Peer Grading over the Web: Enhancing Education in Design Courses

Edward F. Gehringer
North Carolina State University
efg@ncsu.edu

Abstract

We have implemented a peer-grading system for review of student assignments over the World-Wide Web and used it in approximately eight computer science and engineering courses. Students prepare their assignments and submit them to our Peer Grader (PG) system in the form of one or more Web pages. Other students are then assigned to review the pages, with the submitter’s grade being assigned by the peer reviewers. Authors and reviewers can communicate via a shared Web page during the review period, where they post comments and answer questions in a double-blind fashion. Authors are notified by e-mail when a new review of their submission has been posted. They can update their submission in response to the review.

We have experimented with a variety of approaches to facilitating good communication, including multiple deadlines for feedback per review period, and a scheme that rewards reviewers for helping their authors improve their submission. The educational experience has been very rewarding, for several reasons: It makes it more cost-effective to assign design problems, which are very time-consuming for a single individual to review. It forces students to learn to write clearly for their peers, since their grade depends upon it. It can be used to generate problems for future homework and tests, by assigning students to make up a problem involving the course material. It can be used to generate resources for students in a course, as students can be assigned to browse the Web for further material related to each lecture. It promises a scalable solution to managing large courses, because the work of the course staff increases less than linearly with the number of students.

1. Introduction

As technology marches onward toward the 21st century, the rapidity of change forces educators to revise their curricula frequently, while high-tech industry seeks increasingly large numbers of graduates to foster further innovation. Both of these factors increase the demands on engineering faculty, who are faced with large numbers of students and large amounts of new material. To meet these demands, faculty need to get students involved as never before in their own education, helping to educate their peers. For if faculty provide an environment where students are actively engaged in teaching other students, members of large classes may actually benefit from more, rather than less, personalized instruction.

For generations, the academic community has relied on peer review as a way of enhancing the knowledge base and encouraging serious scholarship. It has been praised as a cooperative learning technique. However, the mechanics of peer review have required too much paper-
shuffling to make it practical as a classroom strategy. The era of networked computing—and the World-Wide Web, in particular—has changed all that. In recent years, electronic peer review has made its way into the classroom, especially in writing classes, to eliminate the need for instructors to read and grade hundreds of student essays. However, until now, the tools have rarely if ever supported the kind of work students do in engineering classes.

Peer review in engineering has the same benefits as peer review in writing courses. That is, it gives students experience in communicating with the kinds of audiences they will need to write for on the job after graduation. It allows reviews to be done and grades assigned with minimal faculty involvement. It helps mitigate the problem of inadequate TA support.

But peer review in engineering helps in several other ways. It encourages instructors to assign more design problems, which are very important in an engineer’s education, but very time consuming to grade adequately. In addition, the best work done by students can be turned into resources to help future classes learn. For example, students can be assigned to write research papers on various topics, with several students writing on the same topic. The most highly evaluated paper on each topic can then be presented to the next class of students as background reading on that topic. The writers can be asked to include liberal doses of Web hyperlinks in their papers, so that later students can read not only their work, but also the analyses of experts. Finally, students can be asked to compose problems as well as papers; the best of these can then be assigned to later classes as homework or test questions.

2. The Implementation: PG

Beginning in early 1996, I and my students developed PG, a peer grader program written in Java, which provided for submission of homework and peer review over the Web. In overview, students submit their work over the Web. Reviewers are assigned semiautomatically by the instructor. Reviewers and authors communicate double-blindly via a shared Web page. At the end of the review process, the reviewer assigns a grade to each author whose work (s)he has reviewed. A student’s grade is the average of the grades given by the reviewers, plus an incentive described below to encourage good reviews.

In more detail, students authenticate themselves to the PG system by typing their user-ID and password. The instructor may assign these, but there is also a Kerberos interface so that the student may use his/her login/password for the campus computer system.

A student entering the PG system (Figure 1) has a choice of whether to submit a new page or review pages submitted by others. If (s)he chooses "submit", he is presented with a screen describing how to submit and a browser to select a file. If more than one Web page is to be submitted, the student may either submit them sequentially, assigning different filenames to each, or submit a single Zip file, which PG will unpack into its components. Entire directory hierarchies may be submitted in this manner. The requirement that students submit whole files instead of hyperlinks ensures that the reviewers will not be able to guess their authors’ identities by looking at the pathname in the URL. The ability to submit directory hierarchies allows large projects to be submitted and reviewed as easily as small ones.
The instructor assigns reviewers based upon appropriate criteria. Often reviewers are to be assigned randomly; PG can automatically generate random reviewer mappings. But sometimes it is better to constrain the assignment of reviewers. For example, if the class researches a large number of topics, with each author choosing one of them, it is a good idea to have some reviewers be students who have written on the same topic. In this case, a spreadsheet can be used to create the desired mapping, which can then be imported into PG.

Reviewers communicate with their authors via a shared Web page. There is one such page for each author (Figure 2); the author can view the reviewers’ comments and vice versa. Because there is currently only one such page, each reviewer can also see the other reviewers’ comments and assigned grades (Figure 3). In most situations, this is undesirable, so this policy is slated for change.
For PG-reviewed assignments where students in the class do different work (research different topics, for example), they sign up over the Web for their chosen topic. This allows the instructor to constrain their choices so that a relatively equal number of students choose each topic. Signups are managed by another Java program called Shimmer; Shimmer and PG share the same database format so that the same password file can be used by both.

3. How peer review has been used

PG has been successfully used to review several different kinds of assignments. One kind is research papers, as mentioned above. In an operating systems class, for example, students were asked to research how a particular operating system (e.g., Linux) solved a particular problem (e.g., mutual exclusion). They were encouraged to include hyperlinks to documents describing the strategy in more detail.

In most of my courses since 1997, students have been assigned to research one lecture on the Web. That is, they take the topics covered in the lecture and use search engines to
PG System: ReviewPage

Submitted assignment

Problem:

Submitted assignment

Comments:

Reviewer #0 Wed Dec 09 18:26:46 PST 1998
Assigned Grade: 41
A fairly good idea is conveyed, but could have been more thorough.

Reviewer #3 Sun Dec 13 18:35:49 PST 1998
Assigned Grade: 42
A well written paper but I think the topic could have been covered in more depth.

New Comment:

New grade:

Figure 3: Review page
find other treatments of the same material. The goal is to come up with five to ten good links to the material covered in class. A page of these links is compiled and submitted via PG. At the end of the semester, the best of these links (based on student evaluations) are compiled into a single page, which serves as a resource page the next time the course is taught. a

Taking this process one step further, if the lecture notes are on line (as is the case in most of my courses), students can be assigned to annotate those lectures—that is, to take the on-line notes and insert hyperlinks to explanatory documents at appropriate points. The goal is to help fellow students learn the lecture material if their background is weak or if they just want to explore the topic in more detail. Again, these annotations can be compiled at the end of the semesterb and serve as a resource for later generations of students.

As mentioned above, students have been assigned to make up a problem on some topic covered in the course. These too have been reviewed via PG, and occasionally used on homework or tests in later semesters. The "yield" is generally 20% -25%; that is, about one-fifth to one-fourth of the problems designed by students will be usable on assignments in the future. Being able to read the student evaluations and grades is a great help when looking for problems to assign.

In Fall 1998 for the first time, I had the students do three peer-reviewed assignments—research paper, annotation of a lecture, and made-up problem—but they could do them in any order they chose. In this way, students could be assigned to annotate a lecture or research a topic soon after we covered related material in class. Also, the made-up problems would be better distributed over the topics covered throughout the semester, rather than concentrated on the material covered just before the assignment was given. This seemed to work well, except that students tended to postpone the made-up problem until as late as possible—probably indicating that they considered this the hardest of the peer-reviewed tasks.

Reviewer mappings: A large number of review strategies are possible. In one situation, I assigned a group project to my students (three students were assigned to research a topic and create a study guide and discussion questions). The reviews, however, were done individually. Each student was assigned to review one project, which meant that each team got three reviews. Because review strategies are so varied, it seems better to allow instructors to load spreadsheets of reviewer mappings than to try to have the PG system generate all the mappings itself.

When students choose their topic from a list supplied by the instructor, it seems appropriate for each student to have some reviewers who have chosen the same topic, and some reviewers who have chosen others. For example, if each student reviews four other students, two of those students might have researched the same topic, and two might have written on another topic. This has the advantage that at least some of the reviewers will have a good idea of how hard it

---

a In one case, the author of the textbook used for the course made this compendium one of the resources listed on the textbook’s Website [http://www.bell-labs.com/topics/books/os-book; see "Links to other OSC pages"].

b See, for example http://www4.ncsu.edu/eos/users/e/efg/501/f98/course_locker/www/lectures/annotations.
is to come up with material on that topic, while the other reviewers will be less than experts (and therefore better judges of the clarity of the writing).

If students are assigned to review others who have chosen the same topic, there is a danger that two friends may decide to choose the same topic, and agree to give each other a high score, whether it is merited or not. To frustrate such an attempt, I refrained from assigning students to review each other if they had signed up for the same assignment at almost the same time. Of course, this does not prevent such a ruse, but just makes it less likely that it will succeed.

**The review process:** Perhaps the most frequent complaint from students was that their reviewers did not make comments in time for them to respond to them. If their reviewers waited until the last day before giving them feedback, they would have no time to improve their submission and earn a better grade. This problem was alleviated by assigning two review deadlines: By the first deadline (typically three or four days after the deadline for submitting work), reviewers were required to read the students’ submissions and post some feedback on the shared Web page. By the second deadline (typically one week after the submission deadline), final grades needed to be assigned. In a follow-up survey (detailed below), most students agreed that the two-deadline system worked well.

Peer review will work better if students have an incentive to put some effort into their reviews. To encourage this, the grading formula for student \( x \) takes into account the scores given to the students that \( x \) is reviewing—on the assumption that if \( x \) gets some credit for the work (s)he is reviewing, (s)he will be more motivated to review it carefully. About three-quarters of the student \( x \)’s grade is based on the scores that \( x \)’s reviewers give student \( x \)’s work. The other quarter is determined by the scores received by the authors \( x \) is reviewing (except for the scores given to these authors by \( x \) himself, which are not counted in determining \( x \)’s grade).

In the design phase, we were concerned about how the review system would be affected by students who dropped the course during the review period. An algorithm was designed and implemented to remap the authors and reviewers assigned to the student who dropped, while leaving the rest of the author/reviewer mappings unaffected. In practice, this turned out to be a solution in search of a problem: when students were required to sign up (on the Web) for a specific assignment, very few of them dropped between signup and submission of homework. In cases where they did, their authors were short one reviewer, and their reviewers were short one author. As long as each submission had at least three reviewers, there was no real problem.

PG’s practice of allowing an author and all his/her reviewers to communicate via a single shared Web page makes it susceptible to the "halo" effect, whereby a reviewer is unduly influenced by the comments of previous reviewers. There are advantages to letting reviewers see other reviewers’ comments and the author’s response to them; one reviewer can critically analyze another reviewer’s comments. But the disadvantages probably outweigh the advantages. PG will soon be changed to allow the instructor to deny students the ability to view other students’ comments.
4. Student reaction

To gauge student reaction to PG, the students who used it in Fall 1998 were surveyed. The students had taken an introductory graduate-level operating-systems course. Out of 89 students who finished the course for credit, 41 responded to the survey, a response rate of 46%. The survey was conducted after grades were assigned, though it was not conducted anonymously. Results are given in Table 1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Peer review is helpful to the learning process.</td>
<td></td>
<td>20</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td><strong>4.24</strong></td>
</tr>
<tr>
<td>2 I was satisfied with the reviews of my work.</td>
<td></td>
<td>4</td>
<td>29</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td><strong>3.68</strong></td>
</tr>
<tr>
<td>3 Two review deadlines were imposed, one for the first review and another for the final grade. Did this provide an adequate opportunity for you as an author to respond to the comments of your reviewers?</td>
<td></td>
<td>11</td>
<td>24</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td><strong>4.12</strong></td>
</tr>
<tr>
<td>4 Not all students do the reviews they are assigned to. Did this cause problems for you?</td>
<td></td>
<td>2</td>
<td>8</td>
<td>11</td>
<td>19</td>
<td>1</td>
<td><strong>2.78</strong></td>
</tr>
<tr>
<td>5 Should PG use HTML frames, so that you could see your author’s work at the same time you are writing a review of it?</td>
<td></td>
<td>16</td>
<td>13</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td><strong>4.00</strong></td>
</tr>
</tbody>
</table>

**Key:**

5 = Strongly agree/emphatically yes  
4 = Agree/yes  
3 = Neutral  
2 = Disagree/no  
1 = Strongly disagree/emphatically no

**Table 1:** Responses to Student Survey

Students had quite a positive reaction to peer review. When asked whether they agreed with the statement, "Peer review is helpful to the learning process," respondents’ average response was 4.24 (on a scale of 1 = "strongly disagree" to 5 = "strongly agree"). One of the most positive comments stated,

"I found the peer review system helpful in two ways. First, to make intelligent comments on others’ work I had to carefully review the material presented thus gaining a greater understanding. Second, some of the comments I received pointing out errors in presentations were helpful in clearing up misconceptions that I may have had."
Those students who did not take favorably to peer review usually cited the quality of the reviews as a reason. This response is typical:

"It’s [sic] effectiveness sometimes depends on the sincerity with which the review is done. Sometimes I felt there was a tendency that a reviewer follows the way another reviewer has done the review."

Two questions probed students’ reactions to the reviews they received. Students responded at a mean of 3.68 (1 = "strongly disagree" to 5 = "strongly agree") to the statement, "I was satisfied with the reviews of my work." Most of those who offered an explanation said that they were dissatisfied. One of them stated, "The reviewers graded my work haphazardly without giving any thought to the actual reason of the assignment." Two of the students mentioned the "halo effect," saying that only one reviewer did a real evaluation, and the others tended to follow him or her. As one student put it, "I think grades assigned should not be visible to other reviewers, only to the author. Some reviewers tend to assign grades closer to the ones already assigned. Also, a reviewer who doesn’t look at the work can just assign the average."

In response to the question on satisfaction with reviews, one of the students noted that his work was not always reviewed by all four students who were assigned to review it. Anticipating this, the survey asked a question, "Not all students do the reviews they are assigned to do. Did this cause any problems for you?" The mean response was only 2.78 on a scale of 1 (emphatically no) to 5 (emphatically yes), indicating that it was not a great problem. Perhaps more telling is the fact that 19 students said it was not a problem, while only 9 said it was. Or maybe this says that missed reviews were a problem for a significant minority of students. Whatever the interpretation, it is clear that the quality of reviews could use some improvement.

It is not clear whether more thoroughness in reviews would lead to higher or lower grades. Students who do not take the time to review work thoroughly may fail to find anything wrong with it and assign it a high. But careful reviews may induce students to do better work. In any case, the students appear to have graded more leniently than the instructor and TAs. The mean score on the peer-graded assignments was 89.0%, vs. 84.7% on the other homework assignments.

The two-deadline approach seemed to help students respond to feedback from their reviewers. One question on the survey asked,

"Two review deadlines were imposed, one for the first review, and another for the final grade. Did this provide an adequate opportunity for you as an author to respond to comments of your reviewers?"

The average score was 4.12 on a scale of 1 (emphatically no) to 5 (emphatically yes). One of the comments said that the two-deadline concept had not been adequately explained, but the bulk of the comments addressed anonymity. Most of them supported double-blind reviewing, but one student stated,
"I think that it would also be helpful to have a mechanism whereby both the author and the reviewer could relinquish anonymity in order to talk about the assignment, at least after the first review was written. There were times when I felt I could have provided the author with significant assistance but due to the limitations of the review system I was not able to do so. E-mail and the web are great but sometimes nothing beats a phone call or a face to face meeting."

5. Future PG development

The next stage of PG development will focus on reporting and versioning features. Reporting needs to be improved as follows: Grades in PG are typically based partially on peer reviews of a student’s submission, and partially on the scores received by the authors the student is reviewing. Calculating grades is currently a laborious process involving running perl scripts against the database. Grading needs to be automated, so that students will be able to view their up-to-the-minute grades over the Web.

Students are required to give feedback to the authors they are reviewing by certain deadlines. However, PG provides no way to enforce this; currently, the TA needs to check at the moment the deadline passes and record which students haven’t done their reviews. This function needs to be automated, to generate a report and penalize the reviewers who are late.

To improve the quality of feedback, PG could be expanded to encompass another level of peer review: Each student could be assigned a set of reviews to evaluate, disjoint from the reviews he had written and those that had been written on his/her work. The student would rate each review on a scale of 1 to 5, according to how helpful he thought it would prove to the author. The PG reporting system could factor these into the reviewer’s grade.

There also needs to be a way for authors to submit new versions of their work. The existing system allows authors and reviewers to communicate during the review period, by writing to a shared Web page. But authors need to make revisions to their submissions in response to comments from reviewers. Currently, the only way for an author to submit new work is by overwriting the previous version. This may be confusing to the reviewers, whose comments may now apply to a version that has been superseded.

A versioning system is needed to allow new versions of work to be submitted, but still allow viewing previous versions and previous reviews. Each review should be associated with a particular version—the version that is the most recent at the time it is submitted.

More visibility options need to be added, for example, to either allow or not allow one reviewer to see another reviewer’s comments. This would combat the "halo effect."
6. Summary

The PG project is an effort to make peer grading practical for engineering classes. Students may submit arbitrary hierarchies of Web pages, which are reviewed blindly by other students. For many kinds of non-objective homework, e.g., design problems, it can provide better feedback to students than teaching assistants have time to produce. Not only does PG provide an alternative way of grading homework, but in conjunction with a Web-based signup sheet, it also facilitates collaborative work. For example, it has been used to annotate all the lectures for a semester-long class with hyperlinks to related material from the Web. Student reaction to PG has been very positive, as demonstrated by a post-semester survey. PG has proven itself a valuable tool for enhancing the educational experience in courses as varied as Ethics in Computing and Advanced Object-Oriented Systems. Further information on PG may be found at http://uni22ws.unity.ncsu.edu/PG/basic_documentation.html.

Acknowledgments

The help of the following NCSU students with this project is gratefully acknowledged: Hunter Matthews, Neil Horman, Steve Harris, Karen Proctor, Jason Horne, Drew Puch, and Weigen Liang.

Bibliography


Edward F. Gehringer

Edward Gehringer is an associate professor in the Department of Electrical and Computer Engineering and the Department of Computer Science at North Carolina State University. He has been a frequent presenter at education-based workshops in the areas of computer architecture and object-oriented systems. His research interests include architectural support for persistence and large object systems, memory management and memory-management visualization, and garbage collection. He received a B.S. from the University of Detroit(-Mercy) in 1972, a B.A. from Wayne State University, also in 1972, and the Ph.D. from Purdue University in 1979.