

A Repeated Measures Design for Assessment of Critical Team Skills in Multidisciplinary Teams

Robert S. Thompson
Colorado School of Mines

Introduction

Teamwork education has become increasingly important over the last decade. In 1996, the Accreditation Board for Engineering and Technology (ABET), the sole agency responsible for the accreditation of engineering programs, approved new standards for accreditation reviews. The new standards, Engineering Criteria 2000, require programs to demonstrate specific skills. One specific criterion is the need to demonstrate that graduates have “an ability to function on multidisciplinary teams. 1”

The Colorado School of Mines (CSM) has an established tradition of teamwork education ^{2,3}. Teamwork and critical team skills that are a part of CSM’s Multidisciplinary Petroleum Design course are introduced in the following paragraphs. In this course, students from the disciplines of geology and geological engineering, geophysical engineering, and petroleum engineering work on ill-structured open-ended problems from the petroleum industry. The goal of multidisciplinary team education is to improve team performance. The focus of this paper is on performance feedback of the behaviors that characterize successful teams. Feedback typically comes from faculty but in the case of the teamwork described in this paper, this perspective may be inadequate since faculty may observe only a small fraction of the interactions. Peers working on the teams are afforded a unique view of each other’s behavior and are in a position to provide feedback for improving team performance.

The emphasis on teamwork skills stems from the widespread use of teams in industry. The teams described in this paper are similar to self-directed work teams (SDWT’s). SDWT’s (also referred to as task forces) are characterized by:

1. A limited life,
2. Are usually heterogeneous because of the diverse needs of the project,
3. Have a limited time frame to solve a specific problem,
4. Have members that may not know each other and their capabilities,
5. Must perform non-routine work, and
6. Have a mix of autonomy (self-directed) and dependence (client) ⁴.

These characteristics offer challenges for SDWT's. Seminal research by Morton Deutsch ^{5,6} and later supported by others ⁷⁻¹⁰ document the need for a common goal in group work and the importance of goal strength ¹¹. SDWT’s often have no clear path to a goal. This fact introduces

uncertainty in the task, in how team members will work on it, and in how much time it will consume. SDWT's are often multidisciplinary, and thus possess an element of heterogeneity. The more that uncertainty and heterogeneity pervade a project, the more a task force must be prepared to learn, invent, change, and negotiate. Although uncertainty and heterogeneity offer higher potential creativity, they also bring higher risk. Gersick and Davis-Sacks⁴, p.153 suggest that the "skills of team members-and those who establish and manage the teams-in dealing with uncertainty and heterogeneity strongly influence the ultimate effectiveness of task forces."

Hackman and Morris¹² identified three enabling conditions for effective teamwork: team member effort, team member knowledge and skills, and team task performance strategy. These three variables are identified as mediating variables in the teamwork model presented as Figure 1. The mediating (causal) variables are used to explain the relationship between the initial system components (the SDWT and environment) and the final desired state of effective multidisciplinary problem-solving teams. Team skills such as cooperation, feedback, team leadership, communication, coordination, and back-up behavior are identified as moderating variables in Figure 1 and are assumed to "moderate" the process. That is, these variables have the potential to strengthen (or weaken) the desired outcome¹³. If these skills are missing or poorly developed, the effectiveness of the team is diminished. The team skills of cooperation, feedback, team leadership, communication, coordination, and back-up behavior are the focus of this research.

The foundation for many of the listed team skills is military team research. Morgan, Glickman, Woodard, Blaiwes, and Salas¹⁴ used instructors in the Navy to observe and identify certain behaviors. Effective teams were observed to have a greater proportion of positive skills (compared to negative skills) than was observed for ineffective teams. Effectiveness was rated by the Navy team instructors. Oser, McCallum, Salas, and Morgan as cited in¹⁵, p. 347 observed similar patterns for different Navy teams. In both cases, effective team performance was rated by experts. Airline cockpit crews have also been a focus for teamwork skills research. The common thread in each of the team skill studies is interdependency. On this basis, the results have implications for teams where "interdependence and coordination are required"¹⁵, p. 342. Multidisciplinary teamwork in the petroleum industry is an example.

Using this rationale, six core critical team skills were identified for this research; feedback, back-up behavior, coordination, cooperation, leadership, and team orientation. These critical team skills and the constructs, defined in Table 1, are similar to those identified by Cannon-Bowers et al.¹⁵. Measures of these variables were developed by Morgan et al.¹⁴ for Navy teams that depended on "the integrated performances of teams of individuals who must coordinate their activities in order to contribute to group decision making, unit performance, and operational effectiveness." Anhalt¹⁶ modified the instrument for use as a peer evaluation tool. The instrument used in this research is similar to Anhalt's with only slight modifications being made. The team skills instrument is used to evaluate team skills over time from the perspective of the team members.

A repeated measures design is used to investigate changes in critical team skills over the period of one-semester measured from the perspective of peers. The peer rating system uses a

behavioral anchored rating scale. The following question will be tested: Is there an improvement in critical team skills over the period of a semester for student teams from the perspective of the individual team members? It is hypothesized that critical team skills will improve over the course of a semester.

Method

Participants

Seniors from the disciplines of geology and geological engineering (GE), geophysical engineering (GP), and petroleum engineering (PE) at the Colorado School of Mines were selected for the study. The participants are students in a senior capstone design class that is required for all undergraduate students in the Petroleum Engineering Department. The course is an elective for undergraduate students from the Geology and Geological Engineering and Geophysics Departments. A course objective is development of team skills and critical problem-solving skills in multidisciplinary teams. There were 41 students in the class (33 males and 8 females). Participation in the research was voluntary and students were given the appropriate consent forms before the research began. One male student did not want to be included in the study. Two cases were eliminated because of the negative influence of outliers on the statistical analysis. The distribution by discipline and gender for the 38 students included in the analysis is summarized in Table 2.

Procedure

The research objective of developing a conceptual framework for studying multidisciplinary teams was presented to members of the class at the beginning of the semester. The participants were lead to believe that the scope of the research was broader than an investigation of peer and self-evaluation of critical team skills. The multidisciplinary teams are typically composed of one geologist, one geophysicist, and three petroleum engineers. Team assignments are random with the only constraint being that each team should have at least one geologist and one geophysicist. The number of geology and geophysics students is usually the limiting factor in the team composition and it is the exception to have more than one geologist and geophysicist on a particular team.

There were three major projects during the semester. These projects are preceded by an instruction phase that focuses on the task performance strategy shown in Figure 2, ³. The instruction phase also includes training in meeting management and facilitation, brainstorming, dialogue versus debate, group decision-making, and conflict resolution. The first team project emphasizes the methodology (Figure 2) for multidisciplinary teamwork. The emphasis on the strategy is scaled down during the second problem. Finally, during the last problem, no mention of the strategy is made in the problem statement.

Peer and self-evaluation data were collected at the end of each of the three major projects. Team membership changes for each of the projects. Each member of a team assesses their team members on the six identified critical team skills using a behavioral anchored rating scale (1-5 Likert). The results organized at the team level are presented to each individual team member.

An example is shown as Figure 3. As shown on Figure 3, individuals are assigned a code and the source of individual evaluations is kept confidential. Students are free to exchange ratee codes, but this is their choice. Students are told that peer evaluations are used by the faculty team to make border-line grade decisions at the end of the semester. Thus, there was no explicit grading scheme to integrate the peer evaluations into the final course grade. From the students perspective, the peer evaluation of the critical team skills had a dual purpose, feedback and evaluative.

Design

The final design was a 3-level one-factor within-subjects repeated measures design. Table 3 is a representation of the design. The repeated variable, Occasion, represents the peer evaluations made at the end of each of the three major projects. The dependent variable is the average value for the peer evaluations (does not include self-evaluations) of the team skills measured at the end of each project. A 3-level one-factor within-subjects with a 2-level one-factor (discipline) between subjects repeated measures design was also tested. In this case the disciplines of geological engineering and geophysical engineering were combined into one group. A more complex design utilizing multiple dependent variables, 3-levels for discipline, and 2-levels for gender was eliminated for the reasons discussed in the following paragraph.

MANOVA test procedures require three assumptions. The observations must be independent, the variance-covariance matrices must be equal for all treatment groups, and any linear combination of dependent variables must follow a normal distribution. Other considerations include linearity and multi-collinearity among dependent variables and the sensitivity of outliers. The requirement of equal variance-covariance matrices was sufficient justification to support the decision to combine the multiple dependent variables (six measures of team skills) into one composite dependent variable. A related concern was the high degree of multi-collinearity among the dependent variables. The decision to eliminate the between-subjects factors of gender and to combine the disciplines of geological engineering and geophysical engineering was also made to minimize the risks from violations of the equal variance-covariance matrix requirement. Hair, Anderson, Tatham, and Black¹⁷ suggest that if the ratio of the number in the large group to the number in the small group is less than 1.5, the risks from violations of the equal variance-covariance matrices requirement is minimized. This can be accomplished in this study by combining the disciplines of geology and geophysics. Finally, two cases were eliminated based on the negative impact of outliers on MANOVA procedures¹⁷. Box plots for the Average Teams Skills, shown in Figure 4, were used to identify the two outliers. This figure also shows the trend over time in the peer evaluations.

Results

The first design investigated is the 3-level one-factor (Occasion) within subjects with a 2-level one-factor (discipline) between subjects. Geology and Geophysics students were combined for one group with the other group being petroleum engineering students. This design is represented in Table 4. The results support the statement that there is no significant discipline by team skill interaction ($p = 0.83$) and that there is a significant main effect for team skill using the multivariate tests for significance ($p = 0.004$). The test of significant using the adjusted F values

and the Greenhouse-Geisser Epsilon to correct the degrees of freedom for the main effect, team skill, was also significant ($p = 0.028$). The interaction between discipline and team skill was also insignificant using the adjusted F values ($p = 0.45$). These conclusions are consistent with the simplified model that is described in the following paragraph.

The design was simplified by eliminating the between subjects factor. This is justified since the main research objective is to investigate changes in critical team skills over the period of a semester. Consistent with the previous design, the main effect, team skill, is significant across occasion. This statement is based on multivariate tests of significance ($p = 0.006$) and adjusted F values using the Greenhouse-Geisser Epsilon to correct the degrees of freedom ($p = .037$). The mean values for team skill are 4.24, 4.28, and 4.48 for the first, second, and third projects respectively. One-way ANOVA multiple comparison procedures were used to determine which pairs are significant. Using the Newman-Keuls technique, there is a significant difference between the first and third projects and the second and third projects.

Analysis of self and peer evaluations indicates a general gap between self and peer evaluations with self evaluations consistently being greater than average peer evaluations. The gap was observed to narrow over time indicating a greater agreement at the end of the course. At the team level, large gaps between self and the average peer evaluations, Figure 5, were observed to be an indication of a dysfunctional team.

Discussion

The repeated measures MANOVA analysis supports the hypothesis that, from the perspective of peers, critical team skills improve over the course of a semester. This conclusion is consistent with results reported by McGourty, Dominick, and Reilly¹⁸. Dominick, Reilly, and McGourty¹⁹ also report that exposure to and completion of a feedback instrument itself drives change. Feedback is not a necessary component according to their research. The following paragraphs discuss reliability and validity issues from two perspectives; the critical team skills instrument, and the peer assessment process.

Critical Team Skills Instrument

The reliability of the instrument was estimated by calculating Cronbach's α . The values for Cronbach's α for the first, second, and third measurements are 0.89, 0.97 and 0.94 respectively. High values were expected since the items are all related to the same construct, critical team skills. Based on Cronbach's α calculations, instrument reliability is not an issue.

Content validity for the team skill measures is the assessment of whether the items in the instrument adequately represent the construct team skill. There are numerous examples in the literature to support the identified critical team skills¹⁵. The detailed descriptions for each labeled team skill are also in agreement with the literature. Specifically the identified critical team skills fall into the team generic and task generic categories established by Cannon-Bowers et al.¹⁵. This means that the skills being measured are transportable and suited for our objective

of developing the ability to work on multidisciplinary teams. This does not override the fact that the measured team skills are context driven having taken place on a specific team and task.

Peer Assessment Process

How people perceive others is one of the oldest topics in social and personality psychology²⁰. Validity, or accuracy, was challenged in the mid-1950s by Cronbach and Gage^{21,22}. These psychometricians argued that validity measurement techniques require that sources of variance in interpersonal perception be partitioned such as is done in generalizability theory. Cronbach²³ and Kenny²⁰ present statistical models that are applications of generalizability theory. The model developed by Kenny and La Voie²⁴ is suited for “round-robin” designs where each member of a group rates each other member of the group. Self-evaluations are excluded in the round-robin design. Montgomery²⁵ partitioned the source of variance in a group peer evaluation study using an early version of Kenny’s round-robin statistical program. In the Montgomery study, when the sources of variation are categorized into potential assessment error (rater) versus those that reflect the actual behavioral variation, only 9% of the source of error were assigned to potential assessment error. As reported, these are measures of consensus (agreement) and are not a measure of validity. Validity in the Montgomery study was high. This is based on the correlation of the average of the peer ratings with the average of trained observer ratings being 0.71 and 0.79 for two occasions respectively. Finally, peer ratings depend on the individuals constituting the peer group^{20,26}. This has implications for future research on the reliability and validity of peer assessments on multidisciplinary teams where interdependency is high.

An evaluation of the reliability and validity of peer evaluations was not investigated in this study. Peer ratings, believed to be the best method for providing feedback, are also reported to produce the least valid, reliable, and unbiased measurements when compared to peer nominations and peer rankings^{27,28}. In a peer rating system, each member of a group rates fellow group members on specific dimensions of performance using a rating scale. Kane and Lawler²⁸ reported the median reliability and validity results from 14 studies. The median correlation measuring internal consistency was 0.45 while the median validity coefficient was 0.35. Some of the reasons for the relatively low correlation in peer ratings is believed to be rater bias (how raters rate people in general – e.g. easy grader versus hard grader), real disagreement among raters, and response set or halo error. Halo error in the context of performance appraisal is defined as the tendency of raters to allow a general impression to affect their ratings of individual dimensions²⁹. The result is high inter-dimension correlation and a corresponding decrease in the validity at the item level. A review of the individual peer assessments in this study indicates that halo error is a problem in some of the evaluations. This provides additional justification for using a composite indicator of team skills rather than a more complex design since conclusions at the item level may not be justified.

The purpose of a peer evaluation is also reported to influence the validity of peer evaluations. When the purpose of the peer evaluation is for development compared to evaluation (grading), the validity is enhanced^{29,30}. Peer ratings used for developmental purposes tend to be more

reliable, more valid, and less susceptible to rating bias than ratings collected for evaluative purposes. The assessment data in this study has a developmental and evaluative purpose. This fact may weaken the validity of the study results.

Anonymity of the source of the evaluation versus public disclosure is another critical variable in peer assessments. From the perspective of the feedback recipient, accountability supports the public disclosure of the source especially if the news is good. On the other hand, the opposite is true for the feedback giver especially if the feedback is negative. The source of peer feedback in this study is anonymous and is believed to enhance validity of the study results.

Conclusions and Recommendations

Based on the data collected in this study, the repeated measures design demonstrates that team skills improve over the period of a one semester-long course. Also, identifying large differences between self and peer evaluations is a help in diagnosing dysfunctional teams. As a result, it is recommended that mid-project peer and self-evaluations be conducted.

Finally, it is recommended that generalizability theory be applied to peer ratings in multidisciplinary teams to help answer the following questions:

1. What is the level of consensus in peer ratings?
2. How much of the variance in peer ratings is due to how raters rate people in general?
3. How much of the variance in peer ratings is unique between raters and ratee's?
4. How much of the variance is error (random and can not be explained)?
5. Are the variances stable over time?

Answers to these questions will help define the appropriate role of peer evaluations in teamwork education.

Acknowledgements

The author would like to thank the faculty team for the multidisciplinary course: John Curtis, Geology and Geological Engineering; Robert Benson, Geophysics; and Tom Davis, Geophysics.

Table 1
Critical Team Skills

Back-up Behavior

Back-up Behavior involves assisting the performance of other team members. This implies that members have an understanding of other members' tasks. It also implies that members are willing and able to provide and seek assistance.

Communication

Communication involves the exchange of information between two or more team members using proper terminology. The purpose of communication is to clarify or acknowledge the receipt of information, to propose alternative perspectives, to propose decision criteria, and to make recommendations.

Feedback

Feedback could relate to group process or to task-related issues. Feedback involves the giving, seeking, and receiving of information among members. Giving feedback refers to providing information regarding other members performance. Seeking feedback refers to accepting positive and negative information regarding performance.

Coordination

Coordination refers to team members' executing their activities in a timely and integrated manner. It implies that the performance of some team members influences the performance of other team members. This may involve an exchange of information that subsequently influences another members' performance.

Team Leadership

Team Leadership involves providing direction, structure, and support for other team members. It does not necessarily refer to a single individual with formal authority over others. Team Leadership can be shown by several team members.

Team Orientation

Team Orientation refers to the attitudes that team members have towards one another and the team task. It reflects acceptance of team norms, level of group cohesiveness, and importance of team membership.

Table 2
Number of Subjects Organized by Discipline and Gender

Discipline	Gender	Number of Subjects
Geology & Geological Engineering	Male	8
	Female	3
Geophysical Engineering	Male	3
	Female	2
Petroleum Engineering	Male	20
	Female	2

Table 3
Three-Level One-Factor Within Subjects Repeated measures Design

Subject	Composite Team Skill Measurement		
	Occassion 1	Occassion 2	Occassion 3
1			
2			
:			
:			
38			

Table 4
Three Level One-Factor Within Subjects and One-Factor Between Subjects Repeated measures Design

Discipline	Subject	Composite Team Skill Measurement		
		Occassion 1	Occassion 2	Occassion 3
Geology and Geophysics	1			
	:			
	16			
Petroleum Engineering	17			
	:			
	38			

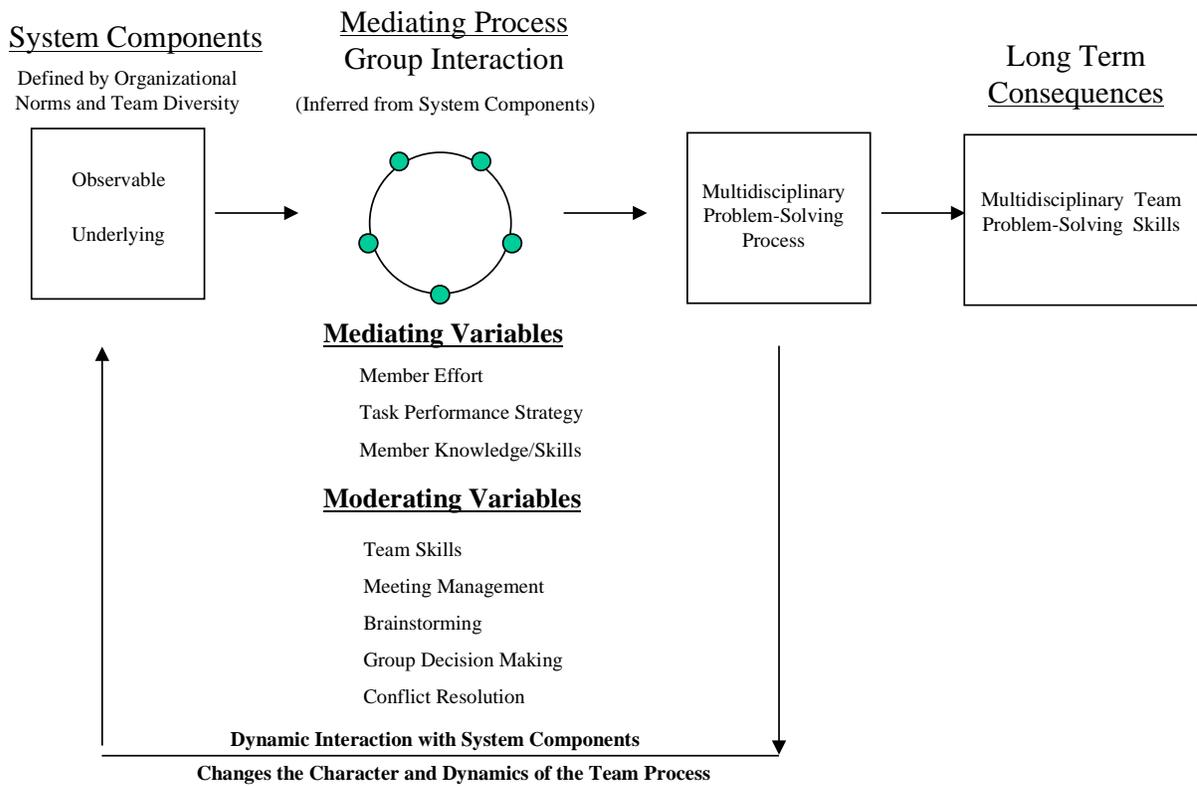


Figure 1: Model for multidisciplinary teamwork

Task Performance Strategy for Multidisciplinary Teams

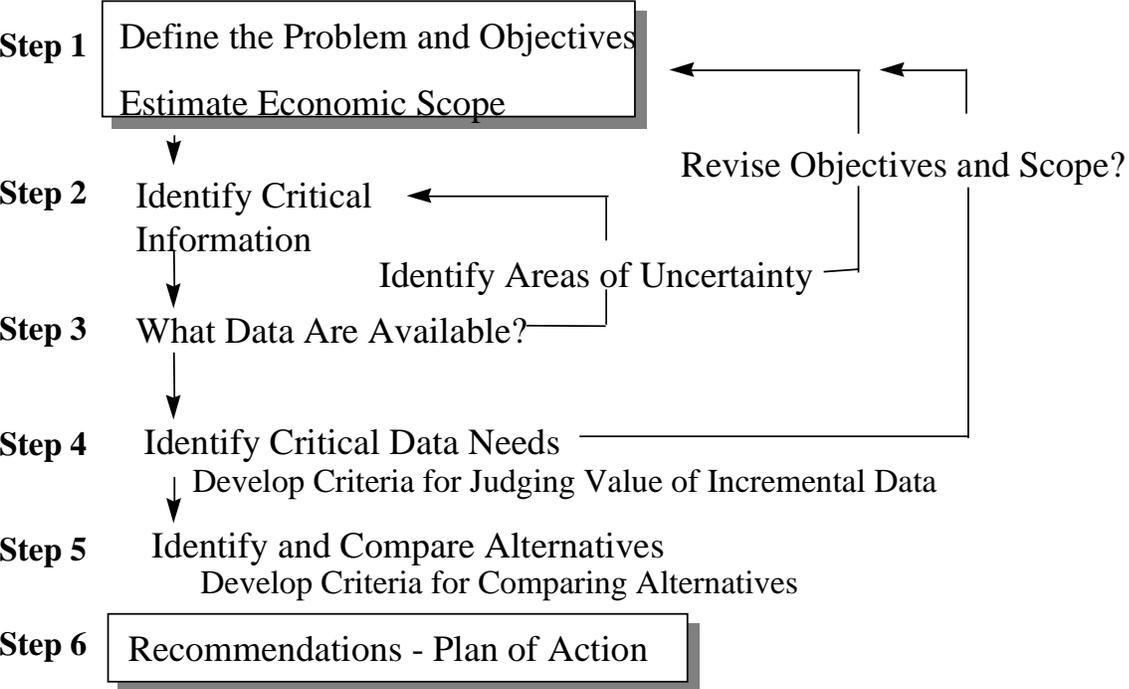


Figure 2: Task performance strategy for multidisciplinary teams

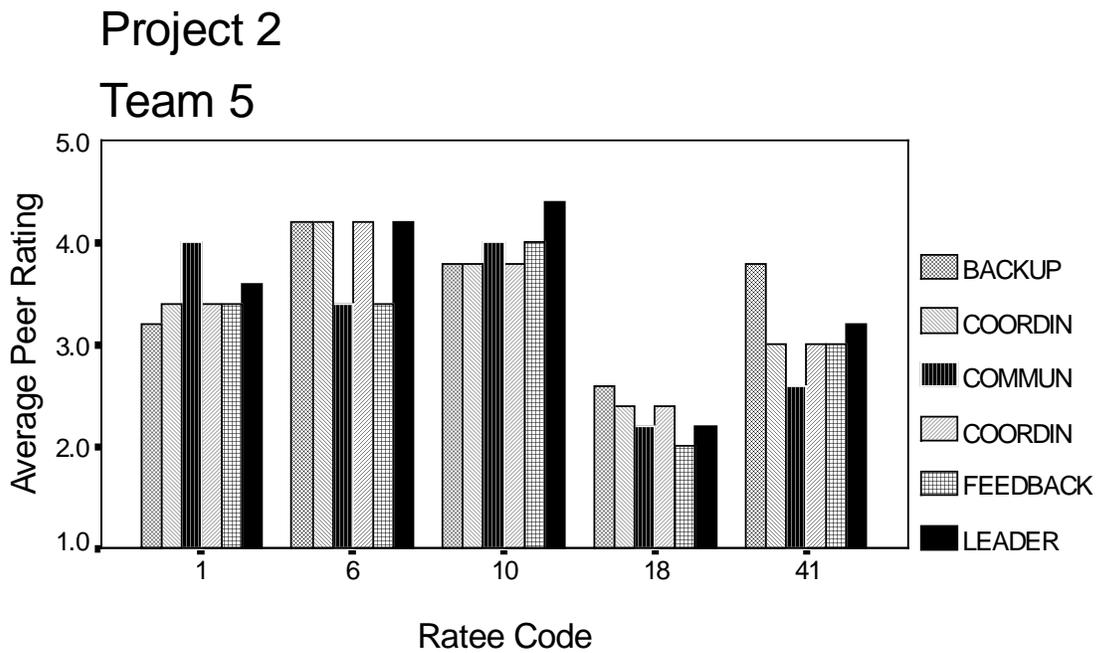


Figure 3: Example of peer evaluation feedback provided at the end of each project.

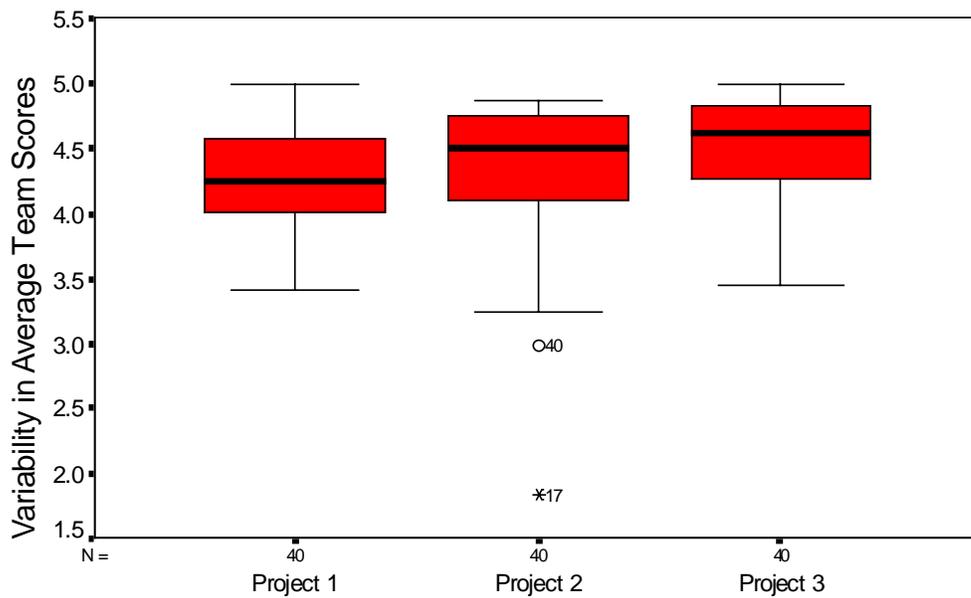


Figure 4: Box plot for average team skills used to identify outliers

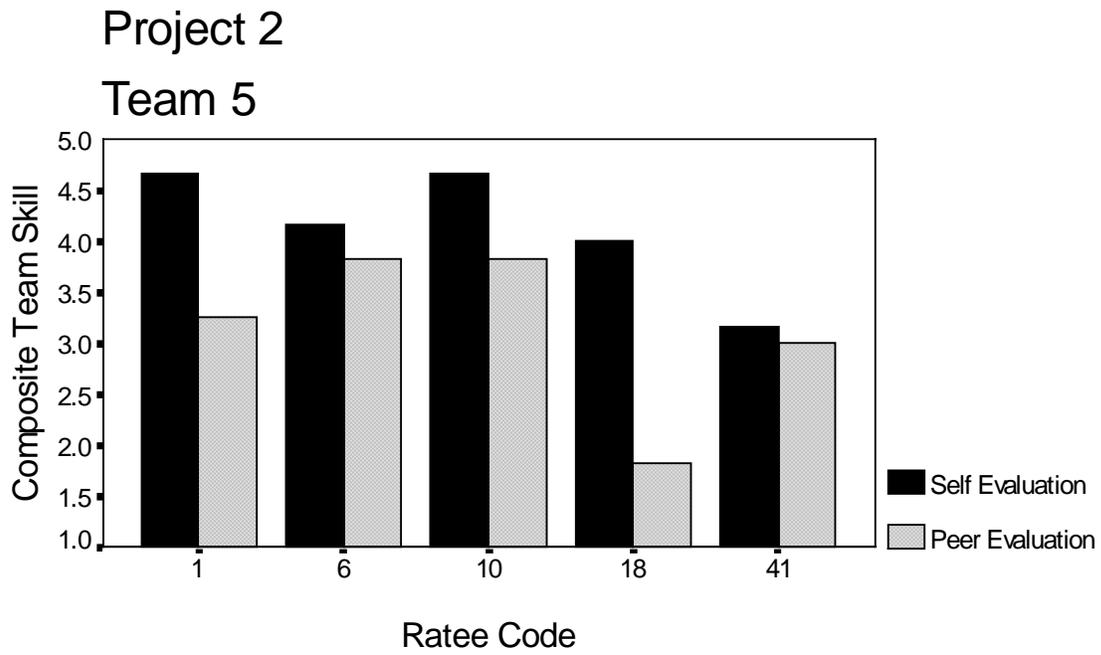


Figure 5: Example of large difference(s) between average peer evaluation and self-evaluation

Bibliography

1. ABET, "ABET engineering criteria 2000," (New York: Accreditation Board for Engineering and Technology, Inc., 1994).
2. B. M. Olds, M. J. Pavelich, and F. R. Yeatts, "Teaching the design process to freshmen and sophomores," *Journal of Engineering Education* (1990): 554-559.
3. John Sutton and Robert S. Thompson, "Multidisciplinary integration: A decision methodology and procedure for instruction" (paper presented at the Frontiers in Education Conference, Tempe, Arizona, November 4-7 1998).
4. Connie Gersick and Mary Lou Davis-Sacks, "Summary: Task forces," in *Groups that work (and those that don't)*, ed. J. Richard Hackman (San Francisco: Jossey-Bass Inc., 1990), 146-153.
5. Morton Deutsch, "A theory of co-operation and competition," *Human Relations* II, no. 2 (1949): 129-152.
6. Morton Deutsch, "An experimental study of the effects of co-operation and competition upon group process," *Human Relations* II, no. 3 (1949): 199-232.
7. David W. Johnson and Frank P. Johnson, *Joining together: Group theory and group skills* (Needham Heights, MA: Allyn & Bacon, 1997).
8. David W. Johnson and Roger T. Johnson, *Cooperation and competition: Theory and research* (Edina, Minnesota: Interaction Book Company, 1989).
9. Robert E. Slavin, *Cooperative learning: Theory, research, and practice* (Prentice Hall, 1990).
10. David W. Johnson, Roger T. Johnson, and Karl A. Smith, *Active learning: Cooperation in the college classroom* (Edina, MN 55435: Interaction Book Company, 1991).
11. Bertram H. Raven and Jan Rietsema, "The effects of varied clarity of group goal and group path upon the individual and his relation to his group," *Human Relations* 10 (1957): 29-45.
12. J. Richard Hackman and Charles G. Morris, "Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration," in *Advances in experimental social psychology*, ed. Leonard Berkowitz (New York: Academic Press, 1975), 45-99.
13. Reuben M. Baron and David A. Kenny, "The moderator-mediator variable distinction in social psychology research: Conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology* 51, no. 6 (1986): 1173-1182.
14. B. B. Morgan et al., "Measurement of team behaviors in a Navy environment," (Orlando, Florida: Navel Training Systems Center, Human Factors Division, 1986).
15. Janis A. Cannon-Bowers et al., "Defining competencies and establishing team training requirements," in *Team Effectiveness and Decision Making in Organizations*, ed. Richard A. Guzzo and Eduardo Salas and Associates, Frontiers of Industrial and Organizational Psychology (San Francisco: Jossey-Bass Publishers, 1995), 333-380.
16. Rebecca L. Anhalt, "Exploring the relationship between personality characteristics and effective teamwork behaviors" (Doctor of Philosophy, Colorado State University, 1995).
17. Joseph F. Hair et al., *Multivariate Data Analysis*, Fourth Edition ed. (Englewood Cliffs: Prentice Hall, 1995).
18. Jack McGourty, Peter Dominick, and Richard R. Reily, "Incorporating student peer review and feedback into the assessment process" (paper presented at the Frontiers in Education 1998, Tempe, AZ, 1998).
19. Peter Dominick, Richard R. Reily, and Jack McGourty, "The effects of peer feedback on team member behavior," *Group & Organization Management* 22, no. 4 (1997): 508-520.
20. David A. Kenny, *Interpersonal perception: A social relations analysis*, ed. Kurt W. Fischer, Tory E. Higgins, and Lawrence A. Pervin, *Distinguished Contributions in Psychology* (New York: The Guilford Press, 1994).
21. Lee J. Cronbach, "Processes affecting scores on "understanding of others" and assumed similarity," *Psychological Bulletin* 52, no. 3 (1955): 177-193.
22. N. L. Gage, George S. Leavitt, and George C. Stone, "The intermediary key in the analysis of interpersonal perception," *Psychological Bulletin* 53, no. 3 (1956): 258-267.
23. Lee J. Cronbach et al., *The dependability of behavioral measurements: Theory of generalizability for scores and profiles* (New York: John Wiley & Sons, Inc., 1972).
24. D. A. Kenny and L. La Voie, "The social relations model," in *Advances in experimental social psychology*, ed. L. Berkowitz (Orlando, FL: Academic Press, 1984), 142-182.
25. Barbara M. Montgomery, "An interactionist analysis of small group peer assessment," *Small Group Behavior* 17, no. 1 (1986): 19-37.
26. D.W. Fisk and J. A. Cox, "The consistency of rating by peers," *Journal of Applied Psychology* 44 (1960): 11-17.
27. K. G. Love, "Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction," *Journal of Applied Psychology* 66 (1981): 451-457.

28. J. S. Kane and E.E. Lawler, "Method of peer assessment," *Psychological Bulletin* 85 (1978): 555-586.
29. Jiing-Lih Farh, Albert A. Jr. Cannella, and Arthur G. Bedeian, "Peer ratings: The impact of purpose on rating quality and user acceptance," *Group & Organization Studies* 16, no. 4 (1991): 367-386.
30. Sonia Maria Rodrigues Calado, "The impact of system design factors on users' reaction to peer feedback" (Doctoral, Georgia State University, 1994).

ROBERT S. THOMPSON

Robert Thompson is an Associate Professor in Petroleum Engineering at the Colorado School of Mines (CSM). Robert received his professional degree in petroleum engineering from CSM and his MBA from the University of Houston. He is currently a PhD candidate in the Educational Leadership and Innovation Program at the University of Colorado at Denver where he is conducting research on cooperative learning in multidisciplinary teams.