Chapter IX

Testing the Validity of the Post and Vote Model of Web-Based Peer Assessment

Bruce L. Mann, Memorial University, Canada

Abstract

Two tests of validity were conducted with undergraduate education students on the post and vote model of Web-based peer assessment. Validity was determined by calculating a Pearson product-moment correlation and corresponding coefficient of determination that compared the average grade assigned by the pre-service teachers with the grade assigned independently by the course instructor. Results of both studies showed that the post and vote model of Web-based peer assessment were valid with these groups, and generalizable to undergraduate classes engaged in similar tasks.

Introduction

The post and vote model is a method of collecting and analyzing peer assessment data using the variety of features provided in off-the-shelf Web tools. This model emerged in the late 1990s, at a time when colleges, universities, and training organizations had
already adopted a Web-based training platform and were becoming acquainted with the features provided in the Web tools. These Web tools included: a bulletin board for posting and replying to discussions, a student viewing area, a questionnaire tool, a message compiler, online chat, student progress tracking, group project organization, student self-evaluation, grade maintenance and distribution, access control, navigation tools, auto-marked quizzes, electronic mail, automatic index generation, course calendar, student homepages, and course content searches (Mann, 1998a, 1998b, 1998c).

Today colleges, universities, and training organizations around the world are using Web-based training platforms to offer students an online education. Since these tools are customarily grouped together under a course name and protected by a password, they can be treated as elements of a “system,” a “Web course management system” (WCMS) (Mann, 1999a, 1999b, 2000a), consistent with systems theory, and the post and vote model is a subsystem, or more accurately a modeling subsystem of the WCMS. More discussion about the post and vote system follows.

**Trial-and-Error Learning and Bulletin Board Data**

Throughout the late 1990s and 2000, publicity and sales promotions in print and on the Web claimed that their Web tools required minimal technical skill, and said they preferred to let educators apply their own methods of course design to their Web courses (Mann, 1998a, 1998b, 1998c). Consequently: 1) the bulletin board tool was overused, with high volumes of student data being saved in the discussion board for reading and grading; and 2) a trial-and-error method of learning Web tools became the de facto method of preparing instructional materials for student learning on the Web, a method which was later re-defined as “an instructor’s gradual phasing-in to Web course management behavior” (Mann, 2000a, p. 23).

**Instructional Design Inertia**

Around the same time that most of us were sifting through hundreds of student discussions and learning Web-tools by trial and error, familiar models and theories of instructional design were being abandoned by practitioners, and when they were used, were regularly misapplied (Gros, Elen, Kerres, van Merrienboer, & Spector, 1997). According to some experts, instructional design models and theories had become static (Boshier et al., 1997), inert (Yang, Moore & Burton, 1995), unusable (Wild & Quinn, 1998), and simply not workable (Winn, 1997) for prescribing interactive learning.

In sum, this was a busy time for instructors, a lack of enthusiasm for applying familiar models and theories of instructional design to Web-based learning, little real Web tool assistance to count on, and hundreds of board postings to read and analyze.
Phase Theory and the Post and Vote Model

It was under these conditions that two frameworks were introduced to help to better explain the situation. First, “phase theory” was published, a descriptive framework for analyzing the Web tools and strategies that an instructor already knows how to use. A detailed explanation of phase theory is available in Mann (1999a, 1999b, 2000a). Second, the “post and vote model of Web-based peer assessment” appeared to help reduce the perceived instructional design inertia and overuse of the bulletin board tool, and settle in to learning a few specific Web tools. A more systematic method of collecting and analyzing the data was needed to manage the high volume of student postings in the discussion board. A key component of the post and vote model was using more of the Web tools together as a system of tools and strategies for the purpose for which they were intended. Another key component was student involvement in the assessment process, which already had a 30-year history and was becoming more prominent (see Falchikov & Goldfinch, 2000; Sluijsmans, Saskia, & van Merriënboer, 2002).

Peer Assessment:
The Benefits and Limitations

Student peer assessment is the process where students evaluate one another’s work (Topping, Smith, Swanson, & Elliot, 2000). Peer assessment requires that students exchange assignments, discuss responses, and rate one another’s work using scoring systems or rating scales devised by the teacher (Airasian, 2001).

Benefits of Peer Assessment

The peer assessor learns from critically analyzing and evaluating other students’ work. The student assessed learns from peer feedback. Peer assessment has the following advantages in face-to-face learning settings (Blumhof & Stallibrass, 1994):

- Helps develop evaluative thinking skills.
- Helps focus on criteria that improve the learning outcomes.
- Helps to motivate students’ own sense of assessment as a result of:
  - Negotiating and agreeing on criteria
  - Understanding the system
  - Having a share in the marking.
• Provides an opportunity to judge their own work and hence improve their own performance because of raised self-awareness of how assessment is undertaken, and what is assessed.
• Heightens awareness of a teacher’s skills. For instance, marking seminars focuses students not only on the content, but also on a range of skills students might otherwise not notice.
• Enhances concentration and increases the learning potential of a situation because students are intimately involved in the processes. By intimate we mean that emotions and values are engaged in an analytic task. The process includes planning and reflection.
• Provides a built-in emphasis on planning and reflection that is of particular value to reflective learners.

Many studies show that peer assessment can help learners develop critical, evaluative, and analytical skills (Anderson, Howe, Soden, Halliday, & Low, 2001; Falchikov, 1986; Pond, Rehan, & Wade, 1995; Searby & Ewers, 1997; Topping et al., 2000). Peer assessment that involves students’ ability to make value judgments, analyze responses, and provide reasoned arguments on other students’ work is beneficial for both the peer assessor and the student being assessed (Falchikov, 1995; Freeman, 1995). Anderson et al. (2001) found that college students who were involved in peer-based critiquing exercises showed significant improvement in their oral and written work in terms of providing justification of their arguments, compared to those in the control group. Falchikov (1986) reported that undergraduate science students involved in peer assessment found that the assessment process “made them think more, learn more, and become more critical and structured” (p. 161). Searby and Ewers (1997) stated, “The most significant reason for the introduction of peer assessment is that it helps students to think critically and to take control of their learning so that they are less dependent on the instructor” (p. 372). In sum, peer assessment that involves students’ ability to make value judgments, analyze responses, and provide reasoned arguments on other students’ work is beneficial for both the peer assessor and the student assessed (Falchikov, 1995; Freeman, 1995; O’Donnell & Topping, 1998).

Limitations of Peer Assessment

Although peer assessment can facilitate many positive outcomes, it can also raise issues of personal anxiety and exposure. In face-to-face teaching conditions, students have said they felt that peer assessment was threatening (Blumhof & Stallibrass, 1994):

• Changes the culture. Students expect the teacher to be the expert and judge, as this is what they have come to expect in their educational experience. Peer assessment is seen to be changing a well-worn and tried ‘traditional system’ and shifting the balance of power from course instructor to students. Though the shift at present is absolutely minimal, the change in culture makes it destabilizing.
Challenges the contract. In economic terms, the informal contract for many students is that students will pay their fees and do a reasonable amount of work, and the course instructor will design a course, teach it, and pass a high proportion of students. Some feel badly treated when the contract is changed.

Causes personal anxiety. Students have reported that they did not feel they had the appropriate skills to undertake the assessment, and this increases anxiety.

Raises problems of exposure. Students are not trained in our social relations to give and take constructive criticism without it feeling like a personal attack on them as individuals. Students dislike the exposure of publicly offering critical feedback as much as they fear receiving it. The main fears that they have heard voiced are that:

- Peer pressure. If you give offence, there is the fear that you may be ostracized from the group.
- Friendship marking. Questions can be raised about the fairness of the marking scheme. Marks may be skewed through friendships or hostilities. Peer assessors also have a tendency to “over-mark” — assign higher marks relative to the course instructor-assigned marks — students’ work (Falchikov, 1986; Kelmar, 1993; Pond et al., 1995; Mowl & Pain, 1995; Rushton, Ramsey, & Rada, 1993; Sluijsmans, Moerkerke, Dochy, & van Merriënboer, 2001).

Makes more visible the inherent difficulties in all marking schemes. These difficulties include:

- The tendency to mark the more easily measurable.
- A tendency to mark towards the middle to avoid obvious offence. Peers are reluctant to indicate weaknesses (critical comments) in their assessment of other students’ work (Falchikov, 1995, 1996; Topping et al., 2000).
- The tendency to be affected by stereotyping and expectations without the safety valve of anonymity marking.
- Exposing the subjectivity of marking schemes by trying to decide on, and weight, criteria.

This inconsistency in peer marking may affect the reliability of the peer assessment process. It was under these conditions that the post and vote model of Web-based peer assessment was introduced, to describe a systematic method of using Web tools and strategies together to collect and analyze student peer assessment data (Mann 1999a, 1999b, 2000a).
The Post and Vote Model

The post and vote model is a partially open, four-step system of Web-based peer assessment. In this research, the post and vote model is both a focal system of peer assessment itself, as well as a subsystem of the Web course management system (WCMS). As the focal system of Web-based peer assessment, the post and vote model readily employs generic, off-the-shelf Web tools provided in most WCMSs, such as BlackBoard (Pittinsky & Chasen, 1997), WebCT (Goldberg & Salari, 1996), or any one of a number of others (Mann, 1999a). For a current listing of WCMSs, consult the EduTools Web site (WCET, 2003). This model requires no extra computer programming, patches, add-ons, or third-party software.

Figure 1 shows a graphical representation of the four steps in the post and vote model, described as a “partially open system” of Web-based peer assessment. The arrows show the direction of input, output, and throughput (processing) between the four elements in the post and vote system. The dashed line surrounding Figure 1 separates the elements “inside the system” from those “outside the system” (in the environment, other systems).

According to systems theory, every system strives toward equilibrium as it constantly fluctuates between too much and too little information. For this reason, every system is necessarily concerned with the communication and control of variety inside and outside itself. “Variety” in this sense is a measure of the number of distinct states a system can be in at any given time. A system can thrive, barely survive, or even die from its variety. “Requisite variety” is a measure of capability inside and outside the system. The more variety of actions available to a system, the larger the variety of perturbations it is able to compensate for from elements inside and outside itself.

Figure 1. A graphical representation of the steps in the post and vote model of Web-based peer assessment
“Systems” are either “closed,” “open,” or “partially open” (Heylighen, Joslyn, & Turchin, 2002). A WCMS is a “partially open system,” only partially open because access to its data and Web tools is protected by a password, making the WCMS “closed” to the general public; yet learning in a WCMS is usually dependent on hypertext links to the outside Web sites or tools, or access to e-mail systems, chat, or other communication tools on other servers. As a partially open system, students and instructors will sometimes require tools from the parent WCMS or those from other Internet systems to improve requisite variety. Filters such as the discussion board and survey tools help to limit the variety entering the system; amplifiers such as e-mail, ftp, and instant messaging help to increase variety, leaving the system for other systems outside itself (in its environment).

The best test of an educational system is the benefit it brings to students and instructors using it. The survival of the post and vote system of peer assessment depends on the requisite variety among the peers working with the WCMS; that is to say, the variety of knowledge and skill of instructor and student assessors together must be greater or at least equal to the variety in each student assignment being assessed. Using another educational metaphor, a teacher working in a school system must possess the requisite variety of knowledge and skill to thrive or at least survive within the variety of knowledge and skill of her students. Group work and role-playing are two well-known filters for maintaining the teacher’s requisite variety in the classroom. Useful amplifiers for increasing teacher variety and maintaining requisite variety inside and outside the classroom include photocopies, homework Web sites, and class e-mail lists. Details on requisite variety and other systems concepts can be accessed from the Principia Cybernetica Web site.

Step 1: Typing into the Board

The student’s first step in using the post and vote model of Web-based peer assessment is to access the discussion board tool, type-in questions and comments about the assignment, and tryout preliminary ideas for possible submission. Even at this early stage, student contributions or “participation grades” can be assigned, their postings to online conferences (the discussion board tool) evaluated using any one of a number of methods of analyzing the discourse. Henri (1992), for example, developed a widely used content analysis model. Five key dimensions were identified for analysis of online discussions, namely: 1) participation rate (e.g., raw number and timing of messages); 2) interaction type (e.g., direct response, “in response to the posting…”); 3) social cues (e.g., “It’s my birthday today”); 4) cognitive skills (e.g., judgment “I disagree with…” and depth of processing (surface-level or deep-level processing); and 5) meta-cognitive skills and knowledge (e.g., providing examples and relating to situations). Hara, Curtis, and Angeli (2000) paralleled Henri’s (1992) recommendations of content analysis in online discussions and proposed more elaborate guidelines to analyze electronic conversations. Their study focused on the social and cognitive processes demonstrated by the students during online discussions. Rose (2002) and Brown (2002) used Henri’s content analysis model in their doctoral research. There are other models, however. For a recent overview of issues and trends in Web-based discourse analysis research, see Rourke, Anderson, Garrison, and Archer (2001).
Step 2: Student Uploads

The second step for the student using the post and vote model is to upload their html-ed assignments into a common viewing area for subsequent viewing by other students. Most WCMSs support individual or group uploads and allow the designer to designate group membership. Alternatively, the WCMS can automatically divide the class into groups, given the desired group size. Students might collaborate (using the e-mail and document-sharing tools) to critique a flow chart or write a research paper or proposal, then “html it” and electronically upload the html file into their student viewing area. Only group members can edit their own content. Once all the papers are uploaded into the student viewing area using the appropriate tool, students are requested to read the submissions of their classmates. Each student uses the survey tool to anonymously grade and comment on other students’ presentations. Student peer assessors can be made “accountable,” told in advance that the instructor will be grading the quality of their peer assessments, and that this grade will be counted toward their final mark. In a recent experimental comparison of accountability (more accountable vs. less accountable) and anonymity (anonymous vs. named) in Web-based peer assessment, Wadhwa, Mann, and Schulz (2004) found that more accountability improved the quality of student peer comments.

Step 3: Student Votes and Comments

Student “voting” is the third step for the student using the post and vote model. Student “voting” in this context means “assigning a number and writing a comment.” The process of assigning a mark occurs as follows: First, individual students view each uploaded assignment in the student viewing area. Once the individual student views an uploaded assignment (e.g., a student Web site developed, a case analyzed, a research paper critiqued, etc.), they open the survey tool to assign a mark and justifying comment to the assignment. Likewise, the instructor independently views the uploaded assignment, and then opens the survey tool to assign a mark and justifying comment to the assignment.

Step 4: Concatenated Votes and Comments

The fourth step for the student using the post and vote model of Web-based peer assessment is to wait for the concatenated votes and comments from instructor and peers.

Figure 2. Post and vote formula for calculating a student’s grades on a Web-based submission

\[
\text{Student Grade} = \frac{\text{Instructor or T.A.} + S1 + S2 + S3 + S4}{4} \\
\]

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
in e-mail. Meanwhile the instructor uses the compile tool in the WCMS to display the concatenated votes and justifying comments. If the class has small numbers, the instructor can highlight and paste the entire page of student peer assessments (grades and justifying comments) into a word processor, and attach the file to the submitting student’s e-mail.

The “final mark” on the assignment is determined as the average of the instructor’s and students’ marks. Figure 2 shows the post and vote formula for calculating a final grade for a student’s educational Web site based on the course instructor’s and student peer assessments. Interpretations of validity, reliability, and generalizability from the data are presented later in this chapter.

Two important components of successful application of the post and vote model are peer anonymity and peer accountability. Anonymity and peer accountability in the post and vote model have been found to reduce peer over-marking and improve the quality of peer comments (Wadhwa et al., 2004).

### Anonymity in Peer Assessment

In the post and vote model, student peer assessors as well as those being assessed can be either anonymous (their names replaced by numbers) or can remain named. Some studies suggest that anonymous peer assessment may reduce social pressure, resulting in peers providing marks that are more accurate and critical comments (Bostock, 2000; Davis, 2000; Falchikov, 1995; Haaga, 1993; Tsai, Liu, Lin, & Yuan, 2001).

There are several different types and degrees of anonymity: complete anonymity, where all participants are completely anonymous to each other; one-way anonymity, where one of the participant’s identities is concealed; and social anonymity, where the absence of social presence depersonalizes individual’s identity. Therefore, an individual’s identity need not be concealed to create anonymity.

The absence of social presence in written electronic communication can lead to social anonymity (Berge & Collins, 1993; Brodia, 1997; Bump, 1990; Liu et al., 2001; Jonassen & Kwon, 2001). However, in graduate online educational courses, students may know each other from another face-to-face course or another online course. Therefore, in graduate online educational courses, social absence may not cause social anonymity. Hence, the effect of friendship and social interactions resulting in peer over-marking and reluctance in providing critical comments remain the same in a face-to-face and online peer assessment. Whereas peer assessment in an online learning environment can encourage students to provide more comments, it may also invite social loafing that can affect the quality of those comments. Zhao (1998) found anonymity to be a “double-edged sword in collaborative learning.” He concluded that while anonymity allowed participants to be more critical, it made them work less.
Peer Accountability in Peer Assessment

Incorporating peer accountability in online peer assessment may reduce social loafing (Davis, 2000; Tsai et al., 2001). Tetlock (1983) defined accountability as “a special type of transmission set in which one anticipates the need not only to communicate one’s opinions, but also to defend those opinions against possible counterarguments.” Topping (2000) suggested that assigning a greater sense of accountability and responsibility on peer assessors may affect the quality of peer comments. Studies on the effect of accountability on student responses found that participants put in more cognitive effort in their responses when they are asked to justify their comments or when others review their decision (Gordon & Stuecher, 1992; Price, 1987). Davis (2000) reported that peer assessors took greater care in marking, since they (peer assessors) knew that they were being assessed on their ability in marking other students’ work (p. 351). In Tsai’s (2001) study, quality of peer assessors’ comments was reviewed and graded by the instructor. This was done to encourage assessors to provide helpful comments. Gordon and Stuecher (1992) examined the differences in students’ responses on teacher evaluation based on degree of accountability. In their study, students were asked to complete two closed-ended questions and one open-ended question evaluating their professor. They found that the linguistic complexity of student responses increased in a high accountability condition, in which they were asked to submit their responses to the faculty as compared to low accountability condition, in which they were asked to submit their responses to their peers.

The real proof of the peer assessment process however, and the post and vote model of Web-based peer assessment in particular, resides in the testing. Two tests of validity were conducted with undergraduate education students (pre-service teachers) on the post and vote model of Web-based peer assessment. Peer accountability in online peer assessment may provide sufficient peer commentary.

Study 1

Participants

The participants in Study 1 were pre-service teachers (n = 39) enrolled in a core course on the Bachelor of Education (Primary/Elementary) program at the largest Canadian university in the Atlantic region. The program was designed to prepare K-6 teachers in a specific discipline, as well as pedagogy courses over five years. A core course on this program was called “Computers and Learning Resources for Primary/Elementary Teachers” with the purpose of teaching pre-service teachers how to integrate computer software and other learning resources into their teaching. Laboratory components were scheduled so that students would learn how to use and implement communications, applications, and curricular software.
Materials

These students used modern computing facilities over a high-speed network designed to maximize the computing experience. A learning resources library provided students with the textual and supplementary materials they required for their course work, including current school texts, as well as teachers’ curriculum guides, children’s literature, resource books, and educational software. Students also had access to a variety of media equipment, educational videos, and multimedia kits. Students were expected to attain high performance levels. This teacher preparation program encouraged examination and discussion of significant educational issues within a framework of critical reflection and analytical practice. Under these conditions, each participant was required to develop a educational Web site of five Web pages, in accordance with two sample educational Web sites and the course instructor’s rubric.

Research Design

In this study, the correlation pre-experimental ex post facto research design described in Cohen, Manion, and Morrison (2000) was implemented to test the validity of the post and vote model of Web-based peer assessment by pre-service teachers. Validity was determined by Pearson product-moment correlations and corresponding coefficients of determination. Working from a rubric, each student developed a paper-based mock-up for an educational Web site, then a five-page online prototype from the mock-up, followed by four peer assessments of other student-made Web sites. The post and vote model was tested with peer assessments of student-made Web sites.

Instrumentation

The instrument for collecting data was a class assignment completed by students as part of their regular course work. The students’ used a rubric for peer assessment of student-developed Web sites developed from other rubrics that had been designed for a similar purpose (e.g., Schrock, 2004). Students were asked to score four of their peer’s Web sites using this rubric out of a maximum score of 30: content and ideas (/10), organization (/5), language use (/5), presentation (/5), technical (/5). Student-assessors were required to use the WebCT Survey tool to enter their peer assessments. Student assessors were made “accountable,” told in advance that the instructor would grade the quality of their peer assessments, and that this grade would be counted toward their final mark. Student accountability in Web-based peer assessment can improve the quality of peers’ comments (Wadhwa et al., 2004).
**Procedure**

The procedure in Study 1 lasted just under six weeks, and was conceptualized in three distinct steps, namely: (1) the paper-based mock-up, (2) the five-page online prototype from the mock-up, and (3) peer assessment.

**Step 1. The paper-based mock-up.** Each student created a “paper-based mock-up” of an educational Web site, essentially Web page frames on paper, as outlined in Brown and Mann (2001). These paper frames were intended to help them think through their educational Web site design, to provide a platform for active discussion and debate about the content. A “Web page frame” is the appropriate chunk of information for presentation to their students at one time. In the computer lab, students were shown how to make the opening frame for their educational Web site, add Web page frames, each one with a title zone, a graphic zone, a text zone, a zone of hypertext links, and a mail zone. Students were advised that the “name, title zone” of their educational Web site should attract and hold their student’s attention. Students were encouraged to include an interesting picture or drawing to evoke curiosity in their students. A “picture or drawing zone” should appear at the top or side of the screen, far above the need to scroll. An “instructional content zone”—also called “task-oriented directions zone”—should provide a goal, learning objective, or challenge for their students. Students were told that it was very important that their educational Web site include a zone for “hints, fault-free questions, or partial answers” to one or two questions, and that their “links zone” should contain only one idea that required the student to click to get a partial answer. An “e-mail or chat zone” close to the bottom of each Web page should get their students sharing information about the content or task with one another in class or from home.

During this time, these students were encouraged to think creatively about educational Web site development. They were encouraged to access the Bulletin Board, post their questions and comments about the assignment, and tryout preliminary ideas for possible submission. Contribution or “participation” grades were also assigned at this stage of the process. Student postings to online conferences were evaluated using the widely used content analysis model by Henri (1992).

**Step 2. The five-page online prototype from the mock-up.** Students were required to transform their paper mock-up into an “five-page online prototype from the mock-up” by developing html documents from paper frames, and uploading them to the Web. Each student uploaded his or her submission for peer assessment using the ftp program provided in the WCMS. Most WCMSs will support individual student uploads and allow the designer to designate group membership. Once all the submissions were uploaded into a public viewing area, students were requested to read the submissions of their classmates.
Step 3. Web-based peer assessment. Each student was required to assess four of their colleague’s educational Web sites, working from a rubric that was adapted from the university regulations:

“Excellent performance with clear evidence of: comprehensive knowledge of the subject matter and principles treated in the course, a high degree of originality and independence of thought, a superior ability to organize and analyze ideas, and an outstanding ability to communicate.” (Section 8.2 University Regulations, University Calendar, 2004-2005)

Using a survey tool, each student voted and commented anonymously on four peer submissions. To clarify: “voting” in this sense meant “assign a mark.” Similarly, the course instructor independently assigned a mark and justifying comment to each submission using the “Survey” tool. Again, everyone was anonymous—student peer assessors, their names replaced by numbers, and those being assessed. Student assessors were also accountable, told in advance that the quality of their peer assessments would be graded by the course instructor and their grade counted toward their final mark. Students were informed of their peer’s grades and comments in a word file attached to an e-mail from the course instructor. Using the “Compile” feature, the course instructor concatenates individual student’s votes and justifying comments. The course instructor pasted the entire results page for each submission (i.e., the final mark and justifying comment) into a word processor and attached it to each student’s e-mail.

Data Collection and Analysis

In Study 1, survey data were generated from individual students clicking on the Web-based survey tool and entering a grade and comments for four classmates. Their student ID number identified them as peer assessor and the student being assessed, so that the assessor and the assessed were unknown to one another (in the context of peer assessment). All students in the study were told that the course instructor would grade their assessments on the quality of feedback on the students’ assignment. Figure 2 shows the Post and Vote formula for calculating a final grade for a student’s educational Web site based on the course instructor’s and student peer assessments.

<table>
<thead>
<tr>
<th>Assessment Content</th>
<th>Assessment Rubric</th>
<th>Course Level &amp; Sample Pre-service teachers (n = 39)</th>
<th>Agreement and Shared Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Web site development</td>
<td>Instructor’s rubric</td>
<td></td>
<td>r = 0.745**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>r² = 0.555</td>
</tr>
</tbody>
</table>

Table 1. Correlations of course instructor and average peer assessments using the Post and Vote method of peer assessment in Study 1
Results and Inter-Rater Reliability

Results of the analysis of Study 1 are summarized in Table 1. The correlations between the course instructor’s scores and the average student peer assessment scores of educational Web sites (n = 39) was very high, at \( r = 0.745, p = .000, \) with almost 56\% of shared variance explained.

Inter-rater reliability between the course instructor and an independent coder was assessed by Pearson Product Moment correlation and found to be significant with the course instructors at \( r = 0.961, p=.000. \) Coding reliability was determined by having a different coder trained to the coding rubric, and then independently code 25\% of educational Web sites. The rater was a Fellow of the School Graduate Studies whose thesis in peer assessment had won the 2004 College Teachers Scholarship.

Study 2

Participants

The participants were an entire class (n = 66) from the same population of pre-service teachers enrolled in a Teacher Education program described in Study 1.

Materials

The materials given to participants in Study 2 were similar to those in Study 1, in that these studies were assigned the same sample educational Web sites and the same course instructor’s rubric. The task assigned to students in this study was more challenging than Study 1 because, in addition to developing an educational Web site consisting of five Web pages, each student had to explain how they developed their Web site on Explorer video.

Research Design and Instrumentation

Study 2 used the same pre-experimental correlation \textit{ex post facto} research design that was implemented in Study 1, except that students assessed one another’s verbalizations about how they were developing their educational Web sites. Working from an instructor-made rubric, each student assessed four of their colleague’s Explorer videos, according to the procedure recommended in Mann (1998, 1997, 1996).

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
Procedure

Study 2 used a procedure similar to Study 1, except that students assessed one another’s Explorer videos, instead of their Web sites. The procedure in Study 2 lasted eight weeks and was conceptualized in four distinct steps, namely: 1) the paper-based mock-up, 2) the five-page online prototype from the mock-up, 3) student-videotaped verbal explanations and demonstrations of how they had developed their educational Web site, and 4) peer assessment. Student assessors were also accountable, told in advance that the quality of their peer assessments would be graded by the course instructor and their grade counted toward their final mark.

Data Collection and Analysis

In Study 2, as in the first study, survey data were generated from individual students clicking on the Web-based survey tool, and entering a grade and comments for four classmates. Their student ID number identified them as peer assessor and the student being assessed, so that the assessor and the assessed were unknown to one another (in the context of peer-assessment).

Results

Summary results of the analysis of Study 2 are shown in Table 2. Analysis of the data revealed that the strength of agreement between course instructor’s scores and student peer assessment scores of the Explorer videos (n = 66), was highly significant, at $r = 0.701$, $p = .000$, with over 49% of shared variance explained.

General Discussion

Results indicated that the computed average of grades and justifying comments assigned by students highly correlated with grades assigned independently by the course instructor. The statistical reliability of this research was high. The calculations were a

<table>
<thead>
<tr>
<th>Assessment Content</th>
<th>Assessment Rubric</th>
<th>Course Level &amp; Sample</th>
<th>Agreement and Shared Variance</th>
<th>$r = 0.701^{**}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explorer video of verbalizations</td>
<td>Course instructor’s rubric (n = 66)</td>
<td>Pre-service teachers</td>
<td>$r^2 = 0.491$</td>
<td></td>
</tr>
</tbody>
</table>
measure of inter-rater reliability insofar as different independent judges or raters evaluated the student work. Most researchers report one of four different methods of determining inter-rater reliability, either: Cohen’s Kappa, Kendall’s coefficient of concordance, the intra-class correlation, or the Pearson product-moment correlation (Huck, 2004, p. 84). Rourke et al. (2001), for example, make an argument for using the Cohen’s Kappa chance-corrected measure of inter-rater reliability over other statistics. Cohen’s Kappa would be a good choice for analysis of the nominal data in Step 1 of the post and vote model, where raters would have to classify each bulletin board discussion into one of Henri’s (1992) five key categories of online discussion, either: participative, interactive, social, cognitive, or meta-cognitive. At Step 4 of the post and vote model however, the judges’ ratings are all raw scores from 0 to 30, and therefore best analyzed by the Pearson product-moment correlation, the most frequently used correlational procedure in educational research (Cohen et al., 2000; Huck, 2004). Reliability is a necessary, but insufficient condition for validity. Validity refers to the degree to which the research accurately reflects or assesses the specific concept that the researcher is attempting to measure. Three types of validity are considered below: face, construct, and content.

The face validity in both studies was high. Face validity is concerned with how a measure or procedure appears. Relative to other methods mentioned earlier, the post and vote model of Web-based peer assessment was a reasonable way to gain assessment information. The interactivity required in the post and vote model answers Schuttler and Burdick’s call (in press) for a faculty model that promotes a facilitative relationship, and an interactive environment for students that can enable a sense of closeness that supersedes distance. Furthermore, the formative feedback aspect of the post and vote model fits Morgan and O’Reilly’s criteria (in press) that a model should provide for student input and negotiation in assessing products and processes. The post and vote model is also consistent with Popham’s (2004) view of assessment as “a formal attempt to determine student status with respect to educational variables of interest” (p. 6). “Student status” in this context was their ability to peer assess their classmates online in the same way they would in a classroom setting—“to make structured judgments about the quality of the work produced by members of your own peer group (such as classmates) and, through raised awareness, increase their ability to self-assess” (Blumhof & Stallibrass, 1994, p. 4).

Construct validity was good in both studies, meaning close agreement between the theoretical concept under investigation and the specific measuring device or procedure. In Study 1, the theoretical concept was “the student-developed Web site,” Web pages completed by students as part of their regular course work, following a procedure published in Brown and Mann (2001). Web site development occurred at Steps 1 and 2 of the post and vote model (see Figure 1). As an educational activity, developing Web sites are believed to have value.

“Building Web pages is among the most constructivist activities that learners can be engaged in, primarily because of the ownership that students feel about their products and the publishing effect” (Jonassen, Peck, & Wilson, 1999, p. 28)
The specific measuring device used by students at Step 2 of the post and vote model was a rubric for assessing student-developed Web sites that was developed from other rubrics that had been designed for a similar purpose (e.g., Schrock, 2004). In Study 2, the theoretical concept was the student verbalizations self-recorded on Explorer Center videos, and Explorer Centers the specific measuring device or procedure. Explorer Centers have been found to be less intrusive than individual workstations, a means of collecting verbal protocols in the absence of the investigator’s tape recorder, and more accurate than conventional transcriptions of observations (Mann, 1998, 1997, 1996). Finally, content validity was assured. Content validity is the extent to which the measurement reflected the intended content domain. The content domain in this research was limited to “the process of developing a student-developed Web site” (Study 1) and “student verbalizations and demonstration of the process of developing Web sites captured on Explorer video” in Study 2.

**Research Design Limitations**

A limitation of the correlation ex post facto research design was that models of online peer assessment could not be directly compared, and therefore claims could not be made about the post and vote model relative to any other models of assessment or learning. Further analyses of these data should include experimental comparisons of the post and vote model with MUCH, NetPeas, and especially an “All-in-the-Discussion Board” condition. Other quantitative studies should look at “grading-over-time intra-rater agreement,” the relationship between the grades and the comments assigned by a peer on a student’s assignment, while carrying out non-anonymous peer-assessment. Qualitative studies could explore “comment-grade intra-rater agreement,” the relationship between the grades and the comments assigned by a peer on a student’s assignment, while carrying out anonymous peer assessment.

**Conclusion**

Both studies in this chapter supported the stated hypothesis, namely: that the post and vote model of Web-based peer assessment would be valid with these groups and generalizable to undergraduate classes engaged in similar tasks. Furthermore, the post and vote model was shown to be a direct way of using a variety of Web tools suited to the tasks for which they were designed, with the result of high validity of student assessment and democratization of the assessment process. Finally, it appears that familiar models and theories of instructional design are now being augmented by research in cognitive load theory (Paas, Renkl, & Sweller, 2003) and multimedia learning (Mayer, 2003; Mayer & Moreno, 2003). It seems that in general educators are becoming more socialized to the new tools and strategies of the trade. It appears too that we may soon count on better Web tool assistance from platform developers.
The other hypotheses, however, will certainly be tested at another time. Some hypotheses will be investigated experimentally and in the court of public opinion. Other hypotheses will be observed under ex post facto conditions. The advent of evolving hardware and software will surely give us more adaptive and nonlinear interactions, and a capacity for more sophisticated Web-supported assessment. In any case, it seems likely that student assessment of Web-based learning will always be a busy time for instructors.

References


Mann, B.L. (1998a, November). Working through phases: Instructional design in WebCT. *Proceedings of the Meeting of the School of Communications and Multimedia, Edith Cowan University, Perth, Australia.*

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.


Mann, B.L. (2000a). Phase theory: A teleological taxonomy of Web course management (pp. 3-26). In B.L. Mann (Ed.). *Perspectives in Web course management*. Toronto, ON: Canadian Scholar’s Press.


Wadhwa, G., Mann, B.L., & Schulz, H. (2004). *Effects of anonymity and peer-accountability during peer assessment in a graduate Web-based education research*


