Five Years of Extra Credit in a Studio-Based Course: An Effort to Incentivize Socially Useful Behavior

Edward F. Gehringer, Zhewei Hu, Yang Song
Department of Computer Science
NC State University
Raleigh, U.S.
{efg, zhu6, ysong8}@ncsu.edu

***Abstract*—In studio-based education, students collaboratively work on projects that allow them to learn aspects of design through experience on authentic artifacts, often for outside clients. In our Object-Oriented Design and Development class, this means work on open-source projects, such as Mozilla Servo, OpenMRS, Sahana Eden, and Apache Ambari, as well as our own Expertiza project. This paper recounts five years of experience awarding extra credit for activity that students engaged in to help fellow students. These activities included doing extra peer reviews of classmates' work, helping other teams with their projects, writing various types of quiz questions, and answering questions on the Piazza message board. The incentives often led to "gaming" behavior, where students engaged in activities simply to earn points, with very little benefit to their fellow students. Consequently, the scoring policy has been changed several times during the five-year period. We show how student behavior changed as the rules changed, how we have enhanced review quality, improved responses and response time on Piazza, and transitioned students away from merely helping with project setup, and toward helping other students with the substance of their projects.**

Keywords—extra credit; studio-based education; peer assessment; Piazza (social network); plagiarism

I. Introduction

Studio-based education [1] is a way of teaching where students learn design by working on real-world projects and critiquing each other's work. Among its benefits are the fact that students have to communicate with the stakeholders (project leaders, customers) rather than just with the course staff. They cannot assume that projects will be completely specified in advance, and have to proceed inductively toward a solution, working with their teammates and clarifying requirements with the other stakeholders for the duration of the project. This is an authentic way for students to learn design, a skill in which many are deficient [2]. Another advantage is that because student teams are working on different projects that no one has ever coded before, they have no opportunity to plagiarize code from their classmates or from the Internet. This means that we don't have to police their work for collusion; rather, we can actively encourage them to help each other past rough spots—for example, how to use a particular database, how to find a library appropriate to their needs and how to invoke its methods.

In the first author's masters-level Object-Oriented Design and Development (OODD) course, students contribute to open-source projects, such as Mozilla, Sahana, and OpenMRS, as well as to the locally developed Expertiza system, a peer-assessment application. As in any course, students who don't perform well on some aspect of the coursework often ask if they can do something to earn extra credit. But it would be a shame if, in a course where the regular assignments were authentic, the extra credit was simply make-work. Five years ago, the author happened on this principle: extra credit will be awarded only for activities that will help someone *else* learn the course material. His decision was influenced by a large body of evidence indicating that students can learn effectively from each other [3]. Several such activities meet this criterion.

- Students can do extra peer reviews of other teams' projects, beyond the two such reviews that are required for each assignment.
- Students can make useful contributions on the Piazza message board/social network, for example, by asking good questions and by answering questions posed by other students.
- Students can help other teams with their programming projects, as described above.
- Students can write quiz questions that are good enough to be included in quizzes that others need to pass to demonstrate their knowledge of some part of the course material.

In later parts of this paper, we will narrate our experience with the first three of these activities, and explain how dysfunctional results led to rule changes that eventually brought about the desired behavior. Our hope is not that other instructors will copy these activities exactly, but rather that they will gain a sense of what kind of cooperation is possible among students, and what restrictions need to be imposed so that students don't fulfill the letter of the rules without making contributions in the spirit of collaboration.

II. Extra Peer Reviews

All homework in our OODD course is done in teams and peer-reviewed using the Expertiza application [4, 5]. This includes writing assignments such as design documents and

978-1-5090-1790-4/16/$31.00 ©2016 IEEE

internal documentation, as well as programming projects. In each semester during the study period, there were four peer-reviewed assignments. To receive full credit for an assignment, each student must complete two peer assessments of other teams' work. Reviews are done in two rounds. The team submits work; the reviewer writes a review; then the team has an opportunity to update their work before the reviewer reviews it a second time. Because teams consist of two to four students depending on the assignment, if each student completes two reviews, each team receives an average of four to eight reviews of their work.

Four reviews is probably not enough for the author to get a good impression of what others think of one's work. So there is some benefit to incentivizing other students to write extra reviews. Though eight reviews are usually sufficient to highlight the main issues in improving the authors' work, there is still some value in having extra reviews done, since the reviewer as well as the author learns from each peer assessment [6].

Extra credit was first awarded in Fall 2011. For each extra peer assessment, the reviewer received 0.2 % credit (this credit was added to the student's final percentage before the final grade was calculated). The reviews were controlled for quality by using meta-reviewing: for each review, a third party (another student in the class) was asked to rate the feedback given by the student reviewer to the authors. This meta-review score was factored into the student's score on that project. Thus, members of the same team might receive different scores for the same project, because part of their grade was determined by the quality of their reviewing. Students were also permitted to do extra meta-reviews; for each extra meta-review, they received a 0.1% bonus on their final average.

This policy was in use for two semesters, Fall 2011 and Fall 2012. It seemed to work reasonably well in 2011, when the mean number of reviews done was 10.84, meaning an average student did 2.84 reviews for extra credit, beyond the two required reviews on each of the four assignments. In Fall 2012, things started to get out of hand, as the mean number of reviews rose to 35.63, with over two dozen students completing more than 50 peer reviews, and one student doing more than 130. More students than ever (18 out of 128) received A+s. But the reviews were not very high in quality. More than half of the students wrote just 0-5 words of textual feedback in their average review. Students were simply filling out the dropdowns, and not justifying the points they awarded. More reviews were being completed than authors could digest; each programming project was reviewed by 17–20 other students. Clearly, reviewing had become more of a game than an educational activity.

Thus, beginning in Spring 2013, the course staff began doing the meta-reviews, instead of the students. No longer were students awarded 0.2% for each extra review; they would now get 0.2% only if the staff gave them 100% on the review. The mean review score was 71.6%. This moved the reviews in the desired direction; most of the students were now writing an average of 34.5 words. The average number of reviews declined to 8, which is exactly the number of required reviews. To some extent, this is due to the fact that the class was very small. In a class of 17 students in two- to four-member teams, there just wasn't that much work to review from other students. The same phenomenon was noticed the next spring, again with a class of 17. One possible explanation is that in a large class, students are competing with each other to try to out-review their classmates.

TABLE I. NUMBER OF REVIEWS DONE PER STUDENT, BY SEMESTER

	# students	Mean # of reviews	Median # of reviews	Std. dev.
Fall 2011	94	10.84	10	4.73
Fall 2012	128	35.63	30	26.05
Spring 2013	17	8.00	7	4.49
Fall 2013	76	19.34	18	9.12
Spring 2014	17	6.44	6	3.16
Fall 2014	79	30.48	31	14.63
Spring 2015	36	23.38	26	9.72
Fall 2015	92	20.03	21	7.57
Spring 2016	54	15.85	16	3.13

TABLE II. MEAN REVIEW SCORE, BY SEMESTER

Semester	Mean review score
Spring 2013	71.62%
Fall 2013	79.68%
Spring 2014	88.26%
Fall 2014	76.91%
Spring 2015	76%
Fall 2015	83.19%
Spring 2016	90.34%

The table starts at Spring 2013, because prior to that, the course staff did not grade reviews.

TABLE III. VOLUME OF REVIEW RESPONSES, BY SEMESTER

Semester	Mean review volume	Median review volume	Std. dev.
Fall 2011	23.25	18.62	19.06
Fall 2012	7.57	5.00	9.03
Spring 2013	34.54	28.5	19.35
Fall 2013	17.54	15.32	10.24
Spring 2014	33.53	26.54	18.16
Fall 2014	16.46	12.66	12.64
Spring 2015	18.64	16.33	6.9
Fall 2015	20.75	18.09	8.6
Spring 2016	39.24	34.86	13.08

One metric for reviews is *volume*, which is defined as the number of *different* words, not including pronouns and function words, that appear in the review. The largest volume

was obtained in Spring 2016, when we made a number of changes to encourage careful reviewing. The next largest volumes were in Spring 2013 and Spring 2014. Both of these were small classes, where students necessarily focused their time on completing a small number of reviews.

In the next two fall semesters (Fall 2013 and Fall 2014), the average number of reviews done again reached dizzying heights. In Fall 2014, enough extra credit was earned (not just from reviewing) to raise the composite class grade-point average by a half a point (from 3.31 to 3.83). Most final projects received 42 or 43 reviews! The volume of text in the average review was 17.54 and 16.46, for the two years respectively, indicating that students were again prioritizing quantity over quality.

To try to limit the number of reviews being done, we imposed a limit of no more than 8 reviews per student project, beginning in Fall 2013. But students soon found a way around the limit. A review did not count as "done" until it had been submitted. But students were allowed to *select* a project for review until the limit had been reached. Students soon figured out that they should select all possible projects to review as soon as the review period started, before anyone had had a chance to submit reviews. Later, they would be able to submit a review they were working on even if the per-project limit had been reached. To foreclose this hack, in 2015 we implemented a limit on the number of outstanding reviews a reviewer could be working on at a time. This limit was set to 3 by default. Students who wanted to select a fourth review would first have to finish one of their three outstanding reviews. This led students to flood the instructor with appeals to increase the number of reviews allowed on each project.

In Fall 2014, we switched from grading by percentages to grading by points. A perfect score was now 1000 points, instead of 100%. In Spring 2015, we imposed a limit on the number of extra-credit points that could be earned by reviewing. Each review was worth a maximum of 5 points, and a maximum of 100 points could be earned by reviewing. This did cause the number of reviews to fall; the average student did 23.38 reviews during the course, and the average final project received about 23 reviews. But the quantity of reviews was still too high, and the quality too low.

The next step, in Fall 2015, was to state that a student could earn extra credit for a maximum of 20 reviews. This had the effect of reducing the amount of extra credit, since a student could earn all 100 extra-credit points only by producing 20 "five-star" reviews. The average number of reviews done per student fell to 20 (including the 8 required reviews, so only 12 of those 20 reviews were for extra credit). But we were still disappointed in the quality of reviews; the average review contained a volume of only 20.75 words, and the average review was given a score of only 83.19%.

Our desire was for students to review fewer projects, but be more careful on each review. Accordingly, in Spring 2016, we raised the value of each review to 10 points, and limited the number of extra reviews to 2 per project on each of the four projects. This forced students to pace themselves throughout the semester, instead of trying to catch up by doing large numbers of reviews of final projects. In addition, we trained students in reviewing with an in-class "calibration" exercise, in which students rated sample projects, and then saw how the instructor rated the same project. The result was that the average review volume rose to 39.24, highest of any semester, and the average review score was 90.34%. Fig. 1 shows how review volume changed over the last three semesters. It shows that the interventions pushed the volume curve farther and farther to the right, indicating that students were writing more complete reviews.

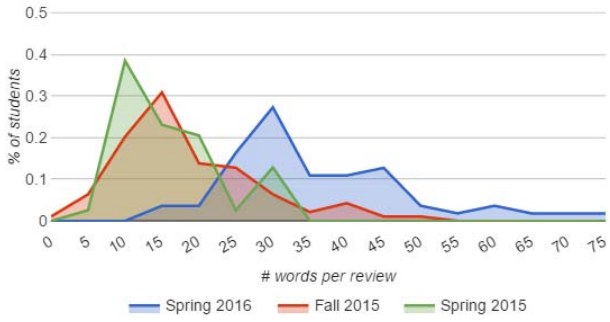

Fig. 1. Distribution of peer-review volume in last three semesters.

TABLE IV. REVIEW POLICIES OVER THE YEARS, DIVIDED INTO CATEGORIES

	Category 1 (F'11-F'12)	**Category 2 (S'13-F'14)**	**Category 3 (S'15)**	**Category 4 (F'15-S'16)**
Staff graded reviews	No	Yes	Yes	Yes
Limited extra points available from reviewing	No	No	Yes	Yes
Effectively limited reviews done by each student	No	No	No	Yes
Average # of reviews done by each student	25.13	21.82	23.38	18.48
Average volume of review response	14.21	20.06	18.64	27.59
Average review score	-	79%	76%	86%

Table IV shows how our review policy evolved over the years in an attempt to guide students to do effective formative reviews instead of a large number of reviews of low quality. The policies are divided into four categories.. Category 1 (Fall 2011-Fall 2012) comprises the semesters when students got the same extra credit for each review, regardless of its quality, and were not limited in the number of reviews they could do. Category 2 (Spring 2013-Fall 2014) are the semesters that we started to grade students' peer reviews

and let them see our feedback. Category 3 (Spring 2015) is the first semester when we limited the total extra points can be earned by doing extra reviews, but left it to the students to decide between doing large number of uninformative reviews, or a smaller number of good-quality reviews. Category 4 (Fall 2015-Spring 2016) is made up the semesters when we limited both the number of extra reviews and the extra points available. Table IV shows the features of each category and the average number of reviews, average volume of reviews, and average review scores. Fig. 2 shows how review volume was distributed for the semesters in each category.

Fig. 2. Distribution of peer-review volume in the four categories.

In summary, Category 1 shows the lowest review volume and the highest number of reviews done by each student. In Category 2, after the teaching staff started to do meta-reviews, the average peer review volume increased significantly (from 14.21 to 20.06, $p < 0.0001$) and the average number of reviews per reviewer dropped. When we limited the extra points available in Category 3, we found that many students opted to earn those points through review quantity, rather than quality. The average volume per review also dropped, but not significantly (from 20.06 to 18.64, $p = 0.60$). After we limited both the number of extra reviews, as well as the number of extra points that could be earned in Category 4, we found a significant increase in the peer-review volume (from 18.64 to 27.59, $p < 0.0001$), and the average number of reviews done by each student dropped.

III. Contributions on Piazza

Piazza is a message board cum social network on which students post questions that can be answered by other students or course staff. Instructors can endorse a student answer as being a "good answer." This saves instructors time, since they can often endorse student answers instead of answering questions themselves. Beginning in Fall 2014, our practice has been to award the top 5 to 10 question answerers (depending on class size) two points for each of their answers. For each endorsed answer, the student who answered it receives four additional points.

The results show that once we started giving extra credit for answering questions, the average responses per student increased from 1.48 to 2.03. Using a t-test, this was found to be significant at the 95% confidence level ($p = 0.0277$). Also, the average response time per question dropped from about 57 min. to 24 min., meaning that students had to wait less than half as long to get their questions answered. Interestingly, the number of posts per student also increased, but by a smaller margin, up from 4 per semester to 5 per semester. One might conjecture that students may have found it more useful to ask questions, because they could get an answer more quickly.

TABLE V. Piazza Statistics, Based upon Whether Extra Credit was Given

	No extra credit for Piazza usage					**Extra credit given**		
	Fall 2011	**Fall 2012**	**Spring 2013**	**Fall 2013**	**Spring 2014**	**Fall 2014**	**Spring 2015**	**Fall 2015**
# students	94	128	17	76	17	79	36	92
"Posts"	597	279	183	226	56	348	199	493
"Posts"/student	6.35	2.18	10.76	2.97	3.29	4.41	5.53	5.36
Avg. based on whether extra credit	4.04					5.02		
Total student responses	227	133	43	79	11	137	68	215
Responses/student	2.41	1.04	2.53	1.04	0.65	1.73	1.89	2.34
Avg. based on whether extra credit	1.48					2.03		
Avg. response time (min.)	46	78	49	46	18	17	29	28
Avg. based on whether extra credit	57.06 min.					23.98 min.		

IV. Extra Credit for Helping Another Team

We have given extra credit for helping another team each semester since 2011. The same amount of credit is given to the helper and the team helped, provided both fill out a report of the interaction, and the reports agree with regard to the amount of time spent and what was done during that time. Since we switched to the point systems, each of these interactions has been awarded 2 to 10 points of extra credit, with the average amount being 5.4 points.

Extra credit for helping another team has been awarded 114 times in the past 5 years. A closer look at these interactions (Table VI) shows that in the first class, credit was awarded more often for helping the other team set up the IDE, project, and development environment on their local machine than for helping others with the substance of their project. In all subsequent semesters, credit was awarded more often for

substantive help on the project. The third category, "helping set up remote system" refers to help deploying the project to a cloud server. When projects are peer-reviewed, the reviewers need to exercise the application to verify that it is working properly. Thus, it is a requirement to run the project in the cloud and submit the URL of the running application so that reviewers can access it.

It is encouraging that most interactions are now for help on the project itself. When teams need to consult with each other about how to set up the environment, that could be a symptom of deficient documentation. Good instructions save time. In recent semesters, we've given the students a VirtualBox environment with the development tools and the project code. This has reduced setup time and allowed the students to get started with the project more quickly.

TABLE VI. EXTRA CREDIT FOR HELPING OTHERS, NUMBER OF TIMES AWARDED

Semester	# helping set up local system	# helping set up remote system	# helping with coding the project	Total number of interactions
Fall 2011	12	2	6	20
Fall 2012	7	1	32	40
Fall 2013	1	1	8	10
Fall 2014	6	2	15	23
Spring 2015	0	0	3	3
Fall 2015	2	4	7	13
Spring 2016	0	0	2	2

TABLE VII. FINAL COURSE GRADES, BY SEMESTER

Letter grade	Fall 2007	Fall 2009	Fall 2010	Fall 2011		Fall 2012		Spring 2013		Fall 2013		Spring 2014		Fall 2014		Spring 2015		Fall 2015	
With extra credit?				with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o
A+	3	2	6	8	5	18	4	0	0	19	4	0	0	39	1	14	1	43	20
A	12	29	45	58	40	38	16	8	8	24	24	6	6	22	18	12	8	26	27
A-	15	11	22	14	24	30	27	4	4	16	20	2	2	12	23	5	6	11	23
B+	7	7	11	6	15	27	39	3	3	8	18	3	3	2	18	2	10	3	6
B	12	24	19	6	12	15	35	2	2	7	5	2	2	2	11	2	3	6	9
B-	1	1	0	1	2	0	5	0	0	2	5	4	4	2	6	1	2	3	7
C+	1		0	1	3	0		0	0	0		0	0	0	2		2		
C	0	1			1									0				1	
C-		1																	
F			2									1	1	4	4	1	1		
Total	51	76	105	94		128	126	17	17	76	76	17	17	79	79	36	36	92	92
GPA	3.54	3.51	3.62	3.84	3.63	3.71	3.4	3.69	3.69	3.81	3.62	3.23	3.23	3.83	3.31	3.84	3.35	3.95	3.75
Increment from extra credit on average					0.21		0.31		0		0.19		0		0.52		0.49		0.2

V. EFFECT ON GPA

Extra-credit work obviously raises students' grades; that's the reason they're willing to undertake it. But what is the magnitude of the effect? To determine this, the first author methodically "backed out" extra credit from all gradebook records over the five-year period. Then each student's grade was determined using the same cut points that were used for that semester (usually 97% for a A+, 93% for an A, etc., or, beginning in Fall 2014, 970 points for an A+, 930 for an A, etc.). Table VII shows the effect.

In Spring 2013 and Spring 2014, the extra credit didn't change anyone's grade. In Fall 2014 and Spring 2015, it had a major effect, approximately half a letter grade. At first glance, it appears that there has been considerable grade inflation over the period, with the most recent semester having the highest average, and the two preceding semesters ranking in the top 4 overall. However, if the averages of the last three semesters without extra credit is compared with the averages in the three "base case" semesters (before extra credit was given), we find an average GPA of 3.56 for the 2007-2009 period, and an average of 3.47 for the semesters since Fall 2014. By this reading, *all* of the "inflation" is due to extra credit.

But the rise in grades only constitutes inflation if it's not associated with learning gains. While we don't have data on learning gains, the activities engaged in for extra credit are ones that have been associated with learning gains in other situations—peer assessment [7] and student interaction in studio-based courses [8]. Past experience indicates that students can be motivated to do substantial amounts of work for modest levels of extra credit. So the overall approach is quite promising, even if verification of learning gains remains for future work.

VI. SUMMARY

This paper has examined three kinds of extra credit in a studio-based object-oriented design and development class. The following are the key findings.

- When asked to do peer assessment, many students do only a very cursory job unless they are given an incentive to do good work. Thus, all peer assessments should themselves be assessed.
- Students are very willing to review extra work if they are given extra credit for it. This makes it easy to get an adequate number of reviews for each project.
- There seems to be a need for a limit on how many reviews each student can do for extra credit, lest the experience become more of a game than a learning activity.
- Concentrating students' attention on a few instances of student work seems to be more effective than allowing them to focus more shallowly on a larger number of projects.
- Giving the top students extra credit for answering questions on Piazza increased the number of questions answered by students, and seems to have cut the response time in half.
- As we've gained experience with the OSS applications our students work on, we've been able to provide more help on setting them up, and that's allowed students to focus on helping with design and coding when they help other teams.
- Copious opportunities for extra credit causes overall grades to rise, but it is not clear whether this represents an increase in learning, or grade inflation.
- Extra credit seems to induce more interactions in a large, studio-based class than a small studio-based class, perhaps because the students have more classmates and teams with which to interact.

Pedagogies of engagement, such as Team-Based Learning [3], Project-Based Learning [9], and Process-Oriented Guided Inquiry Learning [10], have established that students can learn a lot from each other. These approaches have students work problems during class time in cooperation with each other. Our strategy is much different from these, but it too offers a way to involve students in enriching the education of their peers. The evidence suggests that they interact frequently with each other and find the interactions useful.

REFERENCES

[1] C. D. Hundhausen, N. H. Narayanan, and M. E. Crosby, "Exploring Studio-based Instructional Models for Computing Education," in Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education, New York, NY, USA, 2008, pp. 392–396.

[2] C. Loftus, L. Thomas, and C. Zander, "Can Graduating Students Design: Revisited," in Proceedings of the 42nd ACM Technical Symposium on Computer Science Education, Dallas, TX, USA, 2011, pp. 105–110.

[3] M. S. Cracolice and J. C. Deming, "Peer-led team learning," Sci. Teach., vol. 68, no. 1, pp. 20–24, Jan. 2001.

[4] E. Gehringer, L. Ehresman, S. G. Conger, and P. Wagle, "Reusable Learning Objects Through Peer Review: The Expertiza Approach," Innovate: Journal of Online Education, vol. 3, issue 5, pp. 1-7, June/July 2007.

[5] E. Gehringer, "Expertiza: Managing feedback in collaborative learning.," in Monitoring and assessment in online collaborative environments: Emergent computational technologies for e-learning support, IGI Global, 2010, pp. 75–96.

[6] K. Lundstrom and W. Baker, "To give is better than to receive: The benefits of peer review to the reviewer's own writing," J. Second Lang. Writ., vol. 18, no. 1, pp. 30–43, Mar. 2009.

[7] K. Topping, "Peer Assessment between Students in Colleges and Universities," Rev. Educ. Res., vol. 68, no. 3, pp. 249–76, Jan. 1998.

[8] R. Beichner, "The SCALE-UP Project: A Student-Centered Active Learning Environment for Undergraduate Programs," Invited paper for the National Academy of Sciences. Retrieved from http://www7.nationalacademies.org/bose/Beichner_CommissionedPaper . pdf (2008).

[9] J. Krajcik and P. Blumenfeld, Project-Based Learning. 2006.

[10] R. Moog, "Chapter 8: Process Oriented Guided Inquiry Learning," Integrating Cogn. Sci. Innov. Teach. STEM Discip., Sep. 2014.