ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #36480

Team Tenure - the Longitudinal Study of Engineering Student Peer Rating Quality

Chuhan Zhou

Siqing Wei

Siqing Wei received B.S. and M.S. in Electrical Engineering from Purdue University. He is currently pursuing a Ph.D. degree in Engineering Education program at Purdue University. After years of experience serving as a peer teacher and a graduate teaching assistant in first-year engineering courses, he has been a research assistant at CATME research group studying multicultural team dynamics and outcomes. The research interests span 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) to promote DEI. In addition, he also works on many research-to-practice projects to enhance educational technology usage in engineering classrooms and educational research by various methods, such as natural language processing. In addition, he is also interested in the learning experiences of international students. Siqing also works as the technical development and support manager at the CATME research group.

Matthew W. Ohland (Dale and Suzi Gallagher Professor of Engineering Education)

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. He studies the longitudinal study of engineering students and forming and managing student teams and with collaborators has been recognized 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.

Team Tenure - the Longitudinal Study of Engineering Student Peer Rating Quality

Abstract

To achieve the benefit of the growing emphasis on teamwork in engineering courses, it is important to improve how student teams are managed. A prominent way to promote social constructive collaboration is the use of peer evaluations, which enable engineering students to learn from feedback provided from other team members to improve their performance in teams. In this work, we studied the effect of repeated use of a peer evaluation system over time in multiple classes on the quality of peer evaluations in a course late in the students' program. Specifically, we studied the repeated use of the Comprehensive Assessment of Teamwork Effectiveness (CATME) peer evaluation system in a senior level civil engineering course in an Australian University using MANCOVA. Teams whose members had longer team tenure—based on completing more surveys or having more experience in prior teams using CATME—provide more consistent ratings of teammates. This adds to the body of evidence that the quality of ratings improves with repeated use of a peer rating system. We suggest that curriculum decision-makers adapt more widespread use of peer evaluation in courses that involve teamwork to encourage students to develop the valuable skill at providing feedback to peers, which in turn will promote the improvement of team skills.

Introduction

Teamwork is recognized as an essential skill in many fields. Teamwork activities are in increasing demand in industry and education because modern tasks and products are more complicated than before and companies need effective collaboration in their teams to meet their goals [1]–[5]. Teams whose members are dependent of each other in their goals and tasks perform better than individuals [6]–[8]. Working in teams has the potential to provide a higher level of diversity, and teammates could learn from each other and exchange more opinions [3]. Hence, effectively working in teams is a crucial ability for candidates seeking jobs, and recruiters continue to evaluate this ability [9], [10]. According to the Job Outlook survey performed by the National Association of College and Employers (2010), the "ability to work in a team structure" was the most critical skill that recruiters are looking for in college graduates [11].

To equip students or staff with teamwork skills, many studies across a range of disciplines have explored approaches training [12]–[14]. Engineering's accreditation body in the United States, ABET, requires institutions to demonstrate multidisciplinary teamwork skills development in their education [15]. Engineering educators have revised their education approach to prepare students for industry, and the engineering curriculum and prevailing pedagogies have changed – and continue to change – to meet these demands [16], [17]. Among those changes, engineering programs have added courses that teach students teamwork skills and/or require students to work in teams for assignments [17].

To facilitate teamwork training, peer evaluations are commonly used in many college courses and industries to evaluate team members. Many studies use peer evaluations as an indication of teamwork quality [18]–[22]. Peer evaluation offers several advantages, such as developing teamwork skills by creating a sense of responsibility among teammates [22], and motivating teammates to develop good team skills and contribute to the teamwork [23], [24]. The peer evaluation data used in this work were collected using the CATME system and were voluntarily released in deidentified form for research purposes. The CATME (Comprehensive Assessment of Team Member Effectiveness) peer evaluation is widely used in engineering and other disciplines. Launched in 2005 [11], CATME has been used by over 23,000 instructors in different fields at more than 2500 institutions. A main function of the CATME Team Tools suite is its Peer Evaluation system, which is based on five dimensions: Contributing to the Team's Work (C), Interacting with Teammates (I), Keeping the Team on Track (K), Expecting Quality (E), and Having Relevant KSAs (H) [25]. The ratings are performed on a 5-point scale with behavioral anchors at the top, middle, and bottom of the scale, and students are expected to reflect on their own effectiveness as well as provide feedback to their teammates.

In personal psychology, tenure can be defined generally as the length of time period served as a particular role or in organization [26], [27]. In the context of teamwork, tenure has various definitions, including the average amount of time individual members have spent in their team, job, position, or organization (additive tenure); the amount of time team members have been together on the same team (collective tenure); the variability in the amount of time individual team members have spent in their job, team, or organization (team tenure dispersion) [26], [27]. These definitions all capture the essence of what we are studying here – the impact of collectively cumulated experience. Therefore, we argue that team tenure is an appropriate independent variable of interest to investigate how rating quality is related to engineering students' accumulated teamwork experience. We operationalize tenure in two ways: tenure is calculated as the cumulative number of peer evaluations completed using CATME; tenure is calculated as the number of distinct teams they were a member of while completing those peer evaluations. Other research has addressed the connection between teamwork quality and students' experience in teamwork [28]–[31], but results vary and a larger research base can help clarify the relationship. Lewis and colleague found that repeated interactions among individual team members would enhance TMS (transactive memory system) development, which is a memory system for groups to collectively encode, store, retrieve, and communicate knowledge that positively influences teamwork performance [31], [32]. Mulé and colleagues found that team performance benefits from having team members with higher additive tenure, collective tenure, and team tenure dispersion [27], whereas Tesluk & Jacobs concluded that tenure would not have the same impact for all team members [26]. The findings of Tesluk & Jacobs are supported by others – while some team members improve their performance over time, others may get worse, and a few may change unsystematically, as time-based measures of experience cannot account for differences between individuals while intraindividual change occurs [33], [34]. One of the most relevant studies is from Brutus and Donia, which concluded that repeated use of a peer evaluation system would improve the peer rating quality [35]. Yet while Brutus & Donia used peer evaluation data as dependent variable in their studies [28], [35], most other studies of the effect of teamwork experience use other methods, so we decided to use peer evaluation results and quantitative methods to examine the relationship of experience in teams using a common peer evaluation system would improve their peer rating quality. Therefore, the research question is: How does the teamwork tenure of students affect the quality of peer evaluation ratings?

Methodology

The data were collected from a senior civil engineering course in an Australian university that used CATME peer evaluations at multiple points throughout the curriculum. Students in this senior course were assigned to teams of five members and completed three rounds of peer evaluation. In each round of peer evaluation, students rated themselves and each teammate. Among 62 teams for the course, 54 of them (87.1%) did not experience a shift in sample (i.e. all five members in the team are same in all three peer evaluations) and had at least 4 out of 5 members participating in the ratings for all members in all three peer evaluations. In each case, instructors released peer evaluation results to students for students and teams as feedback to improve their individual and collective teamwork behaviors.

In this study, we use the standard deviation of ratings to operationalize rating quality. For having high peer evaluation quality, we expect to see a smaller standard deviation. A smaller standard deviation results from peer ratings that are more similar, which means that members of the team generally agree on the team-member effectiveness of the various members of the team.

Data Cleaning

The sample data were cleaned using two criteria:

(1) we kept only data from teams in which all five members are the same in all three rounds of peer evaluations, and

(2) we removed the data for a team if fewer than four team members completed the survey in a round of peer evaluation. This ensures peer evaluation quality, since too much missing data would reduce the interpretability of our results due to the round-robin nature of peer evaluations [36], [37].

Simple Linear Regression and MANCOVA Analysis

We used statistical methods and models to investigate the relationship between students' rating quality and their experience in teamwork, first constructing scatterplots on all variables. We then conducted a simple linear regression (SLR) analysis predicting the standard deviation of ratings for each dimension as a function of the tenure in the team, which is the average number of CATME surveys completed by team members and the average number of groups in which team members used CATME (measured as separate variables). This computation is performed for each peer evaluation to account for the fact that students are accumulating additional experience throughout the course. We used SLR here since we aim to explore the fundamental relationship between tenure and rating quality, and SLR is suitable for describing the strength of relationship between two numerical variables. The Brown-Forsythe test [38] and Shapiro test [39] were used to verify the assumptions of the constant variance of residuals and normality. We then used multi-variate analysis of covariates (MANCOVA) [40] to explore the pattern of rating quality change across peer evaluations. To achieve more comprehensive results, we conducted two types of analyses: Across-Dimension Analysis (ADA) and in Individual-Dimension Analysis (IDA) for CATME five dimensions described in the introduction. In IDA, the analysis of relationship between rating quality and students' experience is conducted in each single dimension, while we average all the rating scores for individual dimensions then calculate the standard deviations and analyze the relationship in ADA.

We also checked the assumptions for the MANCOVA analysis. First, for each row (observation) we collected contain peer evaluation round, tenure, and standard deviations of different dimensions. For each peer evaluation round, though the teamwork experience of students might be somewhat similar to each other since they are taking the common curriculum, the teams are assembled by the instructor by a process that does not consider tenure, which is an independent variable in this study. Since all ratings are completed individually, it is reasonable to assume that standard deviations of different dimensions for different teams are independent of each other. Second, though the rating results of different CATME dimensions are correlated to each other to the medium degree, the content of dimensions are validated to be distinguished from each other evidenced by the factor analysis results [25]. Therefore, we treat the rating results of CATME dimensions to be independent variables. Third, the independent variable is categorical and the dependent variable is continuous. Fourth, according to the Brown-Forsythe test and Shapiro test conducted above, our data has constant variance of residuals and the vast majority of our data follow a normal distribution. Finally, it is reasonable to assume that dependent variables (standard deviations of different dimensions) are significantly related to covariate (tenure), and our MANCOVA analysis result also verifies this.

Results

The analysis was completed in the R statistical package [41], and below we show scatterplots matrices for the variables in all three peer evaluations. In this figure, the average team tenure is TeamAvg and the average survey tenure is SurveyAvg. The standard deviation of the various dimensions is represented by adding SD to the dimension labels C-H above to become CSD-HSD. The pooled standard deviation for ADA is shown as ASD. We can see that in most cases, especially in Figure 2 and 3, the standard deviations of different dimensions correlate with each other (i.e., when the standard deviation of ratings in one dimension is lower, it would more likely be lower in other dimensions as well). While there is considerable variability, it appears that there is a tendency that when the tenure values are higher, the standard deviations of ratings tend to be slightly lower.

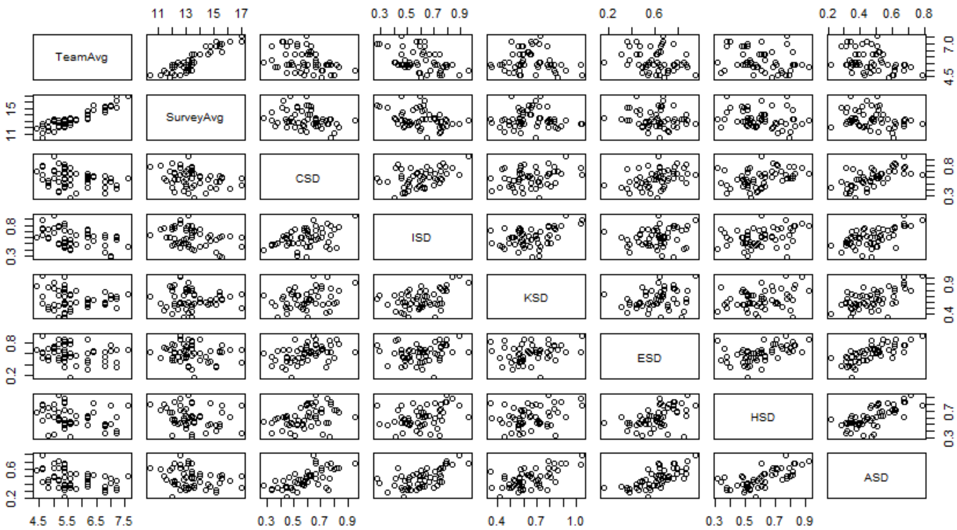

Figure 1: Scatterplot of Variables in Peer Evaluation 1

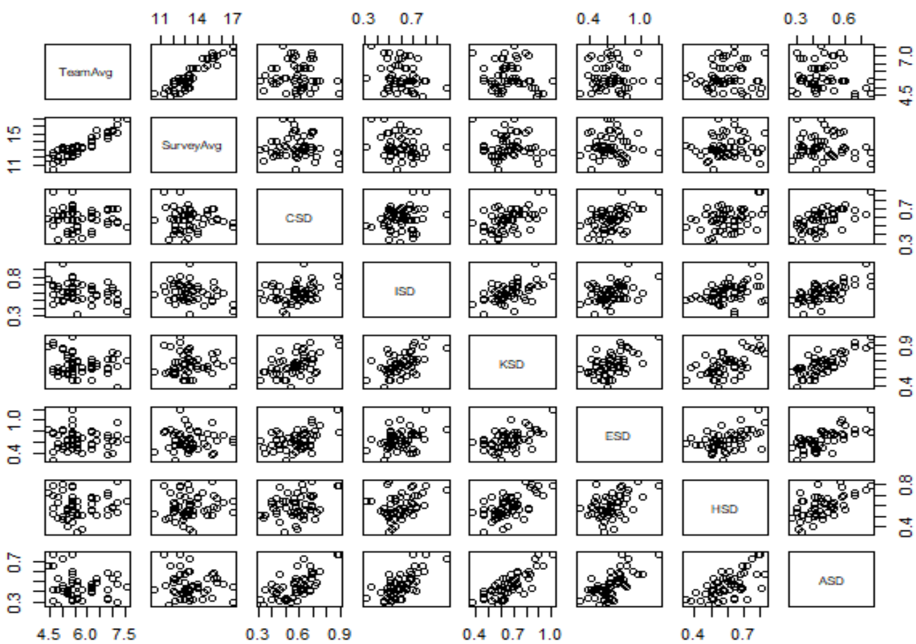

Figure 2: Scatterplot of Variables in Peer Evaluation 2

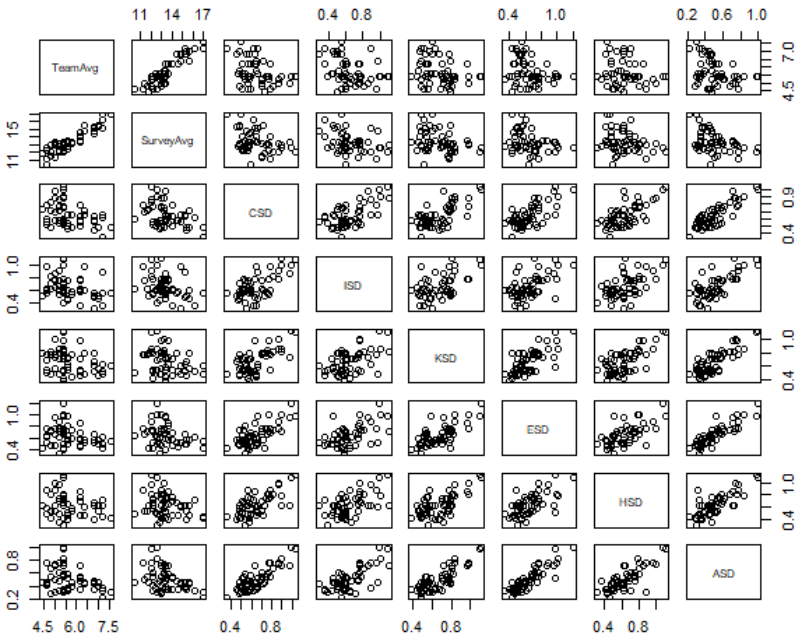

Figure 3: Scatterplot of Variables in Peer Evaluation 3

In the following tables, we assembled all the results in our SLR analysis for all three peer evaluations.

Table 1. SLR Results between Tenure and Rating Quality for PE1, PE2, and PE3

Peer Evaluation	Tenure type	Dimension	p-value	Intercept	Slope	p Shapiro Test	p B-F Test
PE1	Team	**A**	**0.009**	0.78358	-0.0548	0.647	0.828
PE1	Team	**C**	**0.008**	0.95326	-0.0638	0.748	0.983
PE1	Team	**I**	**0.001**	1.06633	-0.0812	***0.042***	0.379
PE1	Team	K	0.386	0.78423	-0.0233	0.776	0.789
PE1	Team	E	0.091	0.87101	-0.0459	0.678	0.698
PE1	Team	H	0.098	0.84385	-0.0417	0.378	0.988
PE1	Survey	A	0.072	0.76425	-0.0219	0.613	0.700
PE1	Survey	**C**	**0.029**	0.99562	-0.0303	0.638	0.846
PE1	Survey	**I**	**0.012**	1.07315	-0.0350	0.148	0.355
PE1	Survey	K	0.789	0.91641	-0.0231	0.537	0.434
PE1	Survey	E	0.268	0.84043	-0.0173	0.926	0.925
PE1	Survey	H	0.106	0.91654	-0.0231	0.473	0.603
PE2	Team	A	0.350	0.56912	-0.0191	0.115	0.600
PE2	Team	C	0.859	0.59931	-0.0038	0.404	0.505
PE2	Team	I	0.052	0.81488	-0.0383	0.568	0.957
PE2	Team	K	0.303	0.79398	-0.0238	0.641	0.684
PE2	Team	E	0.952	0.65053	-0.0018	0.332	0.865
PE2	Team	H	0.741	0.56103	0.0062	0.669	0.319

PE2	Survey	A	0.559	0.55168	-0.0068	***0.047***	0.976
PE2	Survey	C	0.941	0.58975	-0.0009	0.392	0.683
PE2	Survey	I	0.099	0.84546	-0.0186	0.198	0.598
PE2	Survey	K	0.698	0.72685	-0.0051	0.604	0.882
PE2	Survey	E	0.823	0.69048	-0.0037	0.330	0.776
PE2	Survey	H	0.483	0.49645	0.0074	0.522	0.982
PE3	Team	**A**	**0.008**	0.94330	-0.0768	0.051	0.373
PE3	Team	**C**	**0.005**	1.04966	-0.0707	0.599	0.883
PE3	Team	**I**	**0.020**	1.06535	-0.0726	***0.015***	0.914
PE3	Team	**K**	**0.035**	1.02100	-0.0608	0.201	0.714
PE3	Team	E	0.065	0.97436	-0.0565	0.080	0.163
PE3	Team	H	0.180	0.87539	-0.0412	0.281	0.947
PE3	Survey	**A**	**0.002**	1.19702	-0.0516	0.077	0.739
PE3	Survey	**C**	**0.004**	1.19584	-0.0410	***0.015***	0.939
PE3	Survey	**I**	**0.012**	1.25066	-0.0447	0.066	0.988
PE3	Survey	**K**	**0.003**	1.16474	-0.0366	0.226	0.940
PE3	Survey	**E**	**0.025**	1.17365	-0.0389	0.084	0.399
PE3	Survey	H	0.058	1.08108	-0.0328	0.561	0.733

Table notes: Significant p-values for slope contributions are **shown in bold**. Evidence of non-normality is ***indicated by bold italics***.

From the above table, we can see that all slope values, except survey in Dimension H of Peer Evaluation 2, are negative, as we expected. We can see that in ratings for many dimensions, especially in Peer Evaluations 1 and 3, the number of teams and surveys students previously experienced has a significant effect on rating quality (taking $\alpha = 0.05$ as the threshold of significance). When the tenure value is higher, the standard deviation values of ratings are lower. All the data above passed the Brown-Forsythe test and the vast majority of data passed the Shapiro test, meaning that the residuals of the data have constant variance and most of the data follow a normal distribution.

The MANCOVA analysis is a better approach for controlling the system-wide error rate and is conducted using the peer evaluation round as independent variable, standard deviations of different dimensions as dependent variables, and tenure as a covariate. Since tenure is represented by both the number of CATME surveys and the number of teams using CATME each student has experienced, we conducted the primary analysis separately for each measure of tenure. We created Figure 4 using the standard deviations of ratings in each dimension for each peer evaluation to show visually how standard deviations for each dimension differ and change across peer evaluations.

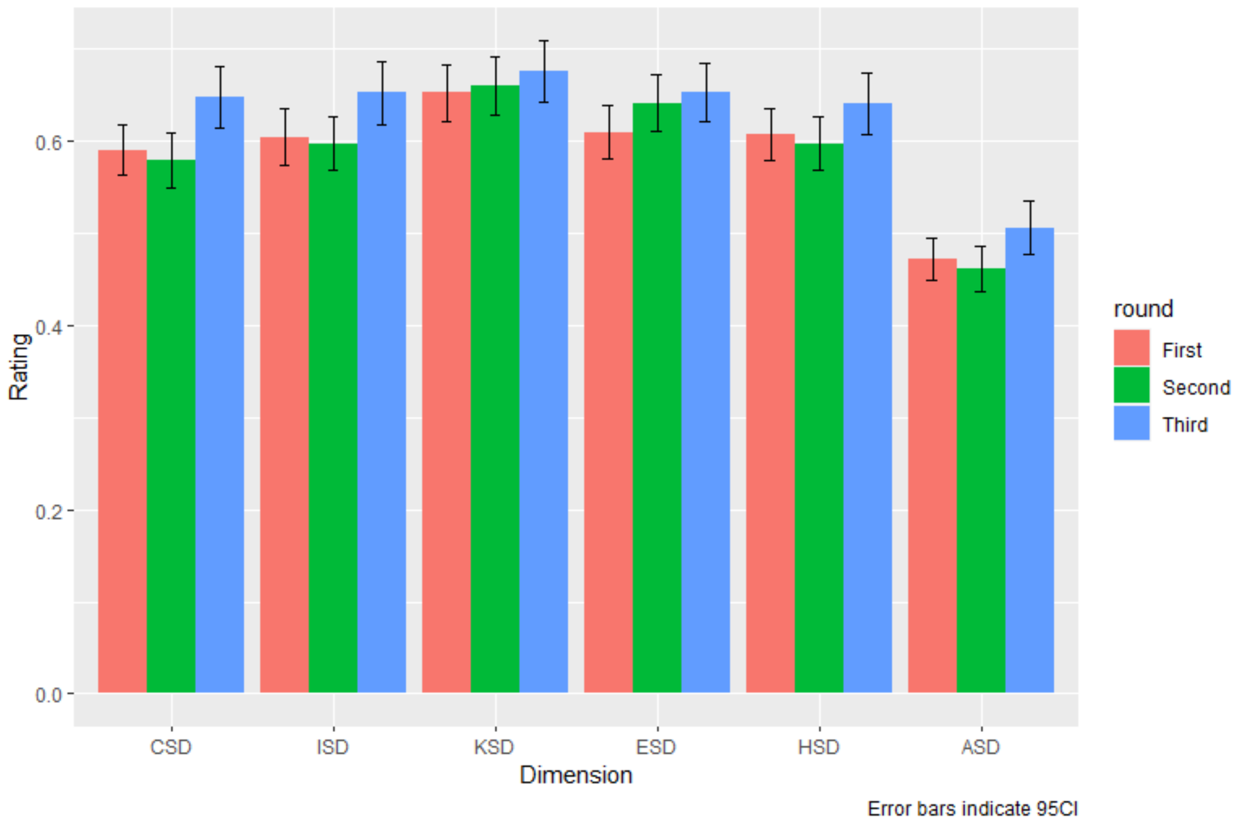

Figure 4. Mean of Standard Deviations Across Dimensions for Different Peer Evaluations

According to the result of the primary MANCOVA analysis, we found that the standard deviations of ratings are not significantly related to the peer evaluation round when the covariate was the average number of surveys the team members had taken, $V = 0.0483$, $F(6, 154) = 1.302$, $p = 0.259$, and when the covariate was the average number of teams the members of the team have been in, $V = 0.0487$, $F(6, 154) = 1.313$, $p = 0.254$. The standard deviation of ratings itself is, however, significantly related to both covariates – the average number of surveys taken ($V = 0.113$, $F(6, 154) = 3.254$, $p = 0.005$) and the average number of teams the members of the team have been in ($V = 0.135$, $F(6, 154) = 4.018$, $p < 0.001$). Hence, from MANCOVA analysis, tenure is argued to serve as a factor related to more consistent ratings by students of their teammates.

Discussion

In the above section, we plotted scatterplots and used SLR and MANCOVA to comprehensively investigate the relationship between peer rating quality and tenure, as well as its relationship with experience in a particular team. From the scatterplots, it appears there is a tendency that when the tenure values are higher, the rating quality is also higher since the standard deviations of ratings tend to be slightly lower, and the statistical results in SLR and MANCOVA confirm this. In addition, the MANCOVA analysis reveals that there is no significant change in this tendency across peer evaluation rounds.

Conclusion

Engineering courses must continue to emphasize teamwork, and previous research finds a relationship between students' experience in teamwork and teamwork quality [28]–[30]. This work adds further weight of quantitative evidence to that research, much of which is qualitative. While the experience of multiple peer evaluation administrations in a single team experience does not have a significant impact on the quality of peer ratings, a longer-term commitment to providing team experiences with a consistent peer evaluation process does have a significant impact. Thus our findings provide support for a more comprehensive teamwork education for engineering students involving multiple experiences that are consistently managed, a recommendation supported by others [28]–[30].

References

[1] S. B. Parumasur, "The Importance of Teamwork, Continuous Top Management Support and Training in Bringing About TQM," *J. Econ. Behav. Stud.*, 2013.

[2] A. A. Ghaferi and J. B. Dimick, "Importance of teamwork, communication and culture on failure-to-rescue in the elderly," *British Journal of Surgery*. 2016.

[3] B. Gummer, "Go team go!: The growing importance of teamwork in organizational life," *Adm. Soc. Work*, 1996.

[4] E. Mohrman, A. Lawler, and G. Benson, *Organizing for high performance: Employee involvement, TQM, reengineering, and knowledge management in the Fortune 1000*. San Francisco, 2001.

[5] D. Cordes, J. Parker, D. Nikles, A. Hopenwasser, C. Laurie, and J. Izatt, "Teaming in technical courses," 2002.

[6] M. Hoegl and H. G. Gemuenden, "Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence," *Organ. Sci.*, 2001.

[7] M. L. Loughry, M. W. Ohland, and D. Dewayne Moore, "Development of a theory-based assessment of team member effectiveness," *Educ. Psychol. Meas.*, 2007.

[8] J. J. DiStefano and M. L. Maznevski, "Creating value with diverse teams in global management," *Organ. Dyn.*, vol. 29, no. 1, pp. 45–63, Sep. 2000.

[9] R. Alsop, "Playing well with others," *Wall Str. J.*, vol. 11., 2002.

[10] National Association of Colleges and Employers, "Job Outlook 2011," 2010.

[11] M. L. Loughry, M. W. Ohland, and D. J. Woehr, "Assessing Teamwork Skills for Assurance of Learning Using CATME Team Tools," *J. Mark. Educ.*, 2014.

[12] C. Jaca, E. Viles, M. Tanco, R. Mateo, and J. Santos, "Teamwork effectiveness factors in healthcare and manufacturing industries," *Team Perform. Manag.*, 2013.

[13] R. Lingard and E. Berry, "Teaching teamwork skills in software engineering based on an understanding of factors affecting group performance," in *Proceedings - Frontiers in Education Conference*, 2002.

[14] O. S. Persons, "Factors Influencing Students' Peer Evaluation in Cooperative Learning," *J. Educ. Bus.*, 1998.

[15] ABET, *2018-19 Criteria for Accrediting Engineering Programs*. Baltimore, MD: ABET, 2019.

[16] S. G. Adams, "Building Successful Student Teams in the Engineering Classroom," *J. STEM Educ. Innov. Res.*, 2003.

[17] A. Kedrowicz, S. Watanabe, D. Hall, and C. Furse, "Infusing technical communication and teamwork within the ECE curriculum," *Turkish J. Electr. Eng. Comput. Sci.*, 2006.

[18] J. E. Bono and A. E. Colbert, "Understanding responses to multi-source feedback: The role of core self-evaluations," *Pers. Psychol.*, 2005.

[19] G. Dai, K. P. De Meuse, and C. Peterson, "Impact of multi-source feedback on leadership competency development: A longitudinal field study," *J. Manag. Issues*, 2010.

[20] J. W. Smither, M. London, and R. R. Reilly, "Does performance improve following multisource feedback? a theoretical model, meta-analysis, and review of empirical findings," *Pers. Psychol.*, 2005.

[21] E. Pfaff and P. Huddleston, "Does It Matter if I Hate Teamwork? What Impacts Student Attitudes toward Teamwork," *J. Mark. Educ.*, 2003.

[22] A. C. Loignon, D. J. Woehr, J. S. Thomas, M. L. Loughry, M. W. Ohland, and D. M. Ferguson, "Facilitating peer evaluation in team contexts: The impact of frame-of-reference rater training," *Acad. Manag. Learn. Educ.*, 2017.

[23] S. A. Hernandez, "Team Learning in a Marketing Principles Course: Cooperative Structures That Facilitate Active Learning and Higher Level Thinking," *J. Mark. Educ.*, 2002.

[24] B. J. Millis and P. G. Cottell, *Cooperative Learning for Higher Education Faculty. Series on Higher Education.* 1998.

[25] M. W. 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.*, 2012.

[26] P. E. Tesluk and R. R. Jacobs, "Toward an integrated model of work experience," *Pers. Psychol.*, 1998.

[27] E. Gonzalez-Mulé, B. S. Cockburn, B. W. McCormick, and P. Zhao, "Team tenure and team performance: A meta-analysis and process model," *Pers. Psychol.*, 2020.

[28] S. Brutus and M. Donia, "Improving the effectiveness of students in groups with a centralized peer evaluation system," *Acad. Manag. Learn. Educ.*, 2010.

[29] J. Biggs and C. Tang, *Teaching for quality learning at university*. 2011.

[30] A. Planas-Lladó *et al.*, "An analysis of teamwork based on self and peer evaluation in higher education," *Assess. Eval. High. Educ.*, 2021.

[31] K. Lewis, D. Lange, and L. Gillis, "Transactive memory systems, learning, and learning transfer," *Organ. Sci.*, 2005.

[32] J. L. Cordery and C. Soo, "Overcoming impediments to virtual team effectiveness," *Hum. Factors Ergon. Manuf.*, 2008.

[33] D. Hofmann, *Task performance and transfer: An assessment of interindividual differences in intraindividual change*. 1992.

[34] D. A. Hofmann, R. Jacobs, and J. E. Baratta, "Dynamic Criteria and the Measurement of Change," *J. Appl. Psychol.*, 1993.

[35] S. Brutus, M. B. L. Donia, and S. Ronen, "Can business students learn to evaluate better? Evidence from repeated exposure to a peer-evaluation system," *Academy of Management Learning and Education*, vol. 12, no. 1. 2013.

[36] D. G. Altman and J. M. Bland, "Missing data," *Br. Med. J.*, 2007.

[37] D. A. Newman, "Missing Data: Five Practical Guidelines," *Organ. Res. Methods*, 2014.

[38] M. B. Brown and A. B. Forsythe, "Robust tests for the equality of variances," *J. Am. Stat. Assoc.*, 1974.

[39] S. S. SHAPIRO and M. B. WILK, "An analysis of variance test for normality (complete samples)," *Biometrika*, 1965.

[40] C. J. Huberty and M. D. Petoskey, "Multivariate Analysis of Variance and Covariance," in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, 2000.

[41] R. C. Team, "R Core Team 2021," *R: A Language and Environment for Statistical Computing.*, 2021. .