

Team Projects And Peer Evaluations

John Kevin Doyle, Benedictine University
Ralph D. Meeker, Benedictine University

ABSTRACT

The authors assign semester- or quarter-long team-based projects in several Computer Science and Finance courses. This paper reports on our experience in designing, managing, and evaluating such projects. In particular, we discuss the effects of team size and of various peer evaluation schemes on team performance and student learning. We report statistical measures of the students' peer evaluations – do they always rate each other strongly or weakly? What are the means and ranges? More importantly, we discuss why we introduced these peer evaluations, and what effect they have had on student commitment and performance. We discuss a small number of cases where student participation was poor, and relate this to the peer evaluation process.

1. INTRODUCTION

The authors assign semester- or quarter-long team projects in several Computer Science and Finance courses. In these courses, a student team consists of a small group with complementary skills who are committed to a common purpose, performance goals, and an approach for which they hold themselves mutually accountable. Our experience in team projects where students receive a single grade for the group's effort led us to seek an alternative approach to designing, managing, and evaluating such projects. This paper discusses the effects of various peer evaluation schemes on team performance and student learning.

The Foundation Coalition [1] has identified a number of key issues in using peer evaluation. One common approach is to give the same grade to every team member on team project submissions. However, giving every individual the same grade for a team assignment is frequently viewed by students as unfair. Students express reservations about this practice when some members of the team are perceived as doing significantly more or less work than others or making more or less significant contributions to the overall success of the group. Peer evaluation provides a process by which the faculty member adjusts individual grades for team assignments by using data collected by asking each team member to evaluate the other members of his or her team.

Both computer science and business educators have addressed various aspects of peer evaluation in recent work. Kaufman, Felder, and Fuller [2] addressed some of these issues in a earlier paper, with particular focus on team projects in engineering. Clark, Davies, and Skeers [3] describe a significantly more complex computer-facilitated process, involving timesheets, self and peer evaluations based on behavioral characteristics of good teamwork, individual contribution reports, and a "quantity" report where each student reports the relative percentage of work contributed by each other student. LeJeune [4] has developed an instructor-weighted assessment mechanism that is combined with peer assessments to arrive at the final grade assignment, an approach that is both more involved and requires more instructor effort than the approach described in this paper. Gueldenzoph and May [5] conducted a literature review and identified best practices for assessment of group members. Williams [6] described experience with compulsory peer assessment methods for online collaborative work. Fellenz [7] developed a protocol for the peer evaluation of the contributions of individuals to student groupwork.

2. COURSE REQUIREMENTS

The courses under study include three undergraduate computer science and computer information systems courses, Systems Analysis and Design, Software Engineering, and Capstone Project, and two finance courses, Advanced Managerial Finance (undergraduate) and Financial Management (graduate). For all courses, team size is usually three or four, although in a few cases, teams as large as five have occurred. The undergraduate courses

include a mixture of traditional and non-traditional students, typically in approximately equal number. The graduate class enrolls working adults. Many of the students have participated in team projects in other courses, and essentially all of the graduate and adult undergraduate students participate in team projects at work. Nonetheless, the instructors discuss the basics of team management, reminding the students of the need to meet, preferably at least once face-to-face each week, certainly via conference calls or email exchanges. In the Capstone Project course, the instructor selects one team member to perform the project manager function.

2.1 Computer Science Courses

In the three computer science courses, the teams are defined in advance of the first class meeting by the instructor, balancing experience, courses taken, and academic performance to ensure that each team has the appropriate mix of talent and strength to succeed. The team project is team selected and instructor approved, and is a “software application,” requiring significant effort in definition, description, and design. Projects are developed in phases with deliverable components due on dates established at the beginning of the semester and published in the course syllabus. The major project components include:

- Project Proposal and Project Overview Statement: a presentation of the project proposal to the class and the instructor, to obtain feedback and recommendations for modifications to the proposal and the acceptability of the proposal. The Project Overview Statement is a short document, following a supplied template, that describes the project.
- Project Plan: covers the WBS (work breakdown schedule), duration and resource estimates, project network diagram, risks and risks management plan, and resource assignments.
- Requirements Specification: a document describing what the project will do in detail. All user interactions with the project, external errors and error handling, systems requirements, etc. are developed.
- System Analysis and Design Document: a document describing how the tasks are accomplished, including what each component does and how the components work together.
- Coding: the actual code implementing the design to achieve the project intent.
- Test Specification: a document describing how the project will be tested including unit testing, system testing, volume testing, integration testing, and acceptance testing. Both test strategy and test specification are required.
- Project Management: tracks, monitors, and manages the project.
- Project Demonstration: a presentation of the project prototype to the department faculty. A demonstration script is provided to the instructor prior to the project demonstration.

Software Engineering and Capstone Project courses both follow this outline. Systems Analysis and Design course replaces the final four components with a presentation of the Systems Analysis and Design document to the class and instructor.

2.2 Finance Courses

In the two finance courses, teams complete a term-long team case analysis. Each team selects a publicly traded company with long-term debt and a history of paying dividends, obtains instructor approval, and analyzes the company’s financial performance, using a series of defined tools. Weekly deliverables are required as follows:

- Company Profile: company history, the nature of the company’s operations, revenue and earnings growth rates for the last two years, and major news from the last year that would materially affect the stock’s performance.
- Statement of Cash Flows: the most recent Statement of Cash Flows and a written analysis with an estimate of FCF, EVA and MVA for the last two fiscal years.
- Pro Forma Income Statement and Pro Forma Balance Sheet: estimated sales growth rate for the next fiscal year using projected sales figure given in Value Line. Includes assumptions about all the accounts based on the company’s financial statements, i.e., comments on each line of the income statement and balance sheet stating how the team derived the line (e.g., % of sales, constant, per annual report, etc.).

- Possible Financing Plan for Next Fiscal Year: computed AFN and also AFN derived from the Pro Forma, and determination of how to fund that AFN. If the company does not need any external funds, the team proposes appropriate use of the surplus funds.
- Stock Average Total Return: the company's stock average total return and its standard deviation for the last ten calendar years, using a dividend yield for the company for these years, and using end of the year closing price (December 31 or the last trading date) from 1997-2006 to compute the capital gain. Prices and dividends are adjusted for stock splits and both the actual and adjusted prices are reported. Also, average total return and standard deviation for the S&P 500 Index is reported (assume zero dividends for the S&P 500 Index).
- Comparison to S&P 500 Index: comparison of the company stock returns and standard deviation with the S&P 500 Index's returns, and estimate of the company's beta for those 10 years.
- Cost of Debt: estimation of the after tax cost of debt using YTM, cost of common stock using constant growth dividend and CAPM models, and preferred stock (if any) using perpetuity model for the company. Also, estimation of the Weighted Average Cost of Capital for the company using Value Line capital structure weights.
- Final Report: all the weekly team assignments are compiled into a single report, addressing the feedback provided by the instructor on each weekly deliverable. Includes a summary "buy" or "sell" recommendation based on the data generated in the project.

3. PEER EVALUATIONS

Both of the peer review processes we describe allow/require the students to rate each other anonymously and privately. That is, no student knows how they are rated by another student, and students do not know how they are rated until the end of the course. While one might argue that allowing a poorly-rated student to see a poor intermediate rating immediately (rather than at the end of the course) might better motivate the student, in the courses under discussion, we chose to report the ratings to the students only at the end of the course. We have since reconsidered this decision, as discussed at the end of Section 4.2. Given the small size of the teams (normally three or four people, so two or three ratings per student), allowing the students to see each peer rating would give them a significant amount of information about how each of their peers rated them. We will discuss later our belief that the process we use does in fact motivate the students.

We have used two principal methods of peer rating. The simplest method involves asking each member of a team to rate each other member of the team at the end of the course. The average rating can then be used as a component of the rated student's course grade.

A more complex method, which we believe to be better, is to have each member of a team rate each other member of the team when each project deliverable is submitted. This gives several assessments of each student, encourages on-going commitment to the team by the students, and allows the instructor to monitor team "morale."

Infrequently, we have addressed poor peer ratings with individual students (maybe once every five or six courses). In all cases, it was clear from other information (e.g., poor homework, poor attendance, consistent tardiness) that the student in question was not committed to the course, and the peer ratings provided another factor which could be used in the discussion. We will address these cases below in the Exceptional Cases section.

We have typically prepared peer rating sheets for each team, with all team members' names, and passed the sheets out to the team members. A sample sheet is shown in Figure 1. The sheets ask the student to print and sign his or her name at the bottom, so we can look for individual biases (we've never seen any). Students do not rate themselves (or, more precisely, we do not record student's self-ratings). We've never had an instance of sabotage or graffiti – every peer rating sheet submitted has been properly completed. We individually email students who missed class and ask for their peer ratings. When the peer ratings are requested for each deliverable in a course, it is helpful to use MS Word's Mail Merge to prepare the rating sheets.

**MBA 651 Financial Management
FINA 380 Advanced Managerial Finance
Peer Rating Sheet**

Here are the members of the XXX YYY Team ZZZ.

- Please rate each person, other than yourself, on their contributions to the February 5 Possible Financing Plan document, on a scale of 1 (poor) to 10 (excellent), by putting an X on the scale in the appropriate column.
- Please print your name in the space provided, and sign in the space provided, and return to me tonight Feb. 5.

Name	1 (poor)	2	3	4	5	6	7	8	9	10 (excellent)
Student A										
Student B										
Student C										
Student D										

My name is _____

Signed _____

Figure 1: Sample student rating sheet.

4. DATA AND DATA ANALYSIS
4.1 Mechanics

Over time, we have moved to using Microsoft Word mail merge to prepare the Peer Rating Sheets. The data file used to prepare the Peer Rating Sheet shown in Figure 1 is displayed in Figure 2:

Date	FooterDate	Deliverable	TeamName	Person1	Person2	Person3	Person4
February 5	2/5/07	Possible Financing Plan	Team XXX	Student E	Student F	Student G	Student H
February 5	2/5/07	Possible Financing Plan	Team YYY	Student I	Student J	Student K	Student L
February 5	2/5/07	Possible Financing Plan	Team ZZZ	Student A	Student B	Student C	Student D

Figure 2: Data file for mail merge operation.

The mail merge template file is displayed in Figure 3:

Here are the members of the XXX YYY «TeamName».

- Please rate each person, other than yourself, on their contributions to the «Date» «Deliverable», on a scale of 1 (poor) to 10 (excellent), by putting an X on the scale in the appropriate column.
- Please print your name in the space provided, and sign in the space provided, and return to me tonight «Date».

Name	1 (poor)	2	3	4	5	6	7	8	9	10 (excellent)
«Person1»										
«Person2»										
«Person3»										
«Person4»										

My name is _____

Signed _____

Figure 3: Mail merge template file.

4.2 Results

Figure 4 shows the peer review results summary from a small Computer Science course. There were seven students, arranged in two teams (one team of three students, one of four) and seven different deliverables from each team over the semester (reported as 7 ratings/pair). There were therefore a total of 126 peer ratings. The mean peer rating in this course was 8.80/10, and the standard deviation was 2.55/10. This mean is lower than we've typically seen, and the standard deviation larger.

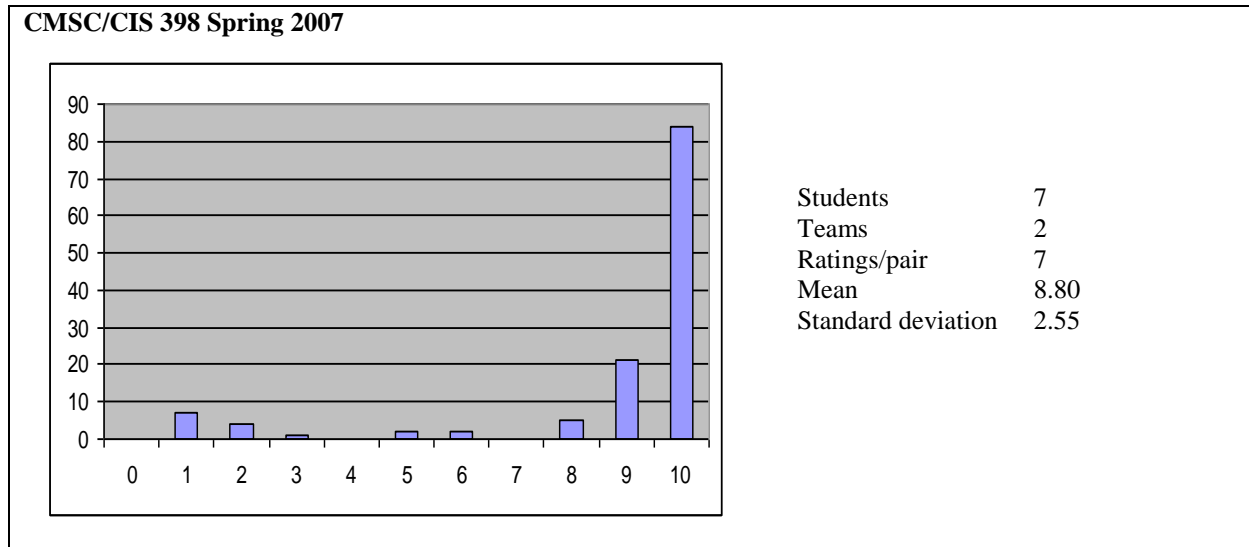


Figure 4: Peer review summary for CMSC/CIS 398 in Spring 2007

Figure 5 shows the peer review results summary for a larger MBA course, with 17 students arranged in four teams. The mean rating was 9.77/10 and the standard deviation 0.63/10.

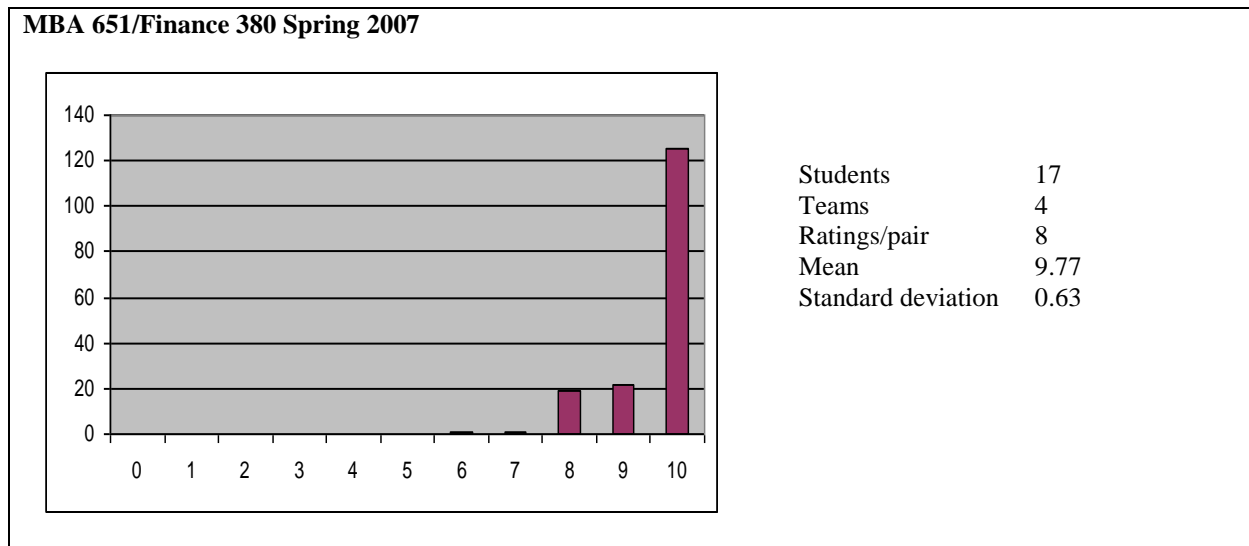


Figure 5: Peer review summary for MBA 651/Finance 380 in Spring 2007

Table 1 shows the overall distribution of peer review ratings, over all nine courses examined in this study, a total of 1,148 peer ratings. Ratings at or above 8, which we view as positive, comprised 94.2% of the total. Ratings at or below 4, which we view as negative, comprise only 2.1% of the total. The overall mean is 9.46. The overall standard deviation is not really meaningful, and so not reported.

Distribution of Scores		
Score	Raw	Percentage
0	2	0.2%
1	10	0.9%
2	7	0.6%
3	2	0.2%
4	3	0.3%
5	6	0.5%
6	17	1.5%
7	20	1.7%
8	84	7.3%
9	99	8.6%
10	898	78.2%
	1148	

Table 1: Distribution of peer review ratings for nine courses.

Of the poor peer ratings, those at or below 4, most were due to two incidents. In one case, a learning disabled student simply stopped effectively participating in the team, and neither the instructor nor teammates were able to convince him to re-engage. In the other case, a student’s marriage unraveled, in a very messy manner, in the middle of a quarter. She would periodically contact her teammates, but was never able to participate effectively thereafter. Other poor peer ratings appeared to mostly be due to individual short-term issues – sickness, business crisis, etc.

One question which naturally arises from this peer rating scheme is whether the peer ratings are consistent. That is, we worried that possibly we would really be recording whether people liked each other, or annoyed each other. Figure 6 shows the coefficient of variation μ/σ among all scores for an individual, from the other members of that individual’s team. So, this is measuring how consistent the ratings are of each individual. If we had a wide range of variation, this would tend to tell us that the ratings are suspect. But in fact, the coefficient of variation is below 0.2 for most individuals, and below 0.8 for almost everyone. We conclude that we actually are measuring peer contributions, as measured through the eyes of teammates.

Another question is whether teams with wide peer review spread (large differences between the top peer rating and the bottom peer rating) performed significantly better or worse than average. If a team was unhappy, did they perform better or worse? We found no significant correlation – the coefficient of correlation was 0.08 – pretty far away from 0.95 significance.

We also asked if low peer rating at the beginning of a course was correlated with improved peer rating later in the course. We do not have enough data to address this. Overall, as discussed above, there’s very little low peer rating, and most of that occurred in the middle or end of the course. Also, we have only recently concluded that we should share peer review ratings with the ratee (suitably sanitized – share the average rating, rather than individual ratings). Therefore, there was no direct feedback to the ratee of peer rating, so expecting performance improvement was somewhat naïve.

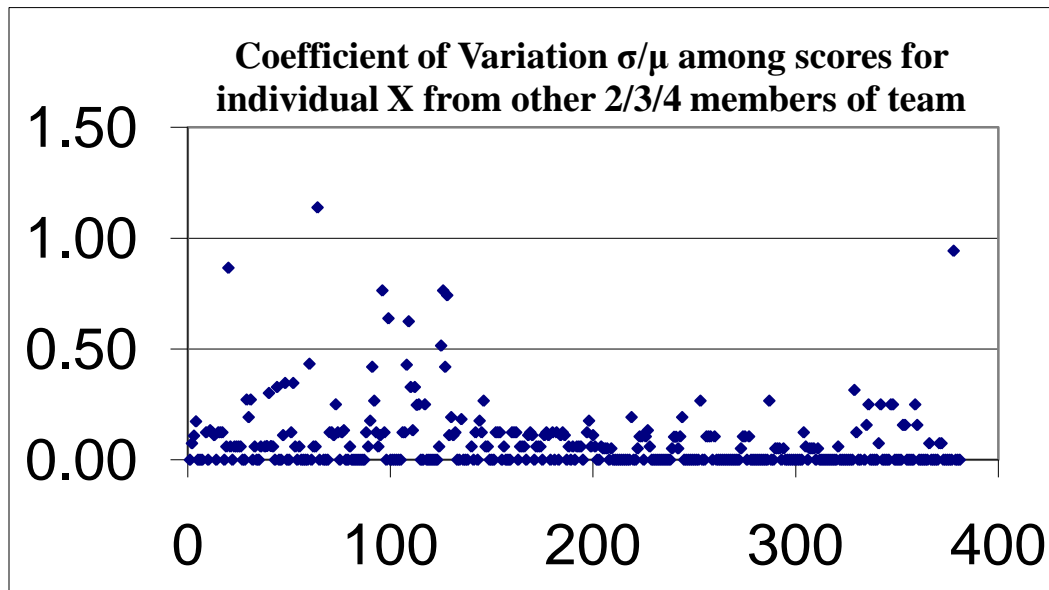


Figure 6: Coefficient of variation for 1148 peer ratings.

4.3 Future Work

We would like to explore whether high peer ratings in the beginning of the course correlate with team performance, whether peer ratings correlate with course grades (they do of course, since peer ratings are a component of course grades, but beyond that?), whether we can detect gender bias (consistent differences between female/female, female/male, male/female, and male/male rating patterns; see [8]), and whether there is any correlation between students who rate all team members the same and the course grades of the raters (suggested by [8]).

We would like to explore introducing peer rating schemes into a “writing-intensive” course. In this course, students peer review each other’s writing (to get them into the habit of getting and giving writing feedback to each other). The students are graded on whether or not they provide feedback to their teammates. We would like to enrich this to ask the author to evaluate the value of the comments they get from their teammates.

5. CONCLUSION

The authors assign course-long team projects in a number of Computer Science and Business courses. Such projects allow the students to accomplish larger and more life-like achievements than individual projects, and to learn more about team success than listening to lectures. We include a peer evaluation component in these courses, in addition to a common grade on the team project and, as appropriate, individual assignments, papers, and tests.

In this paper, we described the logistical aspects of the peer evaluations and our statistical analysis of our results. Most students rated other students very well. A large proportion of the poor ratings arose from a small number of exceptional cases. Team members rating an individual team member had significantly consistent ratings, from which we conclude the ratings were of performance, rather than personality based. We have as yet identified no correlation between peer assessment scores and team performance, although we have identified a number of areas for further investigation.

REFERENCES

1. Peer assessment and peer evaluation, http://www.foundationcoalition.org/publications/brochures/2002-peer_assessment.pdf, accessed November 6, 2007.
2. Kaufman, D.B., Felder, R.M., and Fuller, H. (2000), Accounting for individual effort in cooperative learning teams, *Journal of Engineering Education*, 89(2), 133–140.
3. Clark, N., Davies, P., Skeers, R. (2005), Self and peer assessment in software engineering projects, Proceedings of the 7th Australasian Conference on Computing Education, 42, 91-100.
4. LeJeune, N. (2006), Assessment of individuals on CS group projects, *Journal of Computing Sciences in Colleges*, 22(1), 231-237.
5. Gueldenzoph, L.E. and May, G.L. (2002), Collaborative peer evaluation: best practices for group member assessments, *Business Communication Quarterly*, 65(1), 9-20.
6. Williams, J.B. (2005), Foiling the free riders: early experience with compulsory peer assessment at an online business school, Proceedings of the 9th International Computer Assisted Assessment (CAA) Conference, 497-501.
7. Fellenz, M.R. (2006), Toward fairness in assessing student groupwork: a protocol for peer evaluation of individual contributions, *Journal of Management Education*, 30(4), 570-591.
8. Herbert, N. (2007), Quantitative peer assessment: can students be objective?, Proceedings of the Ninth Australasian Conference on Computing Education, 60, 63-71.

NOTES