based Instruction*, 20, 75–80.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78, 450–462.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4–14.
- Shore, T. H., Shore, L. M., & Thornton III, G. C. (1992). Construct validity of self- and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology*, 77, 42–54.
- Stefani, L. A. J. (1994). Peer, self, and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19, 69–75.
- Sung, Y. T., Lin, C. S., Lee, C. L., & Chang, K. E. (2003). Evaluating proposals for experiments: An application of Web-based self- and peer-assessment. *Teaching of Psychology*, 33, 331–334.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.
- van Kraayenoord, C. E., & Paris, S. G. (1997). Children's self-appraisal of their worksamples and academic process. *Elementary School Journal*, 97, 523–537.