# Optimizing Student Team Skill Development using Evidence-Based Strategies: Year 3: NSF Award 1431694

**Dr. Matthew W. Ohland, Purdue University, West Lafayette (College of Engineering)**

Matthew W. Ohland is Professor of Engineering Education at Purdue University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by the National Science Foundation and the Sloan Foundation and his team received Best Paper awards from the Journal of Engineering Education in 2008 and 2011 and from the IEEE Transactions on Education in 2011 and 2015. Dr. Ohland is Chair of the IEEE Curriculum and Pedagogy Committee and an ABET Program Evaluator for ASEE. He was the 2002–2006 President of Tau Beta Pi and is a Fellow of the ASEE, IEEE, and AAAS.

**Dr. Misty L. Loughry, Rollins College**

Misty L. Loughry, Ph.D. is a Professor of Management at Rollins College in the Crummer Graduate School of Business.

**Dr. David Jonathan Woehr, University of North Carolina, Charlotte**

David J. Woehr is currently Professor and Chair of the Department of Management at The University of North Carolina at Charlotte. He received his Ph.D. in Industrial/Organizational Psychology from the Georgia Institute of Technology in 1989. Dr. Woehr served on the faculty of the Psychology Department in the I/O Psychology program at Texas A&M University from 1988 to 1999 and as a Professor of Management at the University of Tennessee from 1999 to 2011. He has also served as a Visiting Scientist to the Air Force Human Resource Laboratory and as a consultant to private industry. Dr. Woehr is a fellow of the Society for Industrial and Organizational Psychology (SIOP), the American Psychological Association (APA), and the Association for Psychological Science (APS). His research on managerial assessment centers, job performance measurement, work related attitudes and behavior, training development, and quantitative methods has appeared in a variety of books, journals, as papers presented at professional meetings, and as technical reports. Dr. Woehr currently serves as editor for Human Performance as well as on the editorial boards for Organizational Research Methods, and the European Journal of Work and Organizational Psychology

**Dr. Daniel M. Ferguson, Purdue University, West Lafayette (College of Engineering)**

Daniel M. Ferguson is CATME Managing Director and the recipient of several NSF awards for research in engineering education and a research associate at Purdue University. Prior to coming to Purdue he was Assistant Professor of Entrepreneurship at Ohio Northern University. Before assuming that position he was Associate Director of the Inter-Professional Studies Program [IPRO] and Senior Lecturer at Illinois Institute of Technology and involved in research in service learning, assessment processes and interventions aimed at improving learning objective attainment. Prior to his University assignments he was the Founder and CEO of The EDI Group, Ltd. and The EDI Group Canada, Ltd, independent professional services companies specializing in B2B electronic commerce and electronic data interchange. The EDI Group companies conducted syndicated market research, offered educational seminars and conferences and published The Journal of Electronic Commerce. He was also a Vice President at the First National Bank of Chicago [now J.P. Morgan Chase], where he founded and managed the bank's market leading professional Cash Management Consulting Group, initiated the bank's non-credit service product management organization and profit center profitability programs and was instrumental in the breakthrough EDI/EFT payment system implemented by General Motors. Dr. Ferguson is a graduate of Notre Dame, Stanford and Purdue Universities, a special edition editor of the Journal of Engineering Entrepreneurship and a member of Tau Beta Pi.

**Dr. Catherine E. Brawner, Research Triangle Educational Consultants**

Catherine E. Brawner is President of Research Triangle Educational Consultants. She received her Ph.D.in Educational Research and Policy Analysis from NC State University in 1996. She also has an MBA from Indiana University (Bloomington) and a bachelor's degree from Duke University. She specializes in evaluation and research in engineering education, computer science education, teacher education, and technology education. Dr. Brawner is a founding member and former treasurer of Research Triangle Park Evaluators, an American Evaluation Association affiliate organization and is a member of the American Educational Research Association and American Evaluation Association, in addition to ASEE. Dr. Brawner is also an Extension Services Consultant for the National Center for Women in Information Technology (NCWIT) and, in that role, advises computer science and engineering departments on diversifying their undergraduate student population. She remains an active researcher, including studying academic policies, gender and ethnicity issues, transfers, and matriculation models with MIDFIELD as well as student veterans in engineering. Her evaluation work includes evaluating teamwork models, statewide pre-college math initiatives, teacher and faculty professional development programs, and S-STEM programs.

Optimizing Student Team Skill Development using Evidence-Based Strategies: Year 3

NSF Award 1431694

The broad goal of this work is to study the effectiveness of teamwork training methods, experience in teams, and receiving various forms of feedback on the development of team skills and the ability to evaluate teamwork. This is conducted through a series of studies including classroom experiments, lab studies, and analyses of historical data. The research leverages the NSF's prior investment in the Comprehensive Assessment of Team-Member Effectiveness (CATME) system to measure teamwork.[1] The CATME system automates some of the data collection and feedback, providing input to some of the seven empirical studies required to explore these research questions. The entire research protocol is shown in Figure 1. The two outcomes measured in this research are team-member effectiveness and the ability to evaluate the effectiveness of peers.

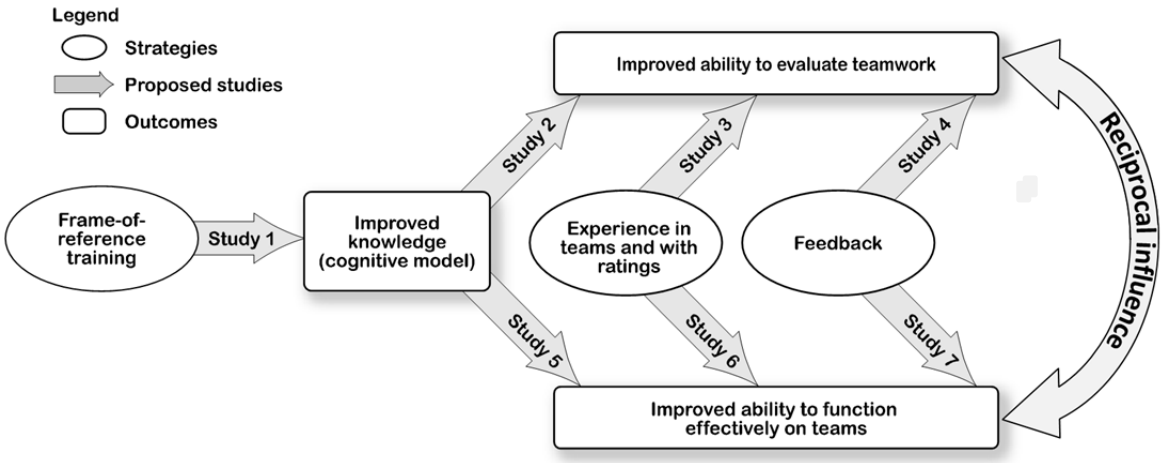


**Figure 1.** Model for improving self- and peer-evaluation skills and teaming skills.

**Interventions to improve rating accuracy**

In laboratory studies, Frame-of-reference training (FORT) is among the most effective methods of improving rating accuracy. FORT is an approach to training in which evaluators learn to rate peers more accurately through practice and feedback in an authentic context, and is described in detail below. Noting that many faculty will not conduct FORT, other approaches to improve rating accuracy were also evaluated in a multi-section course. Since students had no control over which intervention was used in their section, and each section has approximately 120 students, this study approximates a randomized control trial to examine the effect of the various interventions. The effect of team experience and mode of feedback are explored in other studies.

**Frame-of-reference training**

Frame-of-reference training is well-established and empirically supported,[2,3] and is expected to align students' cognitive model of teamwork with the CATME model by teaching students the ways team members can contribute effectively to teams in CATME's five key areas:[1]

- Contributing to the team's work
- Interacting with teammates
- Keeping the team on track
- Expecting quality
- Having relevant knowledge, skills, and abilities

Whereas frame-of-reference training (FORT) is accepted as the most promising approach to rater training and improving rater accuracy, most studies on the effectiveness of FORT were conducted in a laboratory setting,[4] including a laboratory study of FORT developed in this work.[5] Based on the work of Bernardin et al.[6] and Bernardin & Buckley,[7] we have adapted FORT principles for use outside the laboratory, and our approach to FORT in the use of CATME includes five components:

1. ***Define performance.*** The ratings instrument defines what counts in the performance domain. We use the CATME instrument, in which behavioral anchors define performance in five categories of team-member performance derived from research.
2. ***Teach raters the scale.*** Raters will receive training on the rating scale and how it is used.
3. ***Discuss critical incidents.*** To provide a deeper understanding of the rating system, the training will identify critical positive and negative team behaviors using vignettes and role playing.
4. ***Practice rating.*** Raters will practice rating using video or text-based vignettes (rater calibration).
5. ***Give raters feedback.*** Raters receive feedback on the accuracy of their vignette ratings.

**Other training approaches**

Since various training approaches have been used in the classroom, our goal was to compare our FORT design to various other approaches. The following other training materials were included in an experimental design:

- ***Video-based modeling.*** We expect that these video-based materials are more engaging, because they demonstrate teamwork behaviors using scenes from popular movies. Nevertheless, "engaging" does not necessarily mean "more effective." The use of these videos is acceptable under "fair use" copyright guidelines as long as these videos are shown live in class and cannot be downloaded by students.

- ***In-class teamwork reflection.*** This is an exercise to serve as a default comparison case the frame-of-reference and video-based training. It presents some brief scenarios that can occur in teams for teams to discuss how to manage those situations.
- ***Rater practice*** is a feature of CATME that presents a written scenario describing a collection of behaviors of three students. The description appears as a vignette with behaviors presented in no particular order, and students identify which behaviors in the paragraph are related to each dimension and rate each student on all dimensions. The descriptions are available as mouse-over text when completing the CATME instrument. This activity is done out of class, takes little time, and has no control exercise.
- ***Rater accuracy*** training is a single PowerPoint slide that presents three problematic rating patterns – (1) giving all teammates the same rating on all dimensions ("everyone is perfect in every way."), (2) giving different ratings for different dimensions, but the same rating for all teammates ("we're all great at interacting with teammates, and we're all contributing about the same."), and (3) giving different ratings to different teammates without differentiating the different dimensions ("This teammate is terrible at everything, but this other teammate is great at everything."). In this exercise, students reflect on the causes and consequences of those rating patterns.
- ***Peer evaluation feedback.*** This is another short set of slides to remind students of how to interpret feedback from CATME, including sample feedback and its interpretation. This constitutes a discussion of feedback without promoting more accurate rating.

Using the materials described above, the research design included 12 different protocols listed below that were administered in a first-year engineering class at a large Midwestern university. The class was taught in 15 sections, so instructors had some flexibility in choosing their part of the protocol. Instructors teaching multiple sections had identical in-class training protocols, but may have had different out-of-class assignments.

1. Frame-of-reference training + peer evaluation feedback exercise
2. Frame-of-reference training + peer evaluation feedback exercise + rater practice
3. Frame-of-reference training + rater accuracy training
4. Frame-of-reference training + rater accuracy training + rater practice
5. In-class teamwork reflection + peer evaluation feedback exercise
6. In-class teamwork reflection + peer evaluation feedback exercise + rater practice
7. In-class teamwork reflection + rater accuracy training
8. In-class teamwork reflection + rater accuracy training + rater practice
9. Video-based training + peer evaluation feedback exercise
10. Video-based training + peer evaluation feedback exercise + rater practice
11. Video-based training + rater accuracy training
12. Video-based training + rater accuracy training + rater practice

The same students (with less than 10% attrition) enrolled in a follow-on class in the Spring. The training intervention received in the Fall was tracked, but not used to form teams in the Spring, resulting in teams composed of students who received various training protocols in the Fall. This will make it possible to measure the effect of various training and feedback interventions on student's ability to rate their teammates, perform in teams, and effects on a variety of other outcomes including conflict, cohesion, and satisfaction. These follow-on data have not been fully analyzed yet due to the loss of a research team member.

**Measures of improvement resulting from interventions**

There are two main challenges in assessing the quality of peer evaluations: (1) when used in real teams, there is generally no true score available because the team was not observed by an external expert evaluator and (2) the nested nature of peer evaluation data requires the use of statistical models that acknowledge that structure. In our work, the primary measures of peer evaluation quality is obtained using the Social Relations Model (SRM).[8] SRM is a special case of generalizability theory[9] with the model consisting of a two-way, random-effects ANOVA with actor and partner factors, and relationship as the interaction term. In our case, the actor is the target of a peer rating, and the target effect is the amount of variance in peer ratings accounted for by the target. A high target effect is desirable. The other primary variance component is the rater effect, which represents idiosyncratic rating. A low rater effect is therefore desired. The interaction term is the various accounted for by dyadic relationships. While this has by some been considered another source of error (such as collusion), it has been noted that students within a team will vary in the behaviors they exhibit to different team members, so dyadic variance includes a component that represents target behaviors.

**The growth of the CATME user base and its importance to this work**

Faculty at a growing list of institutions is being recruited for intervention studies. Some of these interventions help replicate the findings of the research team's work in other settings, and is being conducted with advice from, but analyzed independently of the research team. In other cases, our team is assisting with analysis. Still other studies are purely quantitative, and are aided by the growing dataset of de-identified data, particularly for studies that require the use of certain follow-on questions, such as team satisfaction and conflict. These studies at other institutions make our findings more robust and extend the capabilities of the research team.

The growth of the CATME user base continues to be strong. From October 1, 2016, to September 30, 2017, another 240,797 unique students have used the CATME system for the first time. In the same year, we reached faculty at 285 new institutions, adding 3,293 new faculty accounts. By mid-November, CATME had served over 1,000,000 unique students, primarily in the United States, and as of publication, 1,090,522 students have used CATME.

Figure 1 shows the trajectory of the growth of CATME's user base in terms of students, institutions, instructors, and countries. This figure is updated monthly on our website.[10]

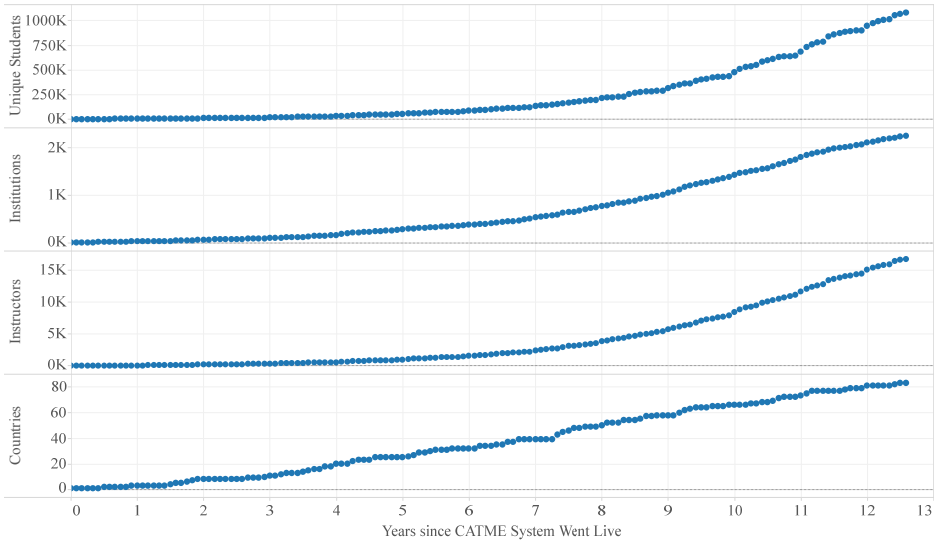


**Figure 1.** Growth of the CATME user community in students, institutions, instructors, and countries

Such a large data resource allows us to restrict our studies to include high-quality data that controls certain confounding issues. We can constrain our study to include only teams of four students at U.S. institutions with complete data. Further, those results can be compared to teams of other sizes to explore how team size affects team dynamics. CATME's user base is also geographically distributed, which helps to ensure that our results are generalizable. Figure 2 shows the institutions in the continental United States where CATME has been used.

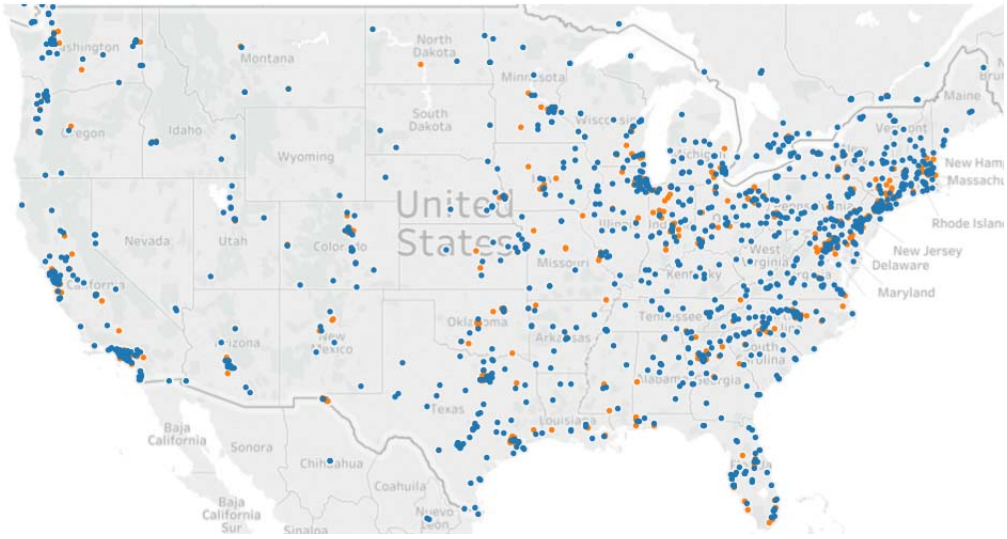


**Figure 2.** Distribution of institutions in the continental U.S. where instructors have had CATME accounts. Pre-college institutions are represented by orange dots.

**The addition of a new feedback mode**

A newly implemented peer-to-peer comments feature added a new mode of feedback and improved accountability to the system. We are studying the extent to which these qualitative data are related to student ratings using CATME's behaviorally anchored rating scale. This work is being published separately. With the addition of peer-to-peer comments to the CATME system, a study is underway to explore the quality of feedback and its consistency with the ratings provided by students.

**References**


[1] Ohland, M.W., Loughry, M.L., Woehr, D.J., Bullard, L.G., Felder, R.M., Finelli, C.J., Layton, R.A., Pomeranz, H.R. & Schmucker, D.G. (2012). The Comprehensive Assessment of Team Member Effectiveness: Development of a behaviorally anchored rating scale for self and peer evaluation. *Academy of Management: Learning & Education*, **11**(4) 609-630.

[2] Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-206.

[3] Woehr, D. J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology, 79(4),* 525-534.

[4] Hersen, M. (2004). Comprehensive Handbook of Psychological Assessment. John Wiley and Sons, NY, p. 380.

[5] Loignon, A.C., Woehr, D.J., Thomas, J.S., Loughry, M.L., Ohland, M.W., & Ferguson, D.M. (2017). Facilitating peer evaluation in team contexts: The impact of frame-of-reference rater training. *Academy of Management Learning & Education*. **16**(4), 562-578.

[6] Bernardin, H.J., Buckley, M.R., Tyler, C.L. and Wiese, D.S. (2000). A reconsideration of strategies in rater training. Research in Personnel and Human Resources Management, 18, 221–74.

[7] Bernardin, H.J., and Buckley, M.R. (1981). Strategies in rater training. Academy of Management Review, 6, 205-212.

[8] Kenny, D. A. (1994). Interpersonal perception: A social relations model. NY: Guilford Press.

[9] Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability of scores and profiles. New York: John Wiley.

[10] Ohland, M.W. (2018). Our user base – CATME. http://info.catme.org/about/our-user-base/