125 YEARS AT THE HEART OF ENGINEERING EDUCATION
ASEE
2018 ASEE ANNUAL CONFERENCE & EXPOSITION
JUNE 24–27, 2018
SALT PALACE CONVENTION CENTER
SALT LAKE CITY, UT

Paper ID #23727

Comparing Peer Evaluations of Teamwork Behavior by K-12 Students versus First-year Engineering Students

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

Daniel M. Ferguson is CATME Managing Director 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. Matthew W. Ohland, Purdue University-Main Campus, 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.

Yuchen Cao, Department of Statistics, Purdue University

Comparing Peer evaluations of Teamwork Behavior by K12 students vs. First year Engineering students

Abstract

The goal of this work-in-process research is to determine the extent to which secondary students in team based courses behave similarly when rating their peers compared to First-Year Engineering (FYE) students. In particular, we are interested in the quality of peer evaluations based on the similarity of the variability or lack thereof in the comparison of peer evaluation ratings.

A person's ability to work effectively in a team or group setting is vital to a college career as well as in a work-life profession and therefore it is often a significant factor in a corporate or government hiring process. Recognizing this need, a number of U.S. undergraduate collegiate STEM programs as well as many K12 instructors, particularly teachers delivering Project Lead The Way (PLTW) courses, use team activity or project-based courses in their curricula. Thousands of undergraduate instructors and K12 teachers also form and manage student teams using online tools including the Team-Maker and the Comprehensive Assessment of Team Member Effectiveness (CATME) tools. CATME contains both scientifically based team formation heuristic tools and a behaviorally anchored peer evaluation instrument that has also been scientifically developed and validated.

In spite of the K12 emphasis on the use of teams, the analysis of teamwork behavior and the assessment of effective use of teams in K12 has not seen the same focus as at the collegiate level. When tools such as CATME's, which was developed for collegiate curriculum, are used, our key question is whether K12 students behave similarly to FYE college students when doing peer evaluations, thereby making the use of assessment tools such as CATME appropriate in K12 contexts.

Introduction

Teamwork and the correct team behavior are key attributes sought after by a large number of companies when hiring new employees [1, 2]. Working in teams not only helps distribute the workload better but leads to greater efficiency, better communication in the future as well as creates a supportive environment for workers that can serve as a platform for even better performance. Hence, teamwork skills training has become more prevalent throughout college programs and in businesses [3]. In fact accreditation bodies in Business, Engineering and Healthcare have mandated teamwork training [4]. Teamwork is defined as "a cooperative or coordinated effort on the part of a group of persons acting together" [5]. Chen argues that many students entering the workplace lack key teamwork skills that hamper their abilities to excel in their job field [6]. Teamwork is not only deemed important in higher education, but has also began to attract the widespread attention of a multitude K-12 instructors in recent times, especially those teaching Project Lead The Way (PLTW) courses. Whereas cooperative learning and other team-based pedagogies have long been commonplace in elementary education, instructors throughout US secondary education have more recently increased the use of small teams in their courses even as such approaches are gaining ground in college classrooms. Wang and MacCann and other researchers have used self-report, peer-report and situational judgment test to assess high-school students' teamwork skills, the psychometrics of the instruments have not been published, so it is unclear the extent to which their development was scientific, whereas the development and validation of CATME have been published [7, 8]. Some measure teamwork as a single-dimension attribute, whereas CATME captures behavior in five dimensions of teamwork. The development of a systematic approach of assessing the quality of teamwork is critical for secondary classrooms just like in higher education. The evidence in this paper suggests that adopting CATME as an assessment tool for use in evaluating the effective use of teams in secondary classes promises a lot of potential, given that there is significant similarity in the ratings and the quality of ratings between secondary and FYE students,

Our primary research goal was to determine how similarly secondary students and FYE students rated themselves and their peers when using CATME peer evaluations. A second research question was: is there a significant difference in the quality of peer ratings across dimensions between these two groups of students? The quality of a rating, in this paper, is defined as having a larger dispersion (measured by standard deviation) in the ratings across teamwork dimensions or in the ratings of team members on specific teamwork dimensions. A larger dispersion demonstrates that students differentiate their perception of teammates' behavior on the various dimensions of teamwork. While it is possible that all team members could exhibit the same level of performance across all five CATME dimensions, this is more likely an indication of assimilation, the tendency to rate all others similarly [9], since it is unlikely that there are no differences in team-member effectiveness among team members on any of the five CATME dimensions [10].

Holding students accountable for ratings given to or by an individual plays an important role in improving the accuracy of the ratings [11, 12] , so various features of the CATME system help instructors evaluate that accountability.

CATME is constructed around five behavioral dimensions: Contributing to the Team's Work, Interacting with Teammates, Keeping the Team on Track, Expecting Quality, and Having Relevant Knowledge, Skills and Attributes (KSAs) [7, 8]. These dimensions are defined briefly as follows:

> **Contributing (C)** to the Team's Work is being able to add value to your team's work/project.
>
> **Interacting (I)** with teammates refers to the way individuals communicate within their teams.
>
> **Keeping (K)** the team on track is similar to being a timekeeper.
>
> **Expecting (E)** quality is taking expectations to the next level and working collaboratively to produce the best possible team outcomes.
>
> **Having (H)** relevant KSAs refers to the base knowledge of individual team members.

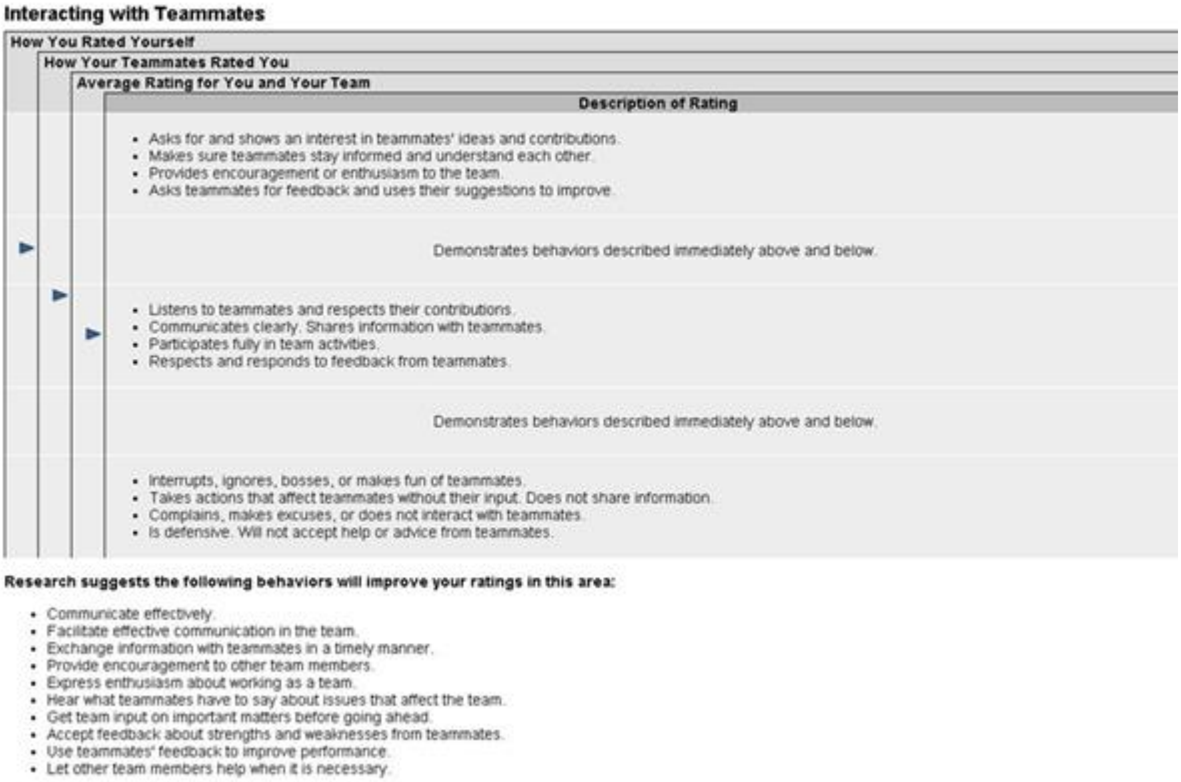
Every aspect of the five teamwork dimensions is equally important to team success and a critical element in the peer evaluations [8].

Peer evaluations facilitate better learning outcomes in upper level education, encouraging students to continue their engagement with constructive team behavior in future team activities[13]. A peer evaluation is an assessment of an individual's contribution to a work activity by their peers, either as students, or as professionals in industry. Peer evaluations help to teach individuals and teams how to act and how to assess one another's performance. Peer evaluations, as facilitated by CATME and its behavior anchored rating system, provide a strong model for facilitating the learning of teamwork behavior [14, 15].

Research Methods

I. CATME Peer Evaluation

CATME peer reviews are conducted online and students receive peer review feedback in an online form. Peer review feedback is shown as pointers to behavioral descriptions, and three arrows indicate the student's self-rating, how that student was rated by peers, and the average rating on that dimension for all team members combined, as shown in Figure 1. This display lacks numbers to emphasize the behavioral focus of the instrument, although ratings are summarized using numbers when reported to instructors [16].



Interacting with Teammates

How You Rated Yourself

How Your Teammates Rated You

Average Rating for You and Your Team

Description of Rating

- Asks for and shows an interest in teammates' ideas and contributions.
- Makes sure teammates stay informed and understand each other.
- Provides encouragement or enthusiasm to the team.
- Asks teammates for feedback and uses their suggestions to improve.

Demonstrates behaviors described immediately above and below.

- Listens to teammates and respects their contributions.
- Communicates clearly. Shares information with teammates.
- Participates fully in team activities.
- Respects and responds to feedback from teammates.

Demonstrates behaviors described immediately above and below.

- Interrupts, ignores, bosses, or makes fun of teammates.
- Takes actions that affect teammates without their input. Does not share information.
- Complains, makes excuses, or does not interact with teammates.
- Is defensive. Will not accept help or advice from teammates.

Research suggests the following behaviors will improve your ratings in this area:

- Communicate effectively.
- Facilitate effective communication in the team.
- Exchange information with teammates in a timely manner.
- Provide encouragement to other team members.
- Express enthusiasm about working as a team.
- Hear what teammates have to say about issues that affect the team.
- Get team input on important matters before going ahead.
- Accept feedback about strengths and weaknesses from teammates.
- Use teammates' feedback to improve performance.
- Let other team members help when it is necessary.

Figure 1: Feedback Screen for Contributing Dimension of the CATME Peer Review Tool

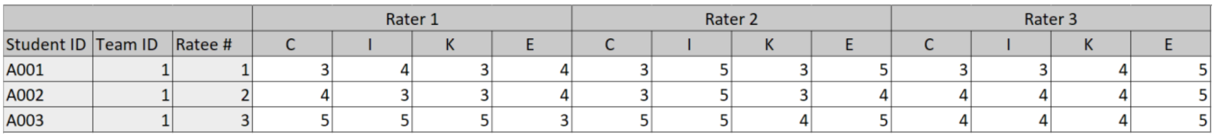
II. Data Collection

For the comparison group of college students several 120 student sections of a Midwestern University FYE program were used from the Fall 2015 semester as the collegiate sample population. FYE students at this institution number over 1,600 (annually) and in Fall 2015 included 23.1% women and 6.4% minorities. Each FYE section enrolls up to 120 students and there were 16 total sections. To conduct this comparison study, we randomly sampled teams from 10 FYE sections totaling 40 teams and160 students for a control or comparison sample.

K12 CATME peer evaluations were completed by students at three different K12 schools from three independent school districts in three different geographic areas of the United States. These schools (K12-1, K12-2, K12-3) used CATME to form teams and conducted multiple peer evaluations during the 2016 – 2017 school years in PLTW, Science and pre-engineering courses. For the K12 experimental sample, K12-1 contributed 26 students from 7 teams, K12-2 contributed 29 students from 9 teams and K12-3 contributed 105 students from 33 teams

III. Data Analysis

The measure of dispersion used in this analysis is defined as the standard deviation of each students' rating of themselves as well as their teammates across the CATME dimensions being used. To be precise, Figure 2 shows a sample of the raw quantitative peer evaluation data. For a team with three members completing a peer evaluation on four dimensions (C, I, K & E), each rater in the team contributes to a 3×4 rating matrix. When all three raters' responses are combined side-by-side, this forms a 3×12 matrix. The standard deviations for each student were calculated by row. In Figure 2, this was the standard deviation for each of the three rows for each of the four dimensions under "Rater 1" label. Then the three row-wise standard deviations were averaged, then placed in a matrix of average dispersions and referred to as the dispersion for the Rater 1's ratings for all the team members including Rater 1's self-rating. This procedure was repeated for Rater 2 and 3 accordingly. The same methods were used to calculate the dispersion matrix for the rest of the secondary school teams as well as the FYE comparison group.



			Rater 1				Rater 2				Rater 3			
Student ID	Team ID	Ratee #	C	I	K	E	C	I	K	E	C	I	K	E
A001	1	1	3	4	3	4	3	5	3	5	3	3	4	5
A002	1	2	4	3	3	4	3	5	3	4	4	4	4	5
A003	1	3	5	5	5	3	5	5	4	5	4	4	4	5

Figure 2 - Raw Peer Evaluation Data

A repeated measure ANOVA was used to compare the differences in dispersions between the K12 and FYE comparison samples for each of two peer reviews. The time delay between reviews was around four weeks for both the K12 schools and FYE comparison group. The difference in dispersion is then calculated using SAS where the standard deviation matrices between two different reviews are compared.

Findings

Table 1: Repeated Measure ANOVA Analysis K12 vs FYE Fall 15

Intervention Sample	Peer Review Time	Intervention Sample	Peer Review Time	Stdv Difference dispersion	P value difference Dispersion	Difference mean rating FYE-K12	P value difference means
K12	2	K12	1	-0.1147	<0.0001	0.08396	0.1358
FYE Control	2	FYE Control	1	-0.01499	0.6190	-0.02942	0.6007
FYE Control	1	K12	1	-0.04555	0.0520	0.06663	0.2996
FYE Control	2	K12	2	0.08410	0.0003	-0.1800	0.0053

N_{K12}=N_{FYE}=160, randomly sampled 160 from FYE, K12 data; 26 K12-1, 29 K12-2, 105 K12-3

Each K-12 class was compared with the FYE Fall 15 sample for two peer evaluations. Table 1 above shows the results from the One-Way ANOVA test.

The difference in dispersion for comparisons across the two reviews showed a significant difference in the K12 time 2 vs time 1 standard deviations comparisons but not the mean scores. In the FYE data, dispersion and mean scores are statistically similar when comparing times 1 and 2.

For time 1 dispersion scores FYE to K12 are different (p= 0.05) whereas mean scores are not (p= 0.30). For time 2 dispersion scores FYE to K12 are different (p= 0.0003) and mean scores are also different (P= 0.005) but the difference in mean ratings changed by 0.25, a large change given the size of our samples.

As defined earlier, a higher standard deviation dispersion value generally suggests that students are better at differentiating one teammate from another and one dimension from another. Regardless of whether the difference between secondary school student rating dispersion and FYE rating dispersion is significant, the difference (Secondary vs FYE) is very small for these comparisons, suggesting that secondary school rating dispersions are at least as good as those provided by FYE students.

A second analytic tool was also used to measure the differences, if any, between self-ratings and others' ratings of that person by team members. The self-ratings were separated from others' ratings of self, their averages across all dimensions were calculated respectively, and compared at each review for every intervention groups. The self-rating and others' ratings are said to be converged if the two ratings for the same individual are not significantly different (using significance level 0.05). A converged peer-evaluation rating across multiple peer evaluations is deemed as a higher quality one because the self and peer ratings have potentially come together either due improved skills in rating teamwork behavior or due to actual changes in the teamwork behavior of the person being rated, potentially as a result of the peer feedback. The samples used in the convergence analysis were the same as those used in the repeated measure ANOVA.

Table 2: Convergence Analysis K12 vs FYE Fall 15 Control

Intervention Sample	Peer Review Time	Difference average	P value
K12	1	-0.4458	<0.0001
K12	2	-0.08291	0.3674
FYE Control	1	0.08576	0.3681
FYE Control	2	0.08909	0.3498

N_{K12}=N_{FYE}=160, randomly sampled 160 from FYE, K12 data; 26 K12-1, 29 K12-2, 105 K12-3

The results of convergence analysis show that in review 1, K12 students tended to rate themselves significantly lower than other teammates' ratings of them, but their ratings converged in review 2, which means their self-ratings became consistent with others' ratings in the second peer evaluations. The self and peer ratings of FYE control group students were similar in both peer evaluation 1 and peer evaluation 2.

The two samples both are converged at time 2. Positive convergence results, as mentioned earlier, are indicators of higher quality peer-evaluation ratings. Our sample data suggests K12 students experience convergence of peer evaluation ratings while our FYE students already provided similar self and peer ratings.

The Social Relationship Model (SRM) [17] was also used to evaluate the key variance components of the peer evaluation ratings. These variances are: rater variance, target variance, dyadic variance, and the team variance. In the following analysis, we mainly consider the rater variance and the target variance. The rater variance measures the tendency of an individual rater to rate his/her teammates consistently. A larger rater effect indicates that raters tend to rate all of their teammates in the same way, which is not deemed as a higher quality rating. The target variance measures the tendency of all team members to rate the same individual consistently. A higher target effect means that all other teammates of one individual rated him/her similarly, which is a more reliable rating. By design, SRM takes only teams with four or more individuals. All teams with sizes less or equal to three were removed before entering the analysis. The analysis was done using R package TripleR [18,19].

Table 3: SRM Analysis K12 vs FYE Fall 15 Control

Variance Component	**K12** Peer Evaluation **1**			**K12** Peer Evaluation **2**			**FYE** Peer Evaluation **1**			**FYE** Peer Evaluation **2**		
	Est.	**SE**	**%**	**Est.**	**SE**	**%**	**Est.**	**SE**	**%**	**Est.**	**SE**	**%**
Rater	0.185	0.059	***25.2***	0.236	0.064	***27.9***	0.206	0.021	***35.2***	0.207	0.018	***43.2***
Target	0.228	0.074	***31.2***	0.267	0.101	***31.6***	0.183	0.023	***31.3***	0.102	0.015	***21.3***
Relationship	0.319	0.060	***43.6***	0.342	0.053	***40.5***	0.196	0.012	***33.5***	0.171	0.010	***35.5***

$N_{\text{K12\_review 1}}$ = 23 teams, 92 individuals, $N_{\text{K12\_review 2}}$ = 29 teams, 116 individuals
$N_{\text{Control\_review 1}}$ = 150 teams, 600 individuals, $N_{\text{Control\_review 2}}$ = 172 teams, 688 individuals

Comparing the SRM results of K12 samples and FYE control group samples, we observe increases in Rater variances between peer evaluation 1 and peer evaluation 2 for FYE and not in the K12 sample. In terms of target variance, in peer evaluation 1, the two samples have comparable target variances, while in review 2, the FYE control sample has smaller target variances. Since lower rater variances and higher target variances are both indicators for higher quality ratings, the SRM analysis on current samples suggest K12 students have at least as good peer-evaluation ability as our FYE control group students.

Conclusion

From this preliminary analysis we conclude that it is likely that secondary students in teams behave similarly to FYE students placed in teams in regards to their ratings of themselves and their peers as measured when using the CATME tool.

Further Research

Due to limited samples available from the three secondary schools, in this analysis we only evaluated two peer evaluations, which is not enough to fully capture any lasting effects from the peer evaluation feedback or peer evaluation trainings that might have taken place. It also will be meaningful to extend the research to more secondary schools/a larger sample, and to analyze a larger number of peer-evaluations per team to more accurately measure any lasting effects.

References

1. National Association of Colleges and Employers. in Job Outlook (2011).
2. Calloway School of Business and Accountancy of Wake Forest University. (2004).
3. Baker M. in AIB Official Blog (2014).
4. ABET. (ed. Engineering Accreditation Commission) (Baltimore, Md., 2003).
5. Dictionary.com. teamwork [Online]. (2017).
6. J. C. Chen & J. Chen. Testing a new approach for learning teamwork knowledge and skills in technical education. *Journal of Industrial Technology* **20**, 37-46 (2004).
7. Loughry, M.L., Ohland, M.W. & Moore, D.D. Development of a theory-based assessment of team member effectiveness. . *Educational and Psychological Measurement* **67**, 505-525 (2007).
8. Ohland, M.W. et al. 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**, 609-630 (2012).
9. LeDoux J. A., Gorman, C.A. & Woehr, D.J. The impact of interpersonal perceptions on team processes: A social relations analysis. *Small Group Research* **43**, 356-382 (2012).
10. Resick C. J. , Dickson M. W. , Mitchelson J. K., Allison L. K. & Clark M. A. . Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy. *Group Dynamics: Theory, Research, and Practice* **14**, 174-191 (2010).
11. Mero, N.P., Motowidlo, S.J. & Anna, A.L. Effects of Accountability on Rating Behavior and Rater Accuracy. *Journal of Applied Social Psychology* **33**, 2493-2514 (2003).
12. Mero, N.P. & Motowidlo, S.J. Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology* **80**, 517-524 (1995).
13. Thomas E. J. . (BMJ quality & safety, 2011).
14. Loughry M. L. , Ohland M. L. & Woehr D. J. . Assessing teamwork skills for assurance of learning using CATME Team Tools. *Journal of Marketing Education* **36**, 5-19 (2014).
15. Ohland M. W. , Layton R. A. , Loughry M. L. & Yuhasz A. G. . Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education* **94**, 319-326 (2005).
16. Ferguson D. M. , Lally C. , Somnooma H. I. , Murch O. & Ohland M. . in Frontiers in Education (Eire, PA, 2016).
17. Kenny, D. A. *Interpersonal perception: A social relations analysis*. (Guilford Press, NY, 1994)
18. Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. TripleR: An R package for social relations analyses based on round-robin designs. Behavior Research Methods, **44**, 455–470. (2012)
19. Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. TripleR: Social Relation Model (SRM) analyses for single or multiple groups (R package version 1.4.1). (2015) Retrieved from http://cran.r-project.org/web/packages/TripleR.