Designing a Peer Evaluation Instrument that is Simple, Reliable, and Valid

Matthew W. Ohland, Misty L. Loughry, Rufus L. Carter, and Amy G. Yuhasz
General Engineering, Clemson University / Management, Clemson University / Institutional Research and Assessment, Marymount University / General Engineering, Clemson University

Abstract
As a result of ABET's EC 2000 Criterion 3, outcome (d), "an ability to function on multi-disciplinary teams"[1] this multi-university research team has focused its attention on teamwork and how it is assessed. Teamwork in engineering is often assessed using a peer evaluation instrument. It is not always clear, however, what characteristics of teamwork these instruments, or the students, are evaluating. In preparation for this multi-year NSF-supported project, the team reviewed peer evaluation literature and instruments. The research team has an ambitious assessment plan that will help develop an instrument that is easy to use and yet meaningful for both faculty and students.

Introduction
In recent years, there has been a great deal of activity in engineering education research aimed at evaluating teamwork. Much of this is a result of the need to measure ABET's EC 2000 Criterion 3, outcome (d), "an ability to function on multi-disciplinary teams."[1] While there has been considerable debate on how to apply the term "multi-disciplinary," the ability to function on a team is central to this outcome.

Though an effort to achieve this ABET outcome is sufficient motivation for many instructors to evaluate teamwork in some way, peer evaluation that assesses each individual's contributions to a team has the additional objective of promoting a productive cooperative learning environment. Cooperative learning (CL) is an instructional paradigm wherein teams of students work on structured tasks (e.g., homework assignments, laboratory experiments, or design projects) under conditions that meet five criteria: positive interdependence, individual accountability, face-to-face interaction, appropriate use of collaborative skills, and regular self-assessment of team functioning. Many studies have shown that when correctly implemented, cooperative learning improves information acquisition and retention, higher-level thinking skills, interpersonal and communication skills, and self-confidence.[2]

Many cooperative learning advocates agree that the approach works best if team grades are adjusted for individual performance. Millis and Cottell give five strong reasons for using peer evaluation to adjust student grades:

1. "Teachers, because they are not the sole arbiters of success or failure, play less of a gatekeeper role responsible for weeding out the unfit and the unworthy. The process of evaluation is shared.
2. Students are in a logical position to judge the individual contributions of their peers far more effectively than an instructor can.

3. Peer feedback is usually directed toward an individual within the context of a specific task. Besides being context-specific, it tends to be delivered promptly which is when feedback is most effective.
4. Peer evaluation builds in accountability: students realize that they are held accountable for their academic achievements and group contributions. They may be able to "psyche out" a teacher, but they can rarely hide from their peers.
5. Students benefit from the process of peer review. They learn valuable lessons about the learning process and about teamwork efforts." [3]

In order to more effectively implement the cooperative learning model, faculty must measure teamwork contribution and use it to appropriately adjust course grades. While each individual's contribution to the team can be measured in other ways, such as by instructor observation, peer evaluation holds the most promise for enhancing the cooperative learning environment, because it engages the students themselves in the evaluation process.

Evaluating teamwork in engineering education

An "autorating" (peer rating) system designed to account for individual performance in cooperative learning teams was developed at the Royal Melbourne Institute of Technology by Robert Brown.[4] Using this system, team members confidentially rate how they and their teammates fulfilled their responsibilities, selecting ratings from nine terms ranging from "excellent" to "no show." Students are instructed to rate team citizenship and not ability or contribution to the project. Numerical values are assigned to each rating ("Excellent" = 100, "Very Good" = 87.5,…,"No show" = 0) and a weighting factor is a student's individual average rating divided by the team average. Weighting factors are generally between 0.95 and 1.05, where 1.0 indicates that a student meets expectations. To give these factors less weight, some take the square root of the factor before using it to compute individual grades from teamwork.

Brown's system was used in two consecutive sophomore-level chemical engineering courses at North Carolina State University as reported by Kaufman, Felder (senior personnel in this study), and Fuller (KFF), who found no gender bias in the ratings but did find a possible racial bias.[5,6] Layton (Co-PI) had been using an instrument designed at North Carolina A&T State University in collaboration with Sam Ofori and Devdas Pai. He experimented with a modified version of the KFF instrument, and partnered with Ohland (PI) to reproduce the KFF study using results from both instruments. Layton and Ohland confirmed the lack of a gender bias and discovered that the same racial bias (blacks rate whites higher than whites rate blacks) was present at NC A&T.[7] In an extension of this study, Ohland and Layton compared the reliability of the two instruments to better understand the quality of the scoring rubric.[8] Layton followed up with a study indicating that an improvement in test administration (giving the students instruction about the elements of team citizenship) eliminated the racial bias.[9] This follow-up study included very few women, resulting in an unusual distribution of scores by gender. Finelli (Co-PI) joined the research team in 2001, using the instrument at Kettering University. The Kettering study showed a restriction of the range of peer ratings, possibly because required co-op participation helps develop improved teamwork skills (the peer ratings were grouped together because co-op experience helps all of them develop their team skills).[10,11] Use of a modified form of this instrument was

recently reported at Smith College. Teams were formed as heterogeneous with respect to MBTI type and the peer rating instrument was used to assign individual grades from group grades.[12]

Perhaps the most comprehensive instrument for assessing teamwork is the *Team Developer*™, a computer-based survey developed by Jack McGourty of Columbia University that provides students with multi-source assessment and feedback. All team members rate themselves and their teammates, and each team member receives a summary report of their ratings. The report highlights the differences between self-perceptions and perceptions of teammates, and the feedback includes recommendations aimed at improving teamwork skills. *Team Developer*™ is available for purchase and has been used on a number of campuses, including New Jersey Institute of Technology (NJIT), the Ohio State University, and the University of Pittsburgh.[13-15]

While the limited success of *Team Developer*™ is an indication of the demand for the meaningful evaluation of teamwork, there is a great need for a simpler instrument to supplement more complex tools such as *Team Developer*™ and verbal protocol analyses that require both expertise and a greater time investment to administer properly.

Investigating Instrument Validity

Validity describes whether an instrument measures what it is supposed to measure.[16] The validity of peer ratings can be (and often is) questioned. Common concerns are that individuals will inflate their self-ratings; team members will agree to give everyone identical ratings to avoid conflict; and gender or racial bias and personal dislikes might influence the ratings. This project will assess and improve the validity of the measurement of teamwork in this context.

The most serious challenge to validity is the lack of an *anchor metric*—a true measure of teamwork. One way to overcome this challenge is by designating some expert observation as an anchor metric. This approach was used by Robert Thompson and his colleagues at the Colorado School of Mines in conducting peer evaluations in the Multidisciplinary Petroleum Design Course, in which faculty observations established the "correct" target at which peer evaluations were aiming.[17] While Thompson's approach is rigorous and has much to offer the discussion of peer evaluation, his conclusions hinge upon the assumption that the faculty observations were a true measure of teamwork, in spite of the fact that the faculty only observes a sample of the team activity. As such, that approach may be inappropriate in this study, and *triangulation* may be a more useful process. This approach uses multiple assessment methodologies to measure an outcome. Each of the measures of teamwork is a surrogate for the true measure, but if the various surrogates concur, each method is strengthened in the process.

The triangulation process in this study will be achieved by measuring the *concurrent validity* of the new peer evaluation instrument compared to other measures of teamwork in multiple experiments. Concurrent validity measures the level of concurrence between two measures of teamwork. If the instrument created in this study shows concurrent validity with a number of other approaches, then triangulation tells us that all of the approaches, including the new instrument, are measuring teamwork. Concurrent validity will be tested with the following instruments:

- Van Duzer and McMartin's instrument[18]
- The BESTEAMS Peer Evaluation form[19]
- The Team Developer™ [13-15,20]

A portion of the Van Duzer and McMartin's instrument asks the rater to distribute a fixed number of points (or %) among the teammates according to their team citizenship. This type of instrument, which will be referred to as a point-distribution system, will be tested for concurrence apart from the rating of various team attributes on independent scales, the other major portion of the Van Duzer and McMartin instrument. The reaction of the rater to these two systems is expected to be very different—the authors believe that point-distribution systems cause students to explicitly take points away from one student to give them to another, making the consequences of downgrading a teammate's citizenship to quantitative and personal.

Content validity asks whether the new instrument measures teamwork, rather than academic ability, popularity, or some other factor. Van Duzer and McMartin describe the importance of content validity for establishing a common and explicit language that defines the traits and the criteria.[18] While content validity cannot be measured in a quantitative sense, a two-step process will be used in this study to ensure content validity. The instrument will be reviewed to make it free from language that will bias the instrument. Before administration, following Van Duzer and McMartin's model, the instrument will be studied with the help of students, who will be individually interviewed using verbal protocols and asked to describe their interpretation of each question and of the rating scale. Representatives of University of Washington's Center for Engineering Learning and Teaching (CELT) will serve as unpaid consultants in the design of the verbal protocol analysis. A letter of support has been provided in supplementary documentation.

Investigating Instrument Reliability

Reliability is a measure of whether an instrument yields the same or similar scores consistently, presuming that all measured factors are unchanged. *Test-retest* reliability measures the correlation of two administrations of the same instrument. The primary drawback of this method is that it is sensitive to the interval between administrations—if the second is administered too soon, the student's memory of the first administration will affect the outcome of the second administration, but if too much time passes before the second administration, the trait being studied may change. The authors expect team citizenship to improve as teams work together, so a time-series design with four administrations in the same class, separated by about 2-½ weeks will be used.

Instruments that collect multiple observations for each subject using the same measures interpreted by different individuals can, and indeed should, use a special form to estimate their reliability, measuring the consistency of the observations from one rater to another—how well different students agree on the rating of a particular teammate. Since peer evaluation within a team yields multiple measures (by each teammate) of the same traits (aspects of team citizenship), this inter-rater reliability will be studied using analysis of variance in a special form by Crocker and Algina.[21] The research team has experience using this method, as published by Ohland and Layton.[8] That study compared the inter-rater reliability of two peer evaluation instruments, the instrument that measured multiple components of team citizenship (and

therefore had multiple items) had a better reliability coefficient for a single rater's score (0.41) than that obtained for an instrument similar to that used by Kaufman, Felder (senior personnel), and Fuller,[5] which used a single, compound item, yielding a coefficient of 0.34. This coefficient is an indication of how well a single rater's score approximates the true score that would be obtained if a large number of raters evaluated the same student—the overall reliability of the scoring method increases with additional raters. In the case of the instruments investigated by Ohland and Layton, the overall scoring reliability scales with the number of raters as follows:

Number of raters	Scoring reliability of multiple-item instrument[8]	Scoring reliability of single-item instrument[8]	Scoring reliability of single-item instrument with improved administration[8]
1	0.41	0.34	0.47
2	0.59	0.51	0.64
3	0.68	0.61	0.73
4	0.74	0.67	0.78
5	0.78	0.72	0.82

Since reliability of the single-item instrument was improved simply by improving the administration instructions given to the students, there are two approaches to choose between:

- Improve and test the reliability and validity of the single-item instrument or
- Improve the administration (and the design) of the multiple-item instrument to yield even higher reliabilities, yet still produce an instrument that is simpler than the *Team Developer*™ and others.

Typically, student teams in engineering classes are comprised of no more than five persons, and four is probably the most common team size. The maximum scoring reliability in the table above for a four-person team is 0.78. This research team will take the second approach, designing a multiple-item instrument with a well-designed administration. The multiple-item instrument should also yield better validity by separately assessing the multiple components of team citizenship, rather than asking students to summarize them in a single measure.

The Need for Further Study of Validity and Reliability

There is little published regarding the validity and reliability of instruments used for peer evaluation in engineering education. The table below makes the need for work in this area clear. Some instruments have no published validity or reliability testing at all.

Instrument	Validity testing	Reliability testing	Format
BESTEAMS	None published	None published	Likert scale, paper-and-pencil
Team Developer™	Trained observers Convergent / discriminant	Inter-rater reliability Internal consistency	Likert scale, computer-based
Thompson	Faculty observation	None published	Not described

Van Duzer and McMartin	None published	None published	Likert scale, paper-and-pencil
Instruments used by this team	None	Inter-rater reliability	Verbal descriptors, paper-and-pencil

Clearly, additional research is needed regarding the validity and reliability of peer evaluation instruments in engineering education. Compounding the weak research base in the validity and reliability of such instruments is that some of what is published is fraught with inappropriate statistical design. The most common error is the use of the t-test to compare the mean of scores on a Likert scale. Since Likert scales do not provide interval data, there is no reason for thinking that the difference between, 2 and 3 on the scale is the same as the difference between 4 and 5 on the same scale. Even the use of the mean can be questionable when using a Likert scale. Another failing of most instruments is that the training process for administration is either unpublished or lacking altogether.

The Need for a Clearly Defined Rubric and Test Instructions

The issues of validity and reliability both point to the critical process of defining a clear scoring rubric and set of administration instructions for the instrument. When discussing the interpretation of reliability coefficients, Kubiszyn and Borich outline four principles:[16]

1) Reliability coefficients are directly proportional to group variability
2) Scoring reliability establishes an upper bound on the instrument reliability
3) More instrument items lead to higher reliabilities (other things being equal)
4) Reliability decreases when scores are collectively too high or too low

Of these, the fourth principle is a subset of the first—if scores are too high or too low, range restriction will occur, causing a reduction in the group variability. The effects of the first principle have already been observed in practice by Ohland and Finelli in a study at Kettering University.[10] In that study, scores lacked variability (perhaps because they were too high, although this was not significant compared to previous studies by Layton and Ohland[7,9]).

These principles underscore the approach this team will take in developing the new instrument. Since the scoring reliability establishes an upper bound on the instrument reliability (per the second principle), the scoring reliability, as measured using inter-rater reliability methods, is a critical parameter. The third principle described above was observed and measured in practice by Ohland and Layton, so the new instrument will use multiple items to measure teamwork.

Van Duzer and McMartin point out that if social comparisons are made, they should be made with respect to an explicit group of known individuals (team members).[18] This calibrates student ratings to a common reference—the typical behavior on the team—but may still fall short as student perceptions of an individual's contribution differ. The authors propose calibrating a peer evaluation instrument in the same way other instruments are calibrated. Since students can vary in so many ways, it is difficult to establish a "standard reference" for each rating. Instead, calibration may be established empirically. If a series of brief case studies is designed that

describe a set of student team behaviors and then provide ratings of the case-study students using our instrument, student perceptions can then be "calibrated" to guard against common threats to validity and reliability. These case studies can be presented in a handout to the students or designed as a "quiz" that is electronically administered using course management software such as WebCT or Blackboard, giving feedback as to how the instrument designers interpret the measurement scale. This case study approach will be evaluated using a multiple-intervention approach, where at least two administrations will differ only in their use of the case studies.

Influencing the Theoretical Development of Peer Evaluation in Engineering Education

Research on the psychological type of engineering students suggests that the engineering culture is dominated by students who are individual achievers, but do not necessarily work well on a team. Two findings of the MBTI engineering consortium study are relevant here:[22]

- Focusing on the logical, analytical, and decisive traits diminishes the development of skills related to listening, understanding, and getting things done through people.
- Typical engineering students are not likely to give feedback to their peers.

Engineering employers regularly cite teamwork skills as foremost in the list of the qualities they look for in graduates and many historical studies aimed at the reform of engineering education have identified teamwork skills as critical.[23-27]

In light of these observations from theory and practice, it is understandable that previous work in peer evaluation in engineering education has focused on the assessment of team citizenship apart from ability. This approach seems to originate from a desire to measure the teamwork skills of students, a motive that has only be strengthened by the impetus of ABET's EC 2000 criteria.[1] While there are many ways to measure students' ability and effort, engineering faculty struggle for a measure of whether or not the student "works well with others."

In contrast to the approach commonly used in engineering education, peer evaluations in management research evaluate team citizenship, ability, and effort. Evaluating the performance of individuals working in teams is an important concern for management scholars.[28-30] Work in organizations is increasingly performed by groups of workers who are responsible for managing themselves, either because the organizational structure has been designed around self-managing work teams or because supervisors manage so many employees that workers are forced to manage themselves most of the time.[31,32] Workers monitoring their peers and enforcing group norms for effort creates accountability to teammates and discourages free riding.[33,34] Team members' monitoring of teammates is critical to effective team functioning.[35,36]

Management researchers who specialize in human resources have studied a personality trait called conscientiousness to predict employee performance. Conscientiousness reflects a tendency to be careful, dependable, responsible and achievement-oriented. Conscientiousness does not suffer from the race-based differences that plague other selection tools that are frequently used to predict performance. Conscientiousness has a small amount of predictive value for task performance and training proficiency.[37,38] Conscientiousness is more strongly associated with contextual performance,[39] which is also called organizational citizenship behavior.[40,41]

Organizational citizenship behavior refers to actions such as helping others, volunteering for extra work, making helpful suggestions, doing things that are right and proper, and complying with work rules. This line of human resources research supports the current research design and holds promise for designing instruments that eliminate racial bias.

In spite of the fact that many large employers use 360-degree performance evaluation systems in which employees rate their coworkers' performance, research about the validity and reliability of peer evaluations is very limited.[42,43] As a result, the outcome of this study has value in the management field as well.

Especially in light of potential personality type differences between engineering students and the employees that are the subject of study in management research, it would be inappropriate to adopt even the best work from management research without challenging whether it is appropriate for use by engineering students. Therefore, the research team will develop a model for how teamwork should be assessed in engineering education.

New peer evaluation instrument development

The new peer evaluation instrument is currently under development. While the final instrument to be developed for peer evaluation within engineering teams will use a "behaviorally anchored rating scale" (BARS) format, the team decided to use a Likert-scale instrument to help define the categories of teamwork that the BARS instrument should measure. In the Fall of 2003, an extensive list of statements, approximately 400, dealing with measures of individual contribution to teamwork were derived from theory in the "teams and groups" literature. All of the items describe behaviors that make individuals more effective team members. The master list was then given to members of the research team and a few graduate and undergraduate students (in management and various engineering disciplines) for their evaluation. These participants were asked to mark any statements that seemed ambiguous or unclear, group the items into categories that seemed similar, and give a name to each category. The research team then reviewed the participants' responses and re-worded items that were not interpreted as intended and wrote new items based on participants' comments. The research team then sorted the 218 items into 36 categories of ways that individuals contribute to teams. The categories seemed to fit into seven broader clusters.

Beginning in the November 2003, this list was administered to volunteer students (undergraduate and graduate students from all majors) at Clemson University. Students were asked to rate each of the statements on a 5-point Likert scale, indicating the "degree to which they agree or disagree that each item describes something that is essential for members of teams to do." When all of the survey data are collected, the researchers will use exploratory factor analysis to empirically derive categories of teamwork as perceived by students. The results of the exploratory factor analysis will also be used to reduce the number of statements to 3-4 items per category that have the highest reliability. It is expected that the final Likert-scale instrument will contain 100-120 items. Preliminary results from the Likert-scale instrument have already been used to develop a draft of the BARS instrument. This BARS instrument will undergo rigorous classroom testing by the members of the research team and then by other engineering faculty. Further results from

the Likert-scale survey and the factor analysis will influence the revision of the BARS scale, and there are other venues in which the reduced-item Likert-scale will find applicability.

Acknowledgements

This material is based upon work supported by NSF DUE-ASA Award Number 0243254, "Designing a Peer Evaluation Instrument that is Simple, Reliable, and Valid."

Author biographies

MATTHEW W. OHLAND
is an Assistant Professor in Clemson University's General Engineering program and is the President of Tau Beta Pi, the national engineering honor society. He received his Ph.D. in Civil Engineering with a minor in Education from the University of Florida in 1996. Previously, he served as Assistant Director of the NSF-sponsored SUCCEED Engineering Education Coalition. His research is primarily in freshman programs and educational assessment.

MISTY L. LOUGHRY
Misty L. Loughry is an Assistant Professor in Clemson University's Management Department. She received her Ph.D. in Management from the University of Florida in 2001. Her research focuses on control in organizations, especially peer monitoring. Prior to her academic career, Dr. Loughry worked in banking for ten years, most recently holding the position of Assistant Vice President of Small Business Lending.

RUFUS L. CARTER
is Coordinator of Institutional Assessment in the Office of Institutional Research, Assessment & Planning of Marymount University. He provides consulting services to this project as a measurement specialist.

AMY G. YUHASZ
is a Postdoctoral Fellow in Clemson University's General Engineering program. She earned her Ph.D. in Industrial Engineering from Clemson University in 2002 studying risk assessment for large industrial capitalization projects. In addition to applying her industrial engineering skills to studying engineering education, she maintains her research interest in industrial risk assessment.

References

1. Criteria for Accrediting Engineering Programs. Published by The Accreditation Board for Engineering and Technology (ABET), Baltimore, Maryland. Last accessed on November 5, 2001; http://www.abet.org/images/eac_criteria_b.pdf (2000).
2. Johnson, D.W., R.T. Johnson, and K.A. Smith, *Active learning: Cooperation in the college classroom*, Edina, MN: Interaction Book Co., (1998).
3. Millis, B.J. and P.G. Cottell, Jr., *Cooperative Learning for Higher Education Faculty*, Phoenix, American Council on Education/Oryx Press, p. 194 (1998)
4. Brown, R.W., "Autorating: Getting individual marks from team marks and enhancing teamwork," *Proc. Frontiers in Education Conference*. IEEE/ASEE, Pittsburgh, November (1995).
5. Kaufman, D.B., R.M. Felder, and H. Fuller, "Peer ratings in cooperative learning teams," *Proc. Amer. Soc. Eng. Ed.*, ASEE, Charlotte, June (1999).
6. Kaufman, D.B., R.M. Felder, and H. Fuller, "Accounting for individual effort in cooperative learning teams," *J. Engineering Education* **89**(2), 133-140 (2000).
7. Layton, R. A. and M.W. Ohland, "Peer evaluations in teams of predominantly minority students," *Proc. Amer. Soc. Eng. Ed.*, Session 2330, ASEE, Washington, DC, (2000).
8. Ohland, M.W., and R.A. Layton, "Comparing the reliability of two peer evaluation instruments," *Proc. Amer. Soc. Eng. Ed.*, St. Louis, MO, session 3530, June (2000).

9. Layton, R.A., and M.W. Ohland, "Peer evaluations revisited: Focus on teamwork, not ability," *Proc. Amer. Soc. Eng. Ed.*, Albuquerque, NM, June (2001).
10. Ohland, M.W., and C.J. Finelli, "Peer evaluation in a mandatory cooperative education environment," *Proc. Amer. Soc. Eng. Ed.*, Albuquerque, NM, June (2001).
11. Finelli, C.J., "Assessing improvement in students' team skills and using a learning style inventory to increase it," *Proc. Frontiers in Education Conference*, IEEE/ASEE, Reno, NV, October (2001).
12. Mikic, B. and D. Grasso, "Socially-Relevant Design; The TOYtech Project at Smith College," *J Engr Ed* **91**(3), July 2002, pp. 319-326.
13. McGourty, J. and K. De Meuse, *The Team Developer: An assessment and skill building program*, J. Wiley and Company, New York (2000).
14. McGourty, J., C. Sebastian, and W. Swart, "Development of a comprehensive assessment program in engineering education." *J. Engineering Education* **87**(4), 355-361 (1998).
15. McGourty, J., "Using multisource feedback in the classroom: A computer-based approach," *IEEE Trans. Ed.*, **43**(2), 120-124, (2000).
16. Kubiszyn, T. and G. Borich, *Educational testing and measurement: Classroom application and practice*, 4th Ed., Harper Collins, New York, (1993).
17. Thompson, R.S., "Relative validity of peer and self-evaluations in self-directed interdependent work teams," *Proc. Frontiers in Education Conference*, IEEE/ASEE, Reno, NV, October (2001).
18. Van Duzer, E. and F. McMartin, "Methods to improve the validity and sensitivity of a self/peer assessment instrument," *IEEE Trans. Ed.*, **43**(2), 153-158, May (2000).
19. Version 1, received from Pat Meade at *Frontiers in Education Conference*, Reno, NV, October (2001)
20. Seat, E., and T.P. McAnear, "Administering, scoring and debriefing Team Developer," *Proc. Frontiers in Education Conference*, IEEE/ASEE, Reno, NV, October (2001).
21. Crocker, L., and J. Algina, *Introduction to classical and modern test theory*, Holt, Rinehart and Winston, Inc., Chicago, p. 143, 157ff, (1986).
22. McCaulley, M. H., Godleski, E. S., Yokomoto, C. F., Harrisberger, L., and Sloan, E. D., "Applications of psychological type in engineering education," *Engineering Education* 394 (Feb. 1983).
23. Hammond, H. P., "Report of the Committee on Engineering Education After the War," J. Engineering Ed. 34, 1944, 589-614.
24. Hammond, H. P., "Report of the Committee on Aims and Scope of Engineering Curricula," J. Engineering Ed. 30, 1940, 555-566.
25. Haddad, Jerrier A., "Engineering Education and Practice in the United States: Foundations of Our Techno-Economic Future," Report of the Committee on the Education and Utilization of the Engineer, National Research Council, Washington, D.C.: National Academy Press, 1985.
26. Grintner, L. E., "Report of the Committee on Evaluation of Engineering Education," *J. Engineering Ed.* 44, September 1955, 26-60.
27. Wickenden, W. E., "Report of the Investigation of Engineering Education," Urbana, IL, Society for the Promotion of Engineering Education, 1930.
28. McGrath, J. E. *Social psychology: A brief introduction.* New York: Holt, Rinehart and Winston (1964).
29. Alchian, A. A., and Demsetz, H. "Production, information costs and economic organization." American Economic Review, **62**, 777-795 (1972).
30. Stevens, M. J., and Campion, M. A. "Staffing work teams: Development and validation of a selection test for teamwork settings." *Journal of Management*, **25**, 207-228 (1999).
31. Howard, A. "A framework for work change." In A. Howard (Ed.), *The changing nature of work* (pp. 3-44). San Francisco: Jossey-Bass (1995).
32. Ilgen, D. R., and Pulakos, E. D. (Eds.). *The changing nature of performance: Implications for staffing, motivation, and development.* San Francisco: Jossey Bass (1999).
33. Hackman, J. R. "Group influences on individuals in organizations." In M. A. Dunette and L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 199-267). Palo Alto, CA: Consulting Psychologists Press (1992).
34. Guzzo, R. A., and Shea, G. P. "Group performance and intergroup relations in organizations." In M. D. Dunette and L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 269-313). Palo Alto, CA: Consulting Psychologists Press (1992).

35. Guzzo, A., and E. Salas (Eds.) *Team effectiveness and decision making in organizations.* San Francisco: Jossey-Bass (1995).
36. Hedge, J. W., Bruskiewicz, K. T., Logan, K. K., Hanson, M. A., and Buck, D. *Crew resource management team and individual job analysis and rating scale development for air force tanker crews* (Technical Report No. 336). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (1999).
37. Barrick, M. R., and M. K. Mount, "The big five personality dimensions and job performance: A meta-analysis," *Personnel Psychology* **44**, 1-27 *(*1991).
38. Salgado, J. F. "The five factor model of personality and job performance in the European community, *Journal of Applied Psychology*, **82**(1), 30-43.
39. Motowidlo, S. J., W. C. Borman, and M. J. Schmit, "A theory of individual differences in task and contextual performance," *Human Performance* **10**(2), 71-83.
40. Smith, C. A., D. W. Organ, and J. P. Near, "Organizational Citizenship Behavior: Its nature and antecedents," Journal of Applied Psychology **68**(4), 653-663.
41. LePine, J. A., A. Erez, and D. E. Johnson, "The nature and dimensionality of organizational citizenship behavior: a critical review and meta-analysis," Journal of Applied Psychology, **87**(1), 52-65.
42. Church, A. H., and Bracken, D. W. "Advancing the state of the art of 360-degree feedback: Guest editors' comments on the research and practice of multirater assessment methods." *Group and Organization Management*, **22**, 149-161 (1997).
43. Fedor, D. B., Bettenhausen, K. L., and Davis, W. "Peer reviews: Employees' dual roles as raters and recipients." *Group and Organization Management*, **24**, 92-120 (1999).
44. Millis, B.J., and P.G. Cottell, Jr., *Cooperative learning for higher education faculty*, American Council on Education Oryx Press Series on Higher Education (1997).
45. Felder, R.M., and L.K. Silverman, "Learning and teaching styles in engineering education," *Engineering Education*, 674-681*,* April (1988).
46. Wankat, P., "An analysis of the articles in the Journal of Engineering Education," *J. Engr. Ed.* **88**(1), 37-42 (1999).
47. http://diamond.gem.valpo.edu/~harvey/teams/index.html