2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #32931

Longitudinal Effects of Team-Based Training on Students' Peer Rating Quality

Mr. Siqing Wei, Purdue University at West Lafayette

Siqing Wei received BSEE and MSEE from Purdue University. He is currently pursuing a Ph.D. degree in Engineering Education program at Purdue University. After years of experience of serving the peer teacher and a graduate teaching assistant in first-year-engineering courses, he is now a research assistant at CATME research group studying how cultural diversity impacts teamwork and how to help students improve intercultural competency and teamwork competency by interventions, counseling, pedagogy, and tool selection (such as how to use CATME Team-Maker to form inclusive and diversified teams). In addition, he also works on many research-to-practice projects to enhance educational technology usage in engineering classrooms and educational research. One feature ongoing project utilizes natural language processing technique to map students' written peer-to-peer comments with their perceived numerical ratings. Siqing also works as the technical development and support manager at CATME research group.

Mr. Chuhan Zhou
Dr. Matthew W. Ohland, Purdue University at West Lafayette

Matthew W. Ohland is Associate Head and the Dale and Suzi Gallagher of 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 for the best paper published in the Journal of Engineering Education in 2008, 2011, and 2019 and from the IEEE Transactions on Education in 2011 and 2015. Dr. Ohland is 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.

Longitudinal Effects of Team-Based Training on Students' Peer Rating Quality

Abstract

This research paper explores how longitudinal team-based training influences engineering students' peer rating quality. Engineering students will be expected to work in teams, and the ability to effectively cooperate and communicate is increasingly recognized by technical corporations. Teamwork is an important outcome included in the accreditation criteria of ABET and is also progressively integrated into engineering courses to a varying extent in engineering programs. However, the cumulative effect of the length of students' exposure to team-based projects, instruction in effective teamwork, and practice in peer evaluation is not well studied. Given the ad-hoc experience and previous literature, which is that student teams need time to overcome barriers and conflicts to perform well, we propose to study peer evaluation quality in a two-course sequence of team-based engineering classes in a large Midwestern public university. We hypothesize that peer evaluation behaviors, including rating scores and quality, for the second team-based course will be better on average compared to the peer evaluation behaviors in the first team-based course. Longitudinal use of a peer evaluation system would be expected to result in more accurate and consistent peer rating and students would get higher peer rating scores. Data from the two consecutive engineering foundation courses were analyzed using ANOVA and the Social Relations Model. Results showed no significant difference between peer rating scores in two consecutive mandatory courses. Rather, peer rating behaviors, or the patterns of peer ratings, restarted in the second course, which would suggest that students need to go through the same process of being better raters each time when they are put into new teams.

This work informs university administrators and instructors that it takes time for student teams as they design curricula that help engineering students improve teaming skills. In particular, if this result generalizes to other course sequences and institutions with the learning objectives related to teamwork competency development, it might suggest that there is a benefit to reforming teams mid-semester in each course to give students additional experiences to apply what they have learned to more teams.

1. INTRODUCTION

Research on improving teamwork skills has been conducted in various disciplines from different perspectives. Providing teamwork training for trauma resuscitation staff could improve the clinical care of trauma patients [1]. Research in plant construction shows that adopting training policies and providing motivators to the company's staff help them achieve better performance and to improve teamwork experience [2]. Organizational competitiveness could be enhanced by using a framework integrated and emphasized on teamwork [3]. In engineering design education, analyzing various pedagogical approaches to combine teamwork experience with reflective activities indicates that engineering students can make a connection between effective teamwork and key engineering design abilities such as open-mindedness, innovation, and communication [4]. In a similar research setting, study suggests that first-year engineering students gradually become more effective team members during a semester and compared to reflections, their teamwork behaviors are the better predictor of their academic performance [5].

Teamwork is integrated into teaching to a varying extent in engineering schools. Many universities have developed engineering foundation courses to prepare novice engineers with fundamental knowledge and skills, including but not limited to, the engineering design process, modeling, and teamwork to meet the ABET requirement of effective communication and teamwork [6] and better prepare engineering students with future professional positions.

Several teamwork frameworks and models have been proposed to provide training or assessments to the individual team members or the team as a whole [7]. Tuckman created the well-known team developmental model to separate the stages of a small team's lifespan by the team's experience at each stage: forming, storming, norming, and performing [8]. The framework might provide expectations and limited insights for typical team dynamics, yet knowledge of this framework has not been shown to improve teamwork. Stevens and Campion developed the Knowledge, Skills, and Abilities (KSA) Teamwork Test, which consists of two categories: interpersonal KSAs (conflict resolution and communication) and Self-Management KSAs (goal setting and task coordination) [9]. This self-assessment instrument focuses on team-based situations and behaviors instead of personality characteristics so that it could only be considered as a reflection of a team member's perception of teamwork skills rather than their behaviors in a team setting. Issues also arise from individuals under- or over-estimating their own capabilities. Numerous problems often show up in team dynamics, such as the preference of work alone, communication barriers, conflicts and cliques, differences in team members' skills, goal levels and motivation, as well as problems like free-riding or social loafing [10].

Designed to help instructors manage student teams, CATME (the Comprehensive Assessment of Team-Member Effectiveness) is a behaviorally anchored rating scale for self and peer evaluation [10]. The five CATME behavioral dimensions are: Contributing to the team's work; Interacting with teammates; Keeping the team on track; Expecting quality; Having relevant knowledge, skills, and abilities (KSAs). There are several benefits from the schema of self and peer evaluations embedded in CATME system: (1) it is one way that instructors can manage teamwork to create better teamwork experiences for their students [11], [12]; (2) it calibrates and improves students' peer evaluation behaviors with CATME teamwork dimensions [10]; and (3) it prepares students to provide constructive feedback to team members, which is a common practice in workplaces [13].

Typical of other methods of peer evaluation, some problems also emerge in the CATME system: (1) self-appraisals usually coexist with leniency errors [14]; (2) people with poor skills can be unable to recognize their skill or performance deficiencies so that they have difficulty providing accurate ratings of themselves and other teammates [15], [16]; (3) Concerns about damaging social relations and the tendency of using social comparison framework greatly influence the peer-to-peer rating results [17]. With many years of development, the CATME system includes frame-of-reference training and a peer rating practice exercise that have both been shown to improve raters' ability to rate accurately [18], [19]. The system also offers the ability to collect peer-to-peer comments that provide additional information to students and instructors.

Using CATME to track peer evaluation behaviors, we are equipped to investigate various topics related to teamwork. Students with different teaming experiences may have a totally different understanding of the process of teamwork [20], which suggests that team members with different levels of experience in a team act differently and judge teamwork behaviors differently. Repeated use of a peer evaluation system has been shown to improve the development of team skills [21]. This

work explores whether that same longitudinal exposure results in more accurate and consistent peer rating performance. Therefore, this paper examines the difference between peer evaluation results based on students' teamwork training and repeated use of peer evaluation of teamwork in two consecutive engineering foundation courses.

2. RESEARCH METHODOLOGY

2.1 Research Background & Data Collection

The research data were collected using CATME [10] from consecutive first-year engineering courses (here called ENGR I and ENGR II) in Fall 2018 and Spring 2019, respectively, at a large university in the Midwestern United States. Students were assigned to teams of three or four (mostly four) members by CATME's Team-Maker feature [22]. Each course required students to complete rater practice training [19] and engage in four rounds of self and peer evaluation. We consider the data were collected in high quality because (1) the average response rate for each round of survey is consistently over 85%; (2) participation in the survey, including providing constructive comments, is counted as part of the student grades: students get up to 2% total grade for each survey completion and meeting team skill expectations during that evaluation period (as determined from peer evaluation results and confirmed by observations of the instructional team); (3) course activities include a substantial amount of team and pair-within-team activities, including a team exam. Survey results are released to students so that they can learn from peer feedback to improve their teamwork behaviors. We treat each round survey as one intervention so that there are eight in total. The first four interventions represent the four rounds of surveys conducted in ENGR I, and the last four interventions (5-8) indicate the four surveys in ENGR II, respectively. We are primarily interested in comparing the developmental trend for each course and the difference in corresponding rounds of peer evaluation results.

2.2 Data Cleaning

Table 1: Clean data summary of teams, students, and gender distribution.

Intervention #	# of teams	# of students	# of Male	# of Female	# of "Other or prefer not to answer"
1	197	788	598	175	12
2	179	716	560	145	9
3	171	684	528	141	14
4	199	796	605	174	13
5	238	952	734	204	10
6	232	928	694	219	11
7	214	856	659	180	13
8	272	1088	838	229	15

We set two rules for data cleaning process:

(1) We only take the data from teams whose members all take both ENGR I and ENGR II to make the results comparable (to avoid a shift in sample population over time);

(2) We delete all the data for a team in a particular round of peer evaluation if fewer than three team members completed the survey that round (fewer data would hinder our ability to measure peer

evaluation quality as described below). This result in a minor fluctuation in the sample over time as shown below.

The demographic description of the cleaned data used for analysis is shown in Table 1. The response rate of teams with complete data is somewhat higher in ENGR II.

2.3 Statistical Models

The dispersion pattern of across CATME dimensions and individual dimension peer rating data is analyzed using ANOVA with a Bonferroni correction to recognize whether differences among the intervention groups are significant; the convergence of self-ratings and peer-rating is tested by Least Square Means [23]–[25].

To investigate the quality of peer ratings, the Social Relations Model (SRM), a conceptual and analytic statistical method, is used to partition the rating variance into rater variance, target variance, relationship variance, team variance, and error variance [18], [25], [26]. Of those, target variance is the most desirable measurement because it describes how much of the variance is related to the team member being rated; a larger target variance indicates more consistent ratings when all of the team members rate a particular team member [18], [25], [26].

3. DATA ANALYSIS

From both Tables 2 and 3 below, we cannot find any difference in the patterns of dispersion or convergence in across all dimensions' peer rating data. Both average score and standard deviation of across dimension rating results are not statistically different from adjacent survey results and from the corresponding round of surveys conducted in both ENGR I and II courses. However, the mean difference and dispersion difference between self and peer evaluation are significant for all interventions. The mean difference indicates that self-rating scores are typically 0.5 lower than peer rating scores, a large effect size. The negative dispersion difference, along with the large effect size, implies that self-ratings are distributed more centrally compared to peer ratings—students in these courses tend not to rate themselves at scale extremes.

Table 2: Dispersion Across Dimensions ANOVA Analysis

Intervention	Intervention	Difference Mean Rating	P-Value	Effect Size (Cohen's d)	Difference Dispersion	P-Value	Effect size (Cohen's d)
1	2	0.02798	1.0000	0.0069	-0.01898	1.0000	0.0030
2	3	0.02807	1.0000	0.0092	-0.03906	1.0000	0.0047
3	4	0.03796	1.0000	0.0223	-0.02517	1.0000	0.0006
5	6	0.02388	1.0000	0.0004	-0.01063	1.0000	0.0004
6	7	0.05866	1.0000	0.0086	-0.01375	1.0000	0.0079
7	8	0.01808	1.0000	0.0006	-0.02863	1.0000	0.0052
1	5	-0.06513	0.9033	0.0421	-0.00371	1.0000	0.0241
2	6	-0.06922	0.8135	0.0349	0.004641	1.0000	0.0276
3	7	-0.03863	1.0000	0.0171	0.02995	1.0000	0.0151
4	8	-0.05851	1.0000	0.0402	0.02649	1.0000	0.0208

Table 3: Self-rating and Peer-rating Across Dimension (Convergence) ANOVA Analysis

Intervention	Difference Mean (Self-Peer)	P-value	Effect Size (Cohen's d)	Difference Dispersion (Self-Peer)	P-value	Effect Size (Cohen' s d)
1	-0.5043	<0.0001	0.4761	-0.4978	<0.0001	1.1806
2	-0.5110	<0.0001	0.4759	-0.4921	<0.0001	1.1963
3	-0.5170	<0.0001	0.4690	-0.5028	<0.0001	1.1936
4	-0.5188	<0.0001	0.4762	-0.5008	<0.0001	1.1837
5	-0.4891	<0.0001	0.4765	-0.4970	<0.0001	1.1664
6	-0.4769	<0.0001	0.4717	-0.4914	<0.0001	1.1681
7	-0.5072	<0.0001	0.4725	-0.5049	<0.0001	1.1760
8	-0.4915	<0.0001	0.4748	-0.4987	<0.0001	1.1739

Table 4: Mean Individual Dimensions ANOVA Divergence Analysis

	Intervention	Intervention	Mean Difference	p-value	Effect Size (Cohen's d)
Contributing to the team's work [C]	1	5	-0.06769	0.0984	0.0390
	2	6	-0.05673	0.3619	0.0453
	3	7	-0.03789	1.0000	0.0198
	4	8	-0.05326	0.3597	0.0319
Expecting quality [E]	1	5	-0.04015	1.0000	0.0267
	2	6	-0.04108	1.0000	0.0425
	3	7	0.04321	1.0000	0.0230
	4	8	-0.08149	0.0462	0.0182
Having relevant KSA [H]	1	5	-0.11760	<.0001	0.0323
	2	6	-0.12770	<.0001	0.0297
	3	7	-0.11280	0.0005	0.0152
	4	8	-0.12710	<.0001	0.0255
Interacting with teammates [I]	1	5	-0.1255	0.0001	0.0384
	2	6	-0.08304	0.0493	0.0462
	3	7	-0.02846	1.0000	0.0255
	4	8	-0.06051	0.2885	0.0323
Keeping the team on track [K]	1	5	-0.08995	0.0248	0.0489
	2	6	-0.05790	0.5928	0.0509
	3	7	-0.04202	1.0000	0.0262
	4	8	-0.06639	0.2111	0.0437

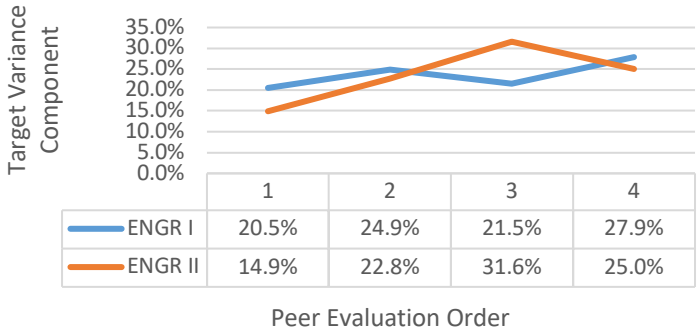

Figure 1. Across-CATME-Dimension Social Relations Model (SRM) Target Variance Analysis

Fig. 1 shows that target variance in ENGR II only exceeds that of ENGR I in the third round of each class. Overall, there is no significant difference between the developmental trend observed in each class. So while in each class, students improve the quality of their ratings, this improvement does not carry over to the next class. Similarly, as shown in Fig. 2, when we analyze target variance by the individual CATME dimensions, we see that the trend in both courses starts with low target variance and improves with each peer evaluation. Not only do students need practice to become better and more consistent raters, but they need practice with each new team. Except for dimension H (having relevant KSAs), there is no significant difference in the trajectory of improvement of target variance between the two courses. Based on our findings, it appears that each time students are assigned to a new team, they have to go through a new process of becoming more accurate raters.

As shown in Fig. 2., target variance of ENGR II's peer rating scores for Having relevant KSAs is higher than that of ENGR I in each round of peer evaluation—the outcome that was hoped for in all dimensions. We propose two possible interpretations of the result: (1) students develop a better understanding on this dimension's scale so that they rate each other more consistently and accurately, even in a new team; (2) the curriculum and learning objectives of ENGR II demands KSAs that are more easily assessed by the students.

However, from previous research findings by our group, we know that the differences are often masked at all-dimension levels so that we need to perform individual dimension analysis to explore the discrepancy [25]. In this case, by performing individual dimension analysis as shown in Table 4, we see various mean differences between ENGR I and ENGR II, with the overall tendency that ENGR II ratings are lower than those in ENGR I, but even those that are significant have a very small effect size.

4. CONCLUSION AND DISCUSSION

By analyzing and comparing the peer evaluation results in ENGR I and II courses, we find no statistically significant differences in the developmental patterns across all-dimension dispersion and convergence ANOVA and SRM analysis.

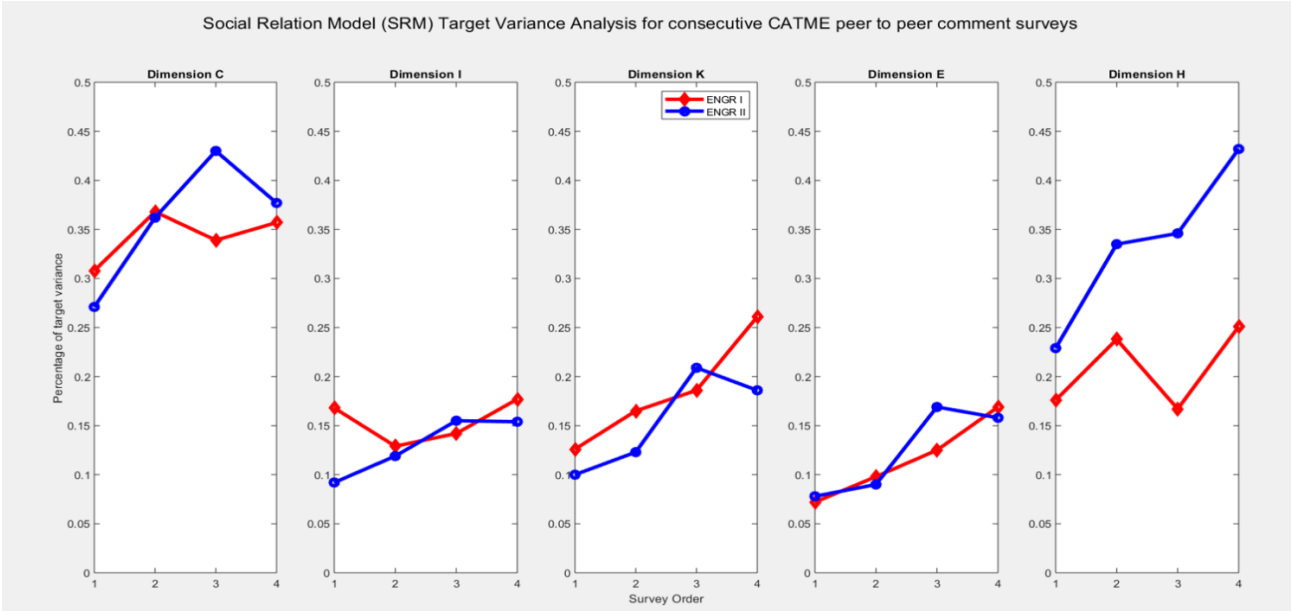

Figure 2. Percentage of target variance for each CATME dimension in ENGR I and ENGR II at four sampling points in the semester

By examining peer rating results from the two courses in two consecutive semesters on the individual dimensions longitudinally, we conclude that student rating quality generally improves in each course, but the trajectory of the improvement of rating quality shows no significant difference between the two courses. Students need time to adjust themselves to become better raters, and separately in each of these team experiences. Several plausible reasons could explain this phenomenon: (1) students need to establish norms for how their teammates will give each other feedback; (2) even if students are well-trained as raters, there is still a tendency toward leniency while getting to know a new set of teammates and how they work; (3) as with much learning, there is some loss with time, particularly a semester transition; (4) students may be uncertain that their new instructor will have the same expectations as their instructor from the previous semester; therefore, they need to spend some time to adjust new circumstances. Further research is needed to explain these observations.

5. LIMITATION

The proposed courses are taught in more than 15 sections per semester by various instructors so that the influence by instructors might influence the peer rating results. We perform the study on only a single cohort of students and through only two consecutive semesters. A longitudinal study of teamwork skill development throughout students' whole undergraduate experience might show different results, particularly if the same peer evaluation instrument were used consistently throughout the curriculum.

References

[1] J. Capella *et al.*, "Teamwork training improves the clinical care of trauma patients," *J. Surg. Educ.*, vol. 67, no. 6, pp. 439–443, 2010, doi: 10.1016/j.jsurg.2010.06.006.

[2] A. A. Tabassi, M. Ramli, and A. H. A. Bakar, "Effects of training and motivation practices on teamwork improvement and task efficiency: The case of construction firms," *Int. J. Proj. Manag.*, vol. 30, no. 2, pp. 213–224, 2012, doi: 10.1016/j.ijproman.2011.05.009.

[3] K. M. Telleria, D. Little, and J. Macbryde, "Managing processes through teamwork," *Bus. Process Manag. J.*, vol. 8, no. 4, pp. 338–350, 2002, doi: 10.1108/14637150210434991.

[4] P. L. Hirsch and A. F. McKenna, "Using reflection to promote teamwork understanding in engineering design education," *Int. J. Eng. Educ.*, vol. 24, no. 2, pp. 377–385, 2008.

[5] S. Anwar and M. Menekse, "Unique contributions of individual reflections and teamwork on engineering students' academic performance and achievement goals," *Int. J. Eng. Educ.*, vol. 36, no. 3, pp. 1018–1033, 2020.

[6] H. J. Passow, "Which ABET competencies do engineering graduates find most important in their work?," *J. Eng. Educ.*, vol. 101, no. 1, pp. 95–118, 2012, doi: 10.1002/j.2168-9830.2012.tb00043.x.

[7] E. Salas, N. J. Cooke, and M. A. Rosen, "On teams, teamwork, and team performance: Discoveries and developments," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 3, pp. 540–547, 2008, doi: 10.1518/001872008X288457.

[8] B. W. Tuckman, "Developmental sequence in small groups," *Psychol. Bull.*, vol. 63, no. 6, pp. 384–399, 1965, doi: 10.1037/h0022100.

[9] M. J. Stevens and M. A. Campion, "The knowledge, skill, and ability requirements for teamwork: Implications for human resource management," *J. Manage.*, vol. 20, no. 2, pp. 503–530, 1994, doi: 10.1177/014920639402000210.

[10] M. Ohland *et al.*, "The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation," *Acad. Manag. Learn. Educ.*, vol. 11, no. 4, p. 609, 2012, doi: 10.5465/amle.2010.0177.

[11] L. E. Gueldenzoph and G. L. May, "Collaborative peer evaluation: best practices for group member assessments," *Bus. Commun. Q.*, vol. 65, no. 1, pp. 9–20, 2002, doi: 10.1177/108056990206500102.

[12] R. S. Hansen, "Benefits and problems with student teams: suggestions for improving team projects," *J. Educ. Bus.*, vol. 82, no. 1, pp. 11–19, 2006, doi: 10.3200/JOEB.82.1.11-19.

[13] V. U. Druskat and S. B. Wolff, "Effects and timing of developmental peer Appraisals in self-Managing work groups," *J. Appl. Psychol.*, vol. 84, no. 1, pp. 58–74, 1999, doi: 10.1037/0021-9010.84.1.58.

[14] E. Inderrieden, R. Allen, and T. Keaveny, "Managerial discretion in the use of self-ratings in an appraisal system: The antecedents and consequences," *J. Manag. Issues*, vol. 16, no. 4, pp. 460–482, 2004.

[15] J. Kruger and D. Dunning, "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments," *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1121–1134, 1999.

[16] A. Jassawalla, H. Sashittal, and A. Malshe, "Students' perceptions of social loafing: Its antecedents and consequences in undergraduate business classroom teams," *Acad. Manag. Learn. Educ.*, vol. 8, no. 1, pp. 42–54, 2009, doi: 10.5465/AMLE.2009.37012178.

[17] R. Saavedra and S. K. Kwun, "Peer evaluation in self-managing work groups," *J. Appl.*

Psychol., vol. 78, no. 3, pp. 450–462, 1993, doi: 10.1037/0021-9010.78.3.450.
[18] A. Loignon, D. Woehr, J. Thomas, M. Loughry, M. Ohland, and D. Ferguson, "Facilitating peer evaluation in team contexts: The impact of frame-of- reference rater training," *Acad. Manag. Learn. Educ.*, vol. 16, no. 4, p. 562, 2017, doi: 10.5465/amle.2016.0163.
[19] D. M. Ferguson, E. Shu, Y. Cao, and M. Ohland, "Examining the effect of a game-like practice tool on the quality of student peer evaluations," in *Frontiers in Education Conference*, 2018, doi: 10.1109/FIE.2018.8659270.
[20] J. R. Rentsch, T. S. Heffner, and L. T. Duffy, "What you know is what you get from experience: Team experience related to teamwork schemas," *Gr. Organ. Manag.*, vol. 19, no. 4, pp. 450–474, 1994, doi: 10.1177/1059601194194004.
[21] M. B. . Donia, T. A. O'Neill, and S. Brutus, "The longitudinal effects of peer feedback in the development and transfer of student teamwork skills," *Learn. Individ. Differ.*, vol. 61, pp. 87–98, 2018, doi: 10.1016/j.lindif.2017.11.012.
[22] R. A. Layton, M. L. Loughry, M. W. Ohland, and G. D. Ricco, "Design and validation of a web-based system for assigning members to teams using instructor-specified criteria," *Adv. Eng. Educ.*, vol. 2, no. 1, pp. 1–28, 2010.
[23] T. L. Poling, D. J. Woehr, C. A. Gorman, and L. M. Arciniega, "The impact of personality and value diversity on team performance," in *Annual Meeting for the Society for Industrial and Organizational Psychology*, 2014.
[24] D. M. Ferguson, M. W. Ohland, C. Lally, H. I. Somnooma, and Y. Cao, "Evaluating the effect of different teamwork training interventions on the quality of peer evaluations," in *Frontiers in Education Conference*, 2018, doi: 10.1109/FIE.2018.8658782.
[25] S. Wei, D. Ferguson, M. Ohland, and B. Beigpourian, "Examining the cultural influence on peer ratings of teammates between international and domestic students," in *American Society for Engineering Education Annual Conference & Exposition*, 2019.
[26] M. D. Back and D. A. Kenny, "The social relations model: How to understand dyadic processes," *Soc. Personal. Psychol. Compass*, vol. 4, no. 10, pp. 855–870, 2010, doi: 10.1111/j.1751-9004.2010.00303.x.