Jens Palsberg and Scott J. Baxter

Teaching Reviewing to Graduate Students

Incorporating the principles and practices of formal review into a Ph.D. education, smoothly and inexpensively, as part of the existing coursework.

LISA HANEY

The computer science research community depends on qualified peer review. For example, when an editor or a program committee decides whether to accept or reject a submission, the decision is often based on a number of reviews. Moreover, an author who receives reviews of a submission may use the reviews as a basis for revision. And while graduate students are not usually asked to write reviews, they will be expected to write them after finishing their Ph.D.'s and enter the computer science community. For these reasons, it is important that new Ph.D. graduates be capable of writing useful reviews.

However, these researchers usually learn the principles and practices of reviewing with little to no practical training because such training is generally not a part of a Ph.D. education. Despite this fact, we believe teaching the review process should be part of a Ph.D. education, and that such training can be integrated smoothly and inexpensively as part of existing coursework rather than be added as an additional course.

Here, we present the design of a graduate course in which the teaching of reviewing is an integral part. We also describe our experiences teaching such a course. We believe our model could be successful in a variety of other contexts.

The Course

In late 2001, we taught the graduate course "Formal Compiling Methods" at Purdue University. The course was a seminar in which 22 papers were presented. There were two papers from 1998, two from 1999, 10 from 2000, and eight from 2001. All but four of the papers were taken from conference proceedings. Thus, they might have been revised, expanded, and submitted to a journal.

Each week, every student wrote a review of a paper, totaling 11 reviews. The students were asked to write the reviews in the style of a review of a submission to *ACM Transactions on Programming Languages and Systems* (*TOPLAS*). A relaxation stipulation of this standard was that the students were not required to check any mathematical proofs, although they were asked to make a judgment of whether the proofs were clear, convincing, and elegant. *TOPLAS* was chosen as the model for review writing because of its reputation as a top journal within the computer science community. The primary purpose of having the students write reviews was to ensure they read the papers thoroughly. A secondary purpose was to initiate students into the process of review writing.

At the beginning of the course, one of two professors (Palsberg) distributed copies of three recent reviews he had written for *TOPLAS* submissions. While it was not implied that these three reviews were the "gold standard" for reviews, they served as useful models from which the students could infer what a journal review resembled. Although it is questionable to distribute reviews written in confidence, we couldn't think of a better way of providing realistic examples. The reviews were distributed in hardcopy and the students were urged to treat them with the utmost discretion. We hope better solutions can be found in the future.

Every week, the students would send the course's second professor (Baxter) a draft of their review; he would then read them, mark them, and give the students feedback in one-on-one conferences. The primary purpose of these conferences was for Baxter to help the students say what they wanted as clearly as possible; a secondary purpose was to eliminate grammatical errors. After that, the students could use Bax-

As one of our colleagues told one of the students: "if you think you were rejected by a graduate student, that is upsetting, but if reviewer comes across as a knowledgeable expert, then rejection is more tolerable."

ter's feedback when revising their reviews before sending them to Palsberg, who would then give them comments on the technical aspects of their reviews. Because Baxter evaluated the reviews before they were sent to Palsberg, a great deal of time was saved on grammatical errors and rhetorical problems. The division of labor meant little communication between Baxter and Palsberg was necessary during the course, allowing each of them to concentrate on their area of expertise.

At the end of the semester, a mock program-committee meeting was conducted. During this meeting, the papers deemed the best were chosen, and then, during final-exam week, each of the students wrote a summary review of one of those top papers in the review style found in *ACM Computing Reviews*. This program-committee meeting and subsequent review turned out to be an effective way to push the students to try and write even stronger reviews than they had written earlier in the course.

Course Evaluation

Obviously, writing reviews cannot compare with actually writing a compiler or doing theoretical exercises on variations of the material. However, writing reviews was the course's only homework. We are happy to report the students took the opportunity to study the material in considerable depth. We are also happy to report that we saw definite improvement in the quality of reviews as the course progressed. In the following we outline four episodes we believe illustrate this improvement.

Episode 1. At the beginning of the course, students struggled with how to position themselves when writing their reviews. However, they quickly learned the importance of trying to project their credibility. As one of our colleagues told one of the students: "if you think you were rejected by a graduate student, that is upsetting, but if the reviewer comes across as a knowledgeable expert, then rejection is more tolerable." While it can be challenging for a novice to judge whether the ideas and the techniques in a paper are new, important, and difficult, the students increasingly learned to rely on what they already know.

Episode 2. Another problem early in the semester was that a number of students began to tell Palsberg they thought the paper was excellent and there was nothing in the paper to criticize. This problem was solved by sending an email message to the class pointing out there is always something a paper can be criticized for, whether it be a lack of scientific motivation, a lack of industry motivation, problems with the proofs, problems with the language, or any number of other potential problems. At this point in the course, Baxter began pushing the students to think about how the review might be useful to both a journal editor and the paper's author. Students were reminded that scientists typically read to find what is most relevant to their particular research agenda, and that a review with no criticism does not help an author or a journal editor pursue that research agenda.

Episode 3. Palsberg graded all the reviews himself, and the grades were weighted so that those reviews written in the second half of the semester were both worth more and were graded more harshly. The students were recommended to ensure their reviews be more than merely summaries, and that criticism should be phrased in terms of specific sections or paragraphs of a paper. In the second half of the semester Palsberg looked at two things he believed were not always done as well as they should have been in the first half of the semester: whether the review had both a thoughtfully written section that discussed the value of the paper as well as a useful list of suggestions for improvement; and whether there was a clear relationship between the body of the review and the overall recommendation (accept, reject, and so forth). Baxter began to focus the conferences on these issues by alerting students to specific parts in their reviews where they could highlight these points.

Episode 4. A welcome side effect of the course was the transfer of writing skills; by improving their ability to write reviews, students also improved their ability to write and reflect on their own papers. The following example helps to illustrate our point: One of the students was writing a paper for his qualification exam while simultaneuously taking our course. He wrote a first draft before the semester started and his advisor told us he suggested a substantial rewrite. This student submitted a later version of his qualifier paper and defended it while taking our course. His advisor told us he was amazed at the dramatic improvement he saw and said the writing on the final version of the qualifier paper was almost as good as what he could have written himself.

The students ultimately gained confidence in several areas, allowing them to write better and more useful reviews. They gained confidence in their ability to read and understand the technical papers presented in the course; they gained confidence in their knowledge of how reviews should be structured, written, and used; and they gained confidence in their technical writing abilities. During the course, the students gave, on average, steadily lower rankings to the reviewed papers. Confidence is certainly not a sufficient condition for improvement in review writing skills, but we believe it is a necessary one.

Conclusion

We have described a graduate seminar in which one of the purposes of the course was to improve students' reviewing skills. For those who may be considering offering a similar course at their institution, we offer the following suggested guidelines.

We spent a considerable amount of time planning the course, although once the semester started, the course took no more time to teach and prepare for than any other graduate course. There was a great amount of time and many email messages devoted to deciding when reviews should be due, how much time would be needed to read them, what the format of the reviews should be, as well as many other details. But probably the most important element in the planning was the initial phone call to Purdue's director of composition. The director of composition was able to identify an appropriate candidate and helped set up the initial meeting with the course's professors.

Because of the large number of the students' written reviews, it would have been a daunting task to teach such a course without a teaching assistant. In an ideal world, the teaching assistant for the course would have been a person from the computer science department who had expertise in teaching writing. However, at least at our institution, no such person existed. But because Baxter worked with the same students throughout the semester, he became well acquainted with the discourse of computer science. If Palsberg had not hired Baxter, and, instead, sent all of the students to Purdue's writing center for comments on their reviews, the reviews probably would not have been written as well as they were.

There is one last suggestion we would like to note. Because most of the students in this course were nonnative speakers of English, it was helpful to hire a teaching assistant who was a second-language writing specialist. However, if such a person cannot be found, a person experienced in the teaching and tutoring of writing would be second best. As we said earlier, a call to the director of composition or the director of the writing center would probably be the best place to start in finding a suitable candidate.

We were quite pleased with the results of the course, although we recognize there are other ways to structure such a course. We leave it to others to use informed judgment in the light of local circumstances to determine what things found in this "On Site" might be most useful to them as they plan and execute their own courses. A longer version of this column can be found at www.cs.purdue.ede/homes/palsberg/paper/661-eval.pdf) **C**

Jens Palsberg (palsberg@cs.purdue.edu) is Associate Head and Professor of Computer Science at Purdue University.
Scott J. Baxter (baxters@purdue.edu) is pursuing his Ph.D. in Teaching English to Speakers of Other Languages at Purdue University.