A Successful Peer Evaluation System for Capstone Projects in Computer Science

Nary Subramanian
Department of Computer Science
The University of Texas at Tyler
nsubramanian@uttyler.edu

Abstract

In team-based capstone projects one of the biggest concerns for the instructor is how to evaluate the work of each member of the team. This is because freeloding is a big possibility especially when close friends are in the same team. At UT Tyler Department of Computer Science, we have been conducting capstone projects as service-learning projects for the past ten years – this means all capstone projects have involved developing a computer program for the industry. This also means that extensive external interaction takes place not always under the eyes of the instructor; therefore, the scope for “piggy-backing” is high. However, we have evolved a system of peer-review based on our experience with over seventy-five projects that allows each member of a team to evaluate other members. We found that this evaluation system has worked well for our projects and individual grades for each phase of the project – the requirements, design, and implementation phases - have always been determined based on this peer-review. Our peer-review system has four ratings: exemplary, satisfactory, marginal, and unsatisfactory. We found that best team players received exemplary ratings consistently over the three phases while students who did not contribute at all usually received unsatisfactories. We also asked students to evaluate themselves as well – while one may be tempted to think that every student is likely to award himself/herself an exemplary, we found that most of the time students were honest in their self-appraisal. We used this self-evaluation as a calibration point. We also gave space for comments and found useful information there – such information was usually not visible to the instructor when observing teams in class. In this presentation we discuss our peer-review system, the evaluation rubric we use, and the lessons learned from our approach.

Introduction

“Service-learning is a form of experiential education in which students engage in activities that address human and community needs together with structured opportunities for reflection designed to address desired learning outcomes.” [1] Thus there are three actors in a service-learning experience: the student, the instructor, and the community client. Students learn, in a service-learning course, not only through in-class instruction but also by performing a service to the community client, being mentored by the community client, and by completing reflection assignments that relate in-class instruction to the service experience. Service-learning benefits include exposure to
industry or practical environment, improved understanding of course materials, better educational performance, improved grades, jobs before or after graduation, external referees for jobs, and others. Service learning is important at UT Tyler where the Center for Teaching Excellence and Innovation has instituted fellowships to promote service learning in the University [2].

Service-learning in Computer Science can take one of several forms: internships, cooperative education, capstone projects, industry visits, summer research experience, and the like. At UT Tyler Department of Computer Science we have been conducting capstone projects as service learning projects which means students, under the supervision of the instructor, develop significant software systems for the community client. Examples of capstone projects include mobile app development, web information systems [3], middleware, and specialized systems based on LabVIEW, Silverlight, or Javascript. Capstone projects course is taken by senior-level students who are typically in their graduating semester in Computer Science and Computer Information Systems and the projects are completed by teams of interdisciplinary students. In the past ten years, at least seventy five projects have been completed by capstone project teams for the community, both within and without UT Tyler. Community clients outside UT Tyler include government agencies such as NASA and Tyler Independent School District, not-for-profit organizations such as Trinity Mother Frances Hospital, and corporates such as Bell Helicopter, Ingersoll Rand, Genesis, and others.

The typical size of a capstone project team is three though teams of four are also common. The team size distribution in the twenty seven projects completed in the last six semesters (Spring 2012 to Fall 2014) is given in Figure 1. Team-based projects reflect the reality of the information technology industry [4, 5] where almost all projects are composed of groups of interdisciplinary personnel. At the beginning of the semester, each team is given an overview of the project they will work on and deadlines are set for project milestones - the requirements phase, the design phase, and the implementation phase, that are usually 4 weeks, 7 weeks, and 12 weeks, away from the semester start. Once the teams are assigned to projects, teams manage themselves - they interact with their client contact as and when required and develop the required deliverables for each milestone.

One of the features of UT Tyler capstone projects is that they are interdisciplinary - each team consists of students from both Computer Science and Computer Information Systems programs. This also means that students are working with other students whom they have never known before. Teams are created by the instructor of the capstone course by discussing the skills and aptitudes of each student with other faculty members in the department - this is because the instructor may not have had the opportunity to have taught every student in the class earlier. Based on this discussion the instructor creates teams and allocates teams to projects. The interdisciplinary nature of teams makes evaluation of individual student’s contribution to the team effort more challenging.
Since projects are done for clients outside class, teams work outside normal class times to meet with their clients or to work on their projects. Since the teamwork happens out of instructor supervision, the scope for “piggy-backing” is high. This problem has been referred to by other names: passenger students [4] or social loafing [6, 7]. Therefore, the problem of assigning individual grades that are different from group grades becomes difficult.

At UT Tyler we have used peer evaluations to answer these challenges and ensure teams achieve their milestones. These evaluations also serve as the means of determining individual grades for students in team projects and this process has worked very well for the past ten years. This follows the industry practice in industry [8, 9] where peer assessment is the basis for group participation and is recognized as a necessary skill in the industry.

This paper is organized as follows: Section 2 discusses our peer evaluation process for capstone projects; Section 3 analyzes these peer evaluations over multiple semesters and for different phases and compares their quality with industry experts’ assessments; Section 4 discusses how these peer evaluations impact individual student grades; Section 5 discusses observations from using our process and compares our approach with related work; and Section 6 concludes the paper.

**Peer Evaluations for Capstone Projects**

Peer reviews are done by each student in a team on all other students in the team. Each student is given one of four ratings by his or her peers: Exemplary, Satisfactory, Marginal, and Unsatisfactory. An example peer evaluation is shown in Figure 2. Following are note-worthy in this peer evaluation:
1. It is very easy to complete - just check mark the cell that the student thinks is most appropriate for their team members
2. Student is given credit for completing the peer review - this provides additional motivation for a student
3. Student is encouraged to rate him/herself as well - this provides an opportunity for students to calibrate their rating of peers
4. Student is encouraged to give comments, if they wish, and they often provide justification for their own performance or for the evaluations of one or more of their peers at the bottom of this table.
5. This review is kept confidential and is not returned to the student - this encourages them to be open and frank about their evaluations.

In order to obtain reviews from all students, these reviews are kept as part of the mandatory project tests which are a form of reflection mechanism used to assess the knowledge gained by each student while performing project work. Therefore, all students are forced to complete the peer-reviews.

Students are also encouraged to give comments - since all teams are self-managed there is very little scope for the instructor to know of any festering issues in a team. Teams email a weekly report every Friday to the instructor highlighting any pending issues and indicate their progress toward the next milestone - however, the issues in the weekly report do not usually include personality issues; they relate to technical or communication problems with the customer. The comments provided during peer reviews allow for deeper insight into the team and the instructor can, if needed, take appropriate action. Typically the intercession by the instructor will be on a instructor-to-student basis but occasionally the whole team sits with the instructor to sort out any pending inter-personal problems. Some of the comments received from one team’s members include the following:

“I put myself there (Satisfactory) because I do feel behind. I am on a tough curve to learn and know as much as Sean does. He lives and breathes all of this, I don’t. So I always
feel I’m playing catch up on the tech side. However, I try to make up for it in things I’m good at.”

“I feel that everyone worked hard in this phase to make sure we could deliver a functioning product that would satisfy customer’s requirements.”

“I think both of my Capstone partners did enough work overall, however, in some areas I do feel like I worked harder.”

As can be seen from these comments, each member of a team had different opinions - however, they all indicate a professional outlook and a sense of respect for fellow members. It is important that they develop this skill as it is not only a goal of our program’s outcomes but is also very much required in industry [8, 9]. Moreover, these comments also help the instructor realize that one or two members are doing the deep technical work while others are doing support work such as documentation and arranging client meetings.

Analysis of Peer Reviews

Figure 3 shows the peer evaluations for six semesters from Spring 2012 to Fall 2014 involving more than 75 students. For the requirements phase and the design phase we see a somewhat mixed trend even though exemplaries are more than satisfactories for most semesters. However, for the implementation phase we find that exemplaries dominate satisfactories. Therefore, in the overall picture, exemplaries are higher than other categories.

By the numbers, there were almost the same number of evaluations for requirements, design, and implementation phases: 254, 258, and 240, respectively, for a total of 752 evaluations (they are not the same for each phase due to some students being absent for the end-of-phase project tests which include these peer-reviews). For the requirements phase, there were 147 exemplaries, 89 satisfactories, 16 marginals, and 2 unsatisfactories. For the design phase there were 161 exemplaries, 81 satisfactories, 15 marginals, and 1 unsatisfactory. For the implementation phase, there were 186 exemplaries, 46 satisfactories, 5 marginals, and 3 unsatisfactories. In total, there were 494 exemplaries, 216 satisfactories, 36 marginals, and 6 unsatisfactories.

In the same period there were 501 evaluations of these projects by industry experts, whom we call industry peers, of which there were 250 exemplaries, 239 satisfactories, and 12 marginals, as shown in Figure 4. Industry experts are called at the end of the semester to view and evaluate the capstone projects - these presentations also include a demonstration of a working prototype of the system. After each presentation the industry experts evaluate the project presented.
However, if you compare the summary evaluations we get a better picture. Since the rubric used for categories may be different between students and industry experts, we added the exemplaries and satisfactories together, and marginals and unsatisfactories together. As shown in Figure 5, we find they are similar - which gives credence to the fact that peer reviews by students can be taken as a reliable yardstick to compare performances by student teams.

It should be clarified here that the evaluation scale used by industry evaluators is the same as the scale used by students for peer reviews - that is, both use ratings of exemplaries, satisfactories, marginals, and unsatisfactories. However, we do not define a rubric for industry evaluators, that is, we do not say what should count for exemplary rating and what should be considered for a satisfactory rating and so on. In the same way, we do not provide a rubric for student peer reviews either - students judge their peers based on their individual perceptions of the contributions of their peers.

**Peer Evaluations Impact Individual Grades**

For each milestone’s deliverables a group grade is awarded - however, individual grades depend on the peer reviews and is a fraction of the group grade. As can be seen from Figure 6, individual project phase grades are determined based on peer reviews. If a student gets ‘Exemplary’ rating from all other students in his/her team, then that student’s individual grade is same as the group grade. If a student does not get all exemplary ratings from his/her peers then the student’s individual grade will be less than that of the group grade. For example, if a student gets at most one ‘Satisfactory’ rating then the maximum individual grade for that student is 97.5% of the group grade - if the group grade is 100 then this student’s grade is 98 (rounded up). If a student gets ‘Unsatisfactory’ rating from two other team members, in a team of size three, then the student can get at most 75% of the group grade. Therefore, students are encouraged to contribute effectively to their teams so that they do not lose points during grading.

These individual grades are computed at the end of each milestone - requirements, design, and implementation, and the final grade for the project is based on these individual grades. The weight distribution is 25% of requirements individual grade, 25% of design individual grade, and 45% of implementation individual grade. A 5% weight is given for the weekly emails from the team and this grade is same for all members of a team.

It should be clarified here that the grades computed using peer reviews affect only the Project Grade component of the Capstone Project course, which typically is only about 25% of the course grade. Student’s final course grade (that is, A, B, C, D, or F grades) is computed based on two midterm exams, a final exam, project tests, and final presentation with demonstration, besides the project grade.
Observations and Related Work

By making the peer review process simple, we found students are more willing to be honest in their evaluations. In fact, they tend to augment their rating with comments to justify their evaluation, even though we do not ask for these comments.

As mentioned earlier, we have not defined a rubric for these ratings: that is, we do not say what qualifies as exemplary or satisfactory and so forth - students seem to intuitively give their peers the rating they think appropriate. It is here that we find asking students to evaluate themselves helps - it serves as a calibration point. If the instructor observes that the student who gives himself/herself a satisfactory rating is a good student in class (in terms of participation, attendance, and grades) and if this student has rated a peer as exemplary then that peer should have done exceptional work for the team.

The effect of peer evaluations on individual grades as shown in Figure 6 has evolved with time. In the beginning we penalized students receiving unsatisfactories much more severely to the extent some received 0% individual grade - over time we have reduced the penalty since we found that all students contribute to a greater or lesser extent to the team. In fact, some students learn from such unsatisfactory students what they themselves should not do to better their prospects in their industry careers.

A point to note from Figure 6 is that we do not include self-classification in deciding individual grades - this is because we found most of the time students awarding themselves a satisfactory evaluations instead of exemplaries while awarding their peers an exemplary. They honestly felt they did not do as much as some of their team members and felt that they could have done better. We preferred to encourage such honest self-evaluation and not to include such evaluations in deciding individual grades of students.

A clarification on the use of Figure 6 is required: for example, during requirements phase, if a student in a team of three, receives satisfactories from both his/her team members, then that student gets at the most 95% of the group grade for the requirements phase. This same student may receive just one satisfactory (and one exemplary) during design phase, in which case the student gets 97.5% of the group grade for the design phase. For the implementation phase, the student quite possibly may receive all exemplaries in which case the student gets 100% of the group grade. Now the final project grade for the student is typically weighted 25% of requirements phase grade, 25% of design phase grade, and 45% of implementation phase grade - these weights emphasize the relative time spent on these phases during the semester. By the way, the remaining 5% of the project grade is reserved for timeliness of weekly project progress emails, as mentioned earlier.

An important observation from the peer reviews is that good team players always received exemplary ratings from their peers - in this sense ‘good’ is determined by the impact such students had on the team. For example, an exemplary student may not have
been the best programmer in the team but may have taken the lead in calling team meetings, establishing contact with the client, creating documentation, or even resolving conflicts.

Often students provided comments to justify their rating as discussed earlier. These comments helped the instructor to intervene when certain team members were not contributing as expected - usually the instructor met with such under-performing students on an individual basis and motivated them to perform better.

![Figure 3. Peer Evaluations for the Past Six Semesters for Different Phases](image-url)
Figure 4. Project Evaluations by Industry Experts

Figure 5. Comparison of Evaluations by Student Peers and Industry Peers
Since students evaluate their peers three times during a project, there is sufficient time for a student to self-correct or for the instructor to intervene in case of students who are contributing below their ability to the team effort. For example, if a student receives a marginal rating from all team members, that student receives lower grade for the milestone and he/she quickly realizes that they will have to work with the team for the rest of the semester to get better grades.

We have found this process of peer evaluation works rather well and we have used it in over seventy five projects spread over ten years. As can be seen from Figure 3, most students (almost 95%) received exemplaries or satisfactories while marginals and unsatisfactories together received only about 5% reviews. Our industry partners are also very satisfied with our students and many of our students are employed by the local industry. As long as students know the process is fair and transparent we believe that they will honestly perform peer evaluations and help complete projects for external clients without instructors needing to micro-manage teams.

**Related Work**

A very similar effort at peer evaluation has been documented in [6, 7] where team members evaluate each other and penalties are applied. In this scheme, each team member evaluates others along five dimensions: Ability to Work in a Group, Amount of Effort, Dependability, Intellectual Contribution, and Overall Contribution to Project, and along each dimension gives a rating from 1 to 5. These peer evaluations are confidential. The penalties applied to individual student is based on the difference between overall team average and the particular student's average score from peers. The decision is taken by a committee of four academics and it appears that sometimes penalties are waived depending on the average scores. An important point in this study is that a team of 4 or 5 members is self-selected and the peer assessment happens only once during the project [7]. As discussed earlier, at UT Tyler we have inter-disciplinary teams that are selected based on faculty opinions on student skills and aptitudes. Moreover, the dominant team size is 3 or 4. Also, each member evaluates other members on their overall performance: exemplary, satisfactory, marginal, or unsatisfactory - the onus of their team member's individual grade is placed on each student. This often leads to justification remarks.
that helps the instructor understand inner working of the team. Also, the penalties are applied strictly by the rubric of Figure 6 by the instructor of the capstone course without any committee discussion. Most importantly, the peer assessment happens thrice during a project and this allows students not contributing during a phase to correct their approach for subsequent phases.

The peer review process adopted by us is similar to [10]; in [10] there are three peer reviews during a project but two of them are informal in the sense that they do not impact individual grades and only the final one impacts individual grades. In our case, all three reviews impact individual grades. However, in [10] students complete a more detailed questionnaire on their team members’ contribution compared to ours.

**Summary and Conclusions**

At UT Tyler Department of Computer Science the capstone projects performed by senior students in their graduating semester has the following characteristics: done by teams, teams are interdisciplinary consisting of students from both Computer Science and Computer Information Systems programs, teams are self-managed, and projects conform to the service-learning paradigm [1] where projects are performed for an external client. Given these features, extensive student interaction for completing the project takes place outside instructor supervision. Therefore, chances for piggy-backing by some students are high and it becomes important that non-contributing students are identified early and corrective action taken. For this purpose we developed a peer review process.

Each team member evaluates other members of the team three time during a project - after requirements, design, and implementation milestones. Each student selects one of exemplary, satisfactory, marginal, and unsatisfactory for his or her fellow team members and this rating is used to calculate individual grades using the rubric of Figure 6. Very often students provide comments to justify their rating. We found that peer evaluations compare very well with a team’s assessment by industry experts at the end of a semester when projects are presented.

We believe that peer evaluation process can help team members to motivate and learn from each other. Our process has worked rather well for us for the past ten years and over seventy-five projects. We believe that others can also adopt this peer evaluation process and impart in their students an essential skill required in their industry careers.

**Acknowledgement**

I thank Prof. Derek Smith, University of Cape Town, South Africa, for kindly sharing references [6] and [7] without which, I believe, this paper would have been incomplete. Prof. Smith also provided helpful hints that helped to better understand the work done by
his group. I also wish to thank the anonymous reviewers of the original version of this paper for their valuable suggestions and comments that helped to improve the quality of this paper.

References

2. [http://www.uttyler.edu/ctei](http://www.uttyler.edu/ctei)

NARY SUBRAMANIAN

Dr. Subramanian currently serves as an Associate Professor in the Department of Computer Science at The University of Texas at Tyler, Tyler, Texas. His research interests include software engineering, systems engineering, and security engineering. He is a Faculty Fellow for Service-Learning at the Center for Teaching Excellence and Innovation at The University of Texas at Tyler.