Assessing Individual Performance Within a Team Using Peer Evaluations

Jerry M. Hatfield and John T. Tester
Northern Arizona University

Introduction

Working in teams is an integral part of the engineering process, and team assignments and activities have become widespread in engineering education. The challenges of assessing team performance and the contributions of individual team members increase rapidly as the number and size of the teams increase. The process can be even more difficult when the teams contain students from multiple engineering disciplines.

Many universities utilize a design project approach when introducing undergraduate students to the use of engineering science in a creative manner. [1,2,3,4] These design project courses are typically "team-oriented;" that is, two or more students are required to work together to form a single design solution to a problem. In such team-oriented design courses, there is a need for an organizational mechanism whereby the students can report the level of participation of the individual team members. The NAU Design4Practice [5,6] curriculum uses peer evaluations as a means by which the students can self-report team participation, one of several methods used in similar engineering courses. [7]

The Design4Practice program at Northern Arizona University is a 13 credit hour sequence of team-based engineering design courses beginning in the freshman year and culminating with a senior capstone design course. The sophomore course [8] divides 60 to 80 students from four different engineering majors into 9 to 12 mixed-discipline teams. Each team is responsible for the design of one of three subsystems that make up the overall design project, with groups of three teams being combined to form three to four integrated project teams, called divisions. Project team and subsystem team activities and products are evaluated and graded. Team grades are then allocated to individual students based on a combination of instructor assessments and peer evaluations.

The peer evaluation process used in the sophomore Design4Practice course was converted from a rather inefficient paper-based system to an on-line automated system in the fall of 2002, resulting in both an easier system to use and generation of a data base for analysis and evaluation of the peer review process itself. This paper addresses some challenges in the peer evaluation process that became visible through the data provided by the new automated system, discusses future plans for addressing these challenges, and provides an overview of the structure and use of the computer-based system in the attached appendix.

Copyright © 2004, American Society for Engineering Education

The Peer Evaluation Process

In team-based design courses, grades for group design activities and deliverables are usually given to the group as a whole. With large teams, it is difficult for the instructor to determine directly how much each student contributed and how to grade their participation. A peer evaluation system can be used to help determine each student's share of the group grade based on estimations of each student's contributions to the group product. [9]

In our rather large sophomore design course, groups of approximately six mixed-discipline students are formed into 12 teams, which are then divided into four divisions. All four divisions compete with each other on the design project assignments. Each student does two types of peer evaluations: 1) an evaluation of each of the other students on the same team, and 2) an evaluation of the other students within the same division and with the same major.

An evaluation consists of assignment of an evaluation value (required) and associated comments (strongly encouraged, but not mandatory) to each peer in a group The evaluation values indicate the relative contribution of each member of a group (team or major) to the tasks during the review period. If all members contributed equally, each would receive an evaluation of 1. Members that contribute more than the average contribution of the group should receive evaluations higher than 1, while less than average contributions should earn less than 1. If there are n members in a group, the total evaluation units awarded by each evaluator must equal n-1 as shown in Figure 1 (since the evaluator does not self-evaluate).

$$Sum_E = \sum_{S=1..n,exceptE} Eval_{ES} = n-1$$

Where E is the evaluator, S is the subject, and n is the number of students in the group

Figure 1: Constraint on Peer Evaluation Values

The average and standard deviation values for the peer evaluation numbers for each student are calculated as shown in Figure 2. The standard deviations are useful in flagging inconsistencies in the evaluations by a peer group.

$$Avg_S = \frac{1}{n-1} \sum_{E=1..n,except:S} (Eval_{ES})$$

$$\sigma_S = \sqrt{\frac{1}{n-1}\left[\sum_{E=1..n,except:S} (Eval_{ES})^2\right] - (Avg_S)^2}$$

Where E is the evaluator, S is the subject, and n is the number of students in the group

Figure 2: Calculation of Peer Evaluation Statistics

The evaluation process produces four sets of evaluations:

1. Evaluation of an individual by members of his/her team.
2. Evaluation of an individual by other students in his/her division with the same major.
3. Evaluation of an individual by the instructor.

Copyright © 2004, American Society for Engineering Education

4. Evaluation of an individual's team by the instructor.

The four evaluation results are combined as shown in Figure 3 and then used to determine what share of team and division product scores are allocated to each student.

$$\begin{pmatrix}\textbf{Individual's}\\ \textbf{Net}\\ \textbf{Evaluation}\end{pmatrix} = \left(\begin{pmatrix}\text{Team's}\\ \text{Evaluation}\\ \text{of Individual}\end{pmatrix} \text{X} \begin{pmatrix}\text{Weight of}\\ \text{Evaluation}\\ \text{by Team}\end{pmatrix} + \begin{pmatrix}\text{Major's}\\ \text{Evaluation}\\ \text{of Individual}\end{pmatrix} \text{X} \begin{pmatrix}\text{Weight of}\\ \text{Evaluation}\\ \text{by Major}\end{pmatrix} + \begin{pmatrix}\text{Instructor's}\\ \text{Evaluation}\\ \text{of Individual}\end{pmatrix} \text{X} \begin{pmatrix}\text{Weight of}\\ \text{Evaluation}\\ \text{by Instructor}\end{pmatrix}\right) \text{X} \begin{pmatrix}\text{Instructor's}\\ \text{Evaluation}\\ \text{of Team}\end{pmatrix}$$

$$\begin{pmatrix}\textbf{Individual's Score}\\ \textbf{for Project}\\ \textbf{Assignment}\end{pmatrix} = \begin{pmatrix}\text{Given Score}\\ \text{on Project}\\ \text{Assignment}\end{pmatrix} \text{X} \begin{pmatrix}\text{Individual's}\\ \text{Net}\\ \text{Evaluation}\end{pmatrix}$$

Figure 3: Calculation of Individual Peer Evaluation and Resulting Project Assignment Score

Descriptions and examples of how the on-line evaluation system is used are provided in the appendix.

Sources of Distortion of Evaluation Values

Ideally, the peer evaluation process should be objective, accurate, and fair. However, there are a number of potential sources of distortion that the instructor must be sensitive to and make adjustments for. These became painfully clear upon review of the evaluation results provided by the new on-line system.

1. Reallocation of evaluation points: The evaluation procedure allocates each team member n-1 evaluation points to distribute to the other members of the team, where n equals the number of students on the team. If there is a weak member on the team, that person receives very few points, and since all n-1 points must be awarded by each evaluator, the other students on the team receive the excess points. For example, consider an extreme case where one member is not performing at all and the other team members are performing equally; the resulting averages for a team of six would be 0, 1.2, 1.2, 1.2, 1.2, and 1.2. Consider another team of five students that are all performing equally; their average evaluation values would all be 1.0. This would imply that the five performing members of the first team were each performing 20% better than the five members of the second team, which may not be true. This distortion can be compensated by adjusting the instructor's team evaluation (0.83 for the first team in this example and 1.0 for the second team.
2. Students not doing an evaluation: When some students on a team do not participate in an evaluation, the resulting evaluations are based on fewer inputs and are more sensitive to impact by the other distortion sources.
3. Difficulty in assessing and quantifying individual contributions: This sometimes results in either no differentiation between team members' performance or in excessively large variations.
4. Not knowing some students: This can be a problem if the teams or divisions are too large.

Copyright © 2004, American Society for Engineering Education

5. Not aware of an individual's contributions: This occurs because not every student works closely with every other student in the team or major group.
6. Not wanting to hurt someone: Some students have difficulty in giving low evaluations, even in extreme cases. It is not unusual to see comments such as "missed a lot of classes, did not seem interested in the project, did practically no work" coupled with an evaluation of 0.9 (1.0 would indicate an average relative contribution).
7. Bias: Gender or ethnic differences can influence the evaluations; these are sometimes difficult to detect.
8. Personality conflicts: These are likely to occur during periods of intense project pressure and can significantly influence the evaluation results. Instructor observations are important in detecting and compensating for this problem.
9. Insufficient involvement in the project: In team design activities, some students tend to hide in the shadows and contribute little to the team. These students typically give everyone a rating of 1.0 and provide comments such as "everyone is doing a great job" or "everyone is contributing equally".
10. Collusion: Two students may agree to give each other high evaluations at the expense of the rest of the team. This can be detected if the instructor's observations are not consistent with the peer evaluations

Improvements Needed

Peer evaluation systems must be designed with the objective of minimizing the effects of various types of biases and process deficiencies. [10,11] Our new on-line evaluation system provides much more data than the old paper-based system and with considerably greater clarity, but in doing so it unmasked many previously undetected problems. The structure of the system implementation works well, but the next task is to refine the data being collected and improve its processing to obtain more objective evaluation results. The Analytic Hierarchy Process, using pair-wise comparisons, is one method to be considered. [12]

References:

[1] Courter, S. S.; Millar, S. B.; Syons, L.; From the students' point of view: Experiences in a freshman engineering design course, *Journal of Engineering Education*, v87, n 3, p 283 – 287, Jul 1998.

[2] Parcover, J. A., McCuen, R. H., Discovery Approach to Teaching Engineering Design, *Journal of Professional Issues in Engineering Education and Practice*, p 236-241, Oct 1995.

[3] Richardson, Jim; Dantzler, John, Effect of a freshman engineering program on retention and academic performance, *32nd Annual Frontiers in Education*, v 3, p S2C/16-S2C/22, Nov 6-9 2002, Boston, MA, United States.

[4] Piket-May, M.; Avery, J.; Service learning first year design retention results, *31st Annual Frontiers in Education*, v 2, p F3C/19-F3C/22, 2001, Reno, NV, United States.

[5] Hatfield, J.; Collier, K.; Howell, S.; Larson, D.; Thomas, G.; Corporate Structure in the Classroom, A Model for Teaching Engineering Design, *1995 Frontiers in Education Conference Proceedings*, November 1-4, 1995, Atlanta, GA.

Copyright © 2004, American Society for Engineering Education

[6] Hartman, D.; Larson, D; Design4Practice, Invited Paper, *Proceedings, The 28th Annual Israel Conference on Mechanical Engineering*, Ben-Gurion University of the Vegev Beer-Sheva, Israel, June 14-15, pp. 419-423.

[7] Campbell, S.; Colbeck, C. L.; Teaching and assessing engineering design: A review of the Research, *Proceedings, American Society for Engineering Education (ASEE) Annual Conference and Exposition*, v 1, Session 3530, p 7PP, 1998.

[8] Larson, D.; Tester, J.; Wecker, S.; Brauer, E.; The Practice and Process of Engineering Design Through Robotics, EGR 286 Engineering Design II – The Process, *AAAI Spring Symposium on Robotics and Education*, March 2001, Palo Alto, CA.

[9] Kaufmann, D.; Felder, R.; Fuller, H.; Peer Ratings in Cooperative Learning Teams, *Proceedings, American Society for Engineering Education (ASEE) Annual Conference*, Session 1430, 1999.

[10] Ohland, M.; Layton, R.; Comparing the Reliability of Two Peer Evaluation Instruments, *Proceedings, American Society for Engineering Education (ASEE) Annual Conference*, Session 3530, 2000.

[11] Layton, R.; Ohland, R.; Peer Ratings Revisited: Focus on Teamwork, Not Ability, *Proceedings, American Society for Engineering Education (ASEE) Annual Conference*, Session 2230, 2001.

[12] Friar, Les; Student Peer Evaluations Using the analytic Hierarchy Process Method, *1995 Frontiers in Education Conference Proceedings*, November 1-4, 1995, Atlanta, GA.

JERRY M. HATFIELD
Jerry Hatfield is an Associate Professor of Electrical Engineering at Northern Arizona University. He is a graduate of the University of California (BSEE) and of the University of Southern California (MBA) and is a registered professional engineer. His areas of interest include freshman programs, multi-disciplinary design, computer aided instruction and testing, computer aided instrumentation systems, and analog and digital circuit design.

JOHN T. TESTER
John Tester is an Associate Professor of Mechanical Engineering at Northern Arizona University. He received his doctorate in Industrial & Systems Engineering at Virginia Tech. Prior to his academic career, he was a Mechanical Engineer at the telecommunications firm, Amtech Systems; he also served as an engineering officer in the U.S. Air Force. His interests include design, advanced manufacturing, renewable energy and engineering education.

Copyright © 2004, American Society for Engineering Education

APPENDIX

The On-Line Peer Evaluation System

The core of the peer evaluation system is a set of CGI programs written in the Perl language and running under the Unix operating system on a web server. The CGI programs provide two-way web-page communications between the system users (students and faculty) and the evaluation processing and database. The system is secure and limits access to validated users only. This system was designed for use in a specific team-based design course, but it has the flexibility to be used for peer evaluations in a wide range of courses. Data is easily transferred between the peer evaluation system and Excel grading spreadsheets.

Setting Up the System

1. The course instructor creates an Excel spreadsheet that contains the following information for each student:
    a. Name
    b. Major
    c. Name of the Team the student is on
    d. Name of the Division the Team is within
    e. The student's university computer system log-in identification.
2. The instructor logs into the evaluation system and pastes the team information from the Excel spreadsheet into a Setup web page as shown in Figure A-1.

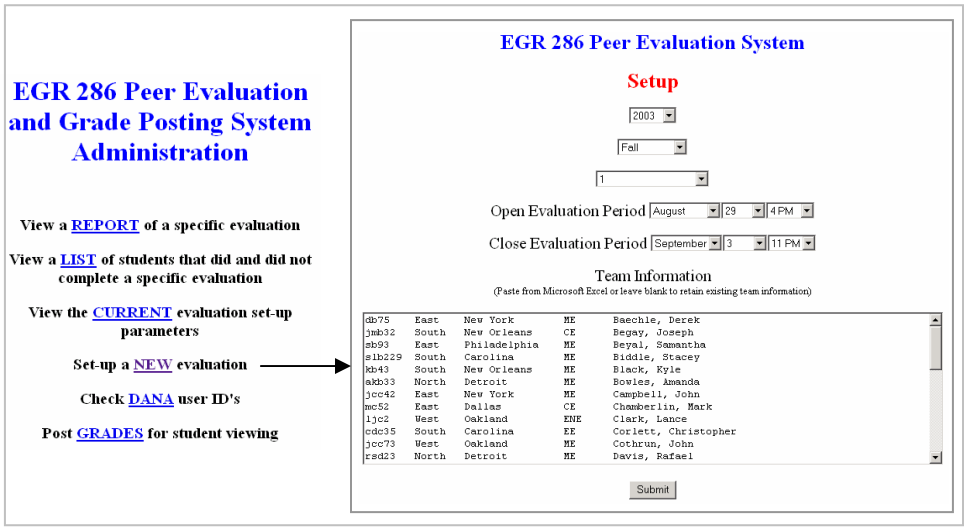

Figure A-1: Setting up an Evaluation

3. The evaluation system stores the team information in a secure file for future use.

Scheduling an Evaluation

1. The course instructor logs into the evaluation system, goes to the Setup web page, and schedules an evaluation by specifying:
    a. An evaluation number

Copyright © 2004, American Society for Engineering Education

b. Opening date and time
c. Closing date and time

2. If there are any changes in the team assignment information (such as a student dropping from the class or changing teams), updated information may be pasted from the Excel team information spreadsheet into the Setup web page.

Performing an Evaluation

1. A student logs into the evaluation system using his/her ID and password; only valid users are allowed access.
2. The evaluation system compares the log-in ID against the team information file and prepares lists of the other students on the evaluator's Team and of the other students with the same major that are members of the same Division. This information is presented to the evaluator through an interactive Evaluation web page.
3. The evaluator assigns an evaluation value and enters comments for each student on the Same Team List and each student on the Same Major List. The evaluation values indicate the relative contribution of each member of a group (Team or Major) to the tasks during the review period. If all members contributed equally, each would receive an evaluation of 1. Members that contribute more than the average of the group should receive and evaluation higher than 1, while others receive less than 1. The average of each group's evaluation values must equal 1; the interactive web page computes the average and will not allow the evaluation to be submitted until it is correct.
4. The evaluation system processes the evaluation, records the results in a secure file, and records that this student has completed the peer review process; once submitted, an evaluator's inputs cannot be changed.

Reviewing the Evaluation Results

1. An instructor logs into the evaluation system through an Administration web page and, upon selection of the report option, is routed to a Report Request web page. Upon selection of the year, semester, and evaluation number, the system then opens the respective evaluation data file and asks for selection of a specific Division for which a report is desired.
2. The system responds with an interactive web page, as shown in Figure A-2, presenting a summary

Peer Summary for South Division
Fall 2003 Evaluation No. 4

Click on a name for details
This table may be selected and pasted into MS Excel for further processing.

Team	Name	Team Rating	Team Deviation	Major Rating	Major Deviation
New Orleans	Begay, Joseph	1.08	0.09	1	0
Carolina	Biddle, Stacey	1.13	0.12	0.89	0.1
New Orleans	Black, Kyle	1.05	0.09	1.19	0.38
Carolina	Corlett, Christopher	1.23	0.21	1	0
Tampa Bay	Gerber, Paul	1	0	1	0
Tampa Bay	Huisman, Richard	1	0	1.26	0.51
Carolina	King, Jeremy	1	0.24	1	0
Tampa Bay	Marquez, Kristen	1	0	1.03	0.05
Carolina	Pfau, Joseph	0.65	0.36	0.7	0.34
Tampa Bay	Robota, Andrew	1	0	1.01	0.09
New Orleans	Ruffenach IV, Peter	1.08	0.13	1.21	0.32
Carolina	Saunders, Brian	0.99	0.19	0.84	0.24
New Orleans	Slakey, Matthew	0.67	0.38	0.9	0.24
New Orleans	Wangmo, Dechen	1.13	0.22	1	0

Select another Division

Figure A-2: Peer Summary Report

Copyright © 2004, American Society for Engineering Education

of the evaluations for all students in the selected Division. The average and standard deviation of the evaluation values are shown for each student from two evaluator groups:

a. Each student is evaluated by the other students on his/her Team.
b. Each student is evaluated by the other students in the Division that have the same major.

3. Two forms of detailed reports are available, as shown in Figures A-3 and A-4:
   a. A report showing details of the evaluations OF the selected student by his/her peers lists the name of each evaluator plus the evaluation value and comments provided by that evaluator.

Evaluations for Ruffenach IV, Peter , South Division, Team New Orleans, ME

Evaluation by other team members

Evaluator Name	Evaluation	Comments
Slakey, Matthew	1	
Wangmo, Dechen	1.3	Helped a lot especially with implementation. Spend weekends and long hours working with a few of us on it.
Begay, Joseph	1	greater worker and communicator
Black, Kyle	1	Peter also did a great job in building the robot. He also was one of the few who put in a great number of hours. He also helped out in putting the final presentation together.

Average evaluation by team = 1.08

Evaluation by others in the same major

Evaluator Name	Evaluation	Comments
Huisman, Richard	2	Put in a ton of hours implementing and rebuilding every subsystem Deserves extra credit
Pfau, Joseph	1.1	Did work on hours and final presentation
Marquez, Kristen	1.1	Worked long hours, always willing to help
Robota, Andrew	1.1	
Slakey, Matthew	1.1	showed a lot of effort
Black, Kyle	1.1	Peter was probably the only ME who spent the most hours in constructing the robot. He too is a great presenter.
Biddle, Stacey	1	

Average evaluation by others in the same major = 1.21

Return to Summary

Figure A-3: Detailed Evaluations OF a Student

Copyright © 2004, American Society for Engineering Education

b. A report showing details for the evaluations BY the selected student of his/her peers lists the name of each peer plus the evaluation value and comments related to that peer.

Evaluations BY Wangmo, Dechen , South Division, Team New Orleans, EE

Evaluation of other team members

Name	Evaluation	Comments
Begay, Joseph	1.1	Worked till late at night helping and doing his share of the work.
Black, Kyle	1.0	performed an expected amount of work and was willing to work if need be.
Ruffenach IV, Peter	1.3	Helped a lot especially with implementation. Spend weekends and long hours working with a few of us on it.
Stakey, Matthew	0.6	Not very co-operative with team members. After built debri-manipulator was not working, he refused to rebuild it again. Was hardly around during the whole implementation phase when we needed people to work.

Evaluation of others in the same major

Name	Evaluation	Comments
Corlett, Christopher	1	Worked extremously for long hours at night and in the morning almost everyday. If the robot was a success, it owes it to Chris. I would like to give him more points but the computer doesnt allow me to. Chris has put in long hours in doing the MECHANICAL as well as electrical part.

Return to Summary

Figure A-4: Detailed Evaluations BY a Student

4. Navigation paths between the report web pages are shown in Figure A-5.

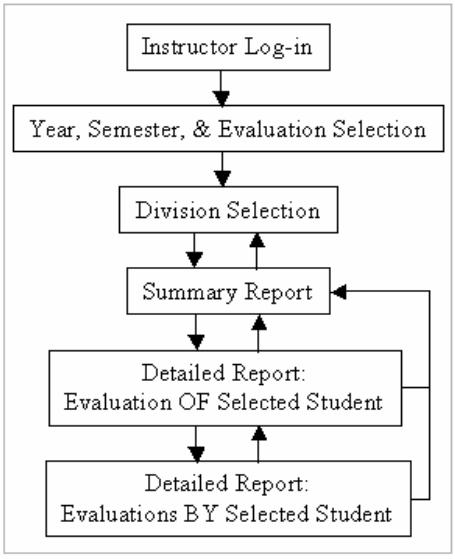

Figure A-5: Report Web Page Navigation Paths

Copyright © 2004, American Society for Engineering Education