Gender differences in students' team expectations and experiences in introductory team-based courses

Margaret Carroll
Mr. James A. Coller, University of Michigan

James Coller is an engineering PhD Candidate at the University of Michigan focusing on the development of a novel multi-layer network approach to understanding design complexity in unmanned maritime vehicles. James also completed his BSE and MSE in Naval Architecture and Marine Engineering in 2017 and 2018 respectively and a MS in Robotics in 2019 at Michigan. His research interests include autonomous robotics for both land and marine environments, ship design for the U.S. Navy, and improving equity and inclusion in engineering learning environments.

James spent three years during his undergraduate education as an instructional aide for the design-build-test-communicate course described within this paper. Since leaving the instructional staff, James continues to work with the course to research educational methods. James is also an occasional guest lecturer for this course teaching about field robotics and team communication challenges.

Dr. Laura K Alford, University of Michigan

Laura K. Alford is a Lecturer and Research Investigator at the University of Michigan. She researches ways to use data-informed analysis of students' performance and perceptions of classroom environment to support DEI-based curricula improvements.

Ms. Roxanne Pinsky, University of Michigan
Mr. Sangam Munsiff, University of Michigan
Mr. Charles William Schertzing, University of Michigan
Ms. Julia T. Toye, University of Michigan
Mr. Magel P. Su, California Institute of Technology

Magel P. Su is a PhD student in the Department of Applied Physics and Materials Science at the California Institute of Technology under the direction of Professor Harry Atwater. He has a B.S.E in materials science and engineering and a minor in chemistry from the University of Michigan. At Michigan, he was a member of the Ultrafast Laser - Material Interaction Laboratory and the Engineering Honors Program. He also served as an instructor for several courses including Introduction to Engineering, Introduction to Materials and Manufacturing, and Structural and Chemical Characterization of Materials.

Dr. Robin Fowler, University of Michigan

Robin Fowler is a lecturer in the Program in Technical Communication at the University of Michigan. She enjoys serving as a "communication coach" to students throughout the curriculum, and she's especially excited to work with first year and senior students, as well as engineering project teams, as they navigate the more open-ended communication decisions involved in describing the products of open-ended design scenarios.

Gender Differences in Students' Team Expectations and Experiences in an Introductory Project-Based Team-Based Course

Abstract

This paper is student-led engineering education research on gender differences in student responses to experience working on teams in an introductory engineering course. We are past and present teaching assistants for a team-based, project-based introductory engineering course at a large Midwestern research university. We strive to create and maintain an equitable and inclusive environment so that our diverse students can feel comfortable and excel. We regularly assess our students' perceptions of their team experience, but we are aware that some social groups can be at a disadvantage when it comes to teamwork. We chose to look at gender in this particular research endeavour. Our assessment data is derived from an online team support tool hosted by our institution. To better use this tool and improve support for our teams, we investigated the relationship between gender and student rating of teamwork behaviors.

The main research objective was to address the following question: Is there a relationship between gender and peer-reported teamwork behaviors in the class as assessed at the end of the term? We focused on the individual's contribution of enacted ideas, contribution of valuable ideas, reliability, and listening skills as assessed by their peers. We find that women are rated higher on both reliability and listening skills.

The data set is gathered by the online team support tool hosted by our institution and is approved by our Institutional Review Board (IRB) [HUM# 00135376]. The data set is from two semesters of a team-based, project-based introductory engineering course with a total of 118 student responses analyzed (N = 118).

Introduction and Motivation

There are many examples in the engineering education literature of teamwork being particularly fraught for women. Specifically, women sometimes end up completing less technical work and more project management work [1, 2, 3]; they are sometimes spoken over in conversations [4, 5]; and they are sometimes evaluated by themselves and their peers according to different standards than their peers who are men [6, 7, 8, 9]. All these examples led us to examine how gender differences impacted team-dynamics in our course.

In many classes, peer evaluations are implemented as a way of holding students accountable for their individual efforts to the team. This effort can be problematic in first year engineering

courses because students do not yet have the technical background to allow for interdisciplinarity, resulting in tasks that are less differentiated and more integrated [10]. While it is difficult to measure effective teamwork [11], when instructors adopt peer assessment, social loafing decreases [10]. A decrease in social loafing greatly improves a team's efficiency and efficacy.

Therefore, many courses, including ours, have adopted peer assessments in order to combat social loafing. In our course and many other team project engineering courses, these peer assessments are sometimes used to affect the student's grade [11]. However, this practice is potentially problematic; students sometimes over-value their own work and undervalue the work of others [12]. Perhaps most worrisome is that student ratings of others may be impacted by their teammates' identity characteristics [8, 9, 13, 14, 15, 16, 17]. Note that not all similar studies find these effects (see, for example: [18, 19]).

This research examines the potential problem of using peer assessments in team-based classes as a grade influence due to teammates' diverse identity characteristics. In particular, we investigated students' ratings of teammates regarding the enactment of their ideas, their listening ability, their reliability, and their possession of relevant valuable skills.

Course Background

The course used for data collection and examination of team-dynamics in this research was a first-year Design, Build, Test, & Communicate engineering class. Sixty students are enrolled in the class each semester and attend a lecture semi-weekly. Students are also divided into three lab and discussion sections of twenty students. The discussion sections focus on technical communication (written and oral) while the labs build hands-on skills. The course is built around a semester long project that challenges students to build an underwater remotely operated vehicle (ROV). The project culminates in an end of semester competition and presentation to faculty, peers, and industry representatives. While the ROV project is the highlight of the class, the main goal of the class is to help students understand how to work in teams of four or five students and effectively communicate both within the team and to external stakeholders.

Faculty and Staff

The course is instructed by two or three co-instructors: one technical lecturer and one or two technical communication lecturers. They share lecture time, and the technical communication lecturer(s) also act as the smaller twenty person discussion section lead(s). The labs are led by a professional lab manager and four instructional assistants (IAs) who are selected by the faculty from upper-class students who excelled in the class when they took it during their first year. Three of the IAs are assigned to the three lab sections, and the fourth runs supplementary labs and supports the faculty and students. These many levels of instructors provide teams multiple avenues for asking questions, raising issues, and reaching help.

Peer Mentors

As an additional layer of personalized help, peer mentors are upper-class students who have previously taken the course and volunteer to come back and aid in mentoring an individual ROV team. Each ROV team in the class has 1-2 peer mentors that help guide them through the project, the presentations, and in managing team meetings outside of class time. Peer mentors have provided indispensable insight to the rest of the faculty on understanding struggling teams and ways to help teams that are more hesitant to ask questions to faculty and staff. For more information on the class peer mentoring and structure, see [20].

Methods

Students complete a Mid-term Survey and an End of Term Survey for the two major team evaluations in the project, assessing themselves and their teammates on a variety of project contribution and teamwork behaviors. More information on the survey tool is available at https://teamsintandem.com.

Students in the course were required to complete the End of Term Survey for course credit. There are nine survey questions that address positive teamwork behaviors for individual students to reflect on their own and team members' contribution to the project. In this study, we focused on the four questions shown below in Table 1.

Table 1: Survey questions used for assessment. "$TeamMember" was replaced with each individual's name.

Where would you place $TeamMember on each of these scales? [slider with 9 stops]			
[Enacting Ideas]	Our project didn't include many ideas from $TeamMember.	←→	Many of $TeamMember's ideas were used in our project.
[Listening]	$TeamMember was more likely to speak up with their own ideas than to listen and encourage others.	←→	$TeamMember was a great listener who helped encourage our other members.
[Reliability]	$TeamMember was often late to meetings, was distracted while we were collaborating, or was generally unreliable.	←→	$TeamMember always showed up, responded to messages, and was generally reliable.
[Valuable]	$TeamMember was still gaining the skills needed for our project.	←→	The skills $TeamMember brings to the team are incredibly valuable

By the end of the semester, every student should ideally have had the opportunity to evenly divide the work of the term project. This research focuses on the End of Term Survey data because it gives a more holistic representation of how each peer evaluated one another throughout the whole semester, as opposed to Mid-term Survey data that evaluates students for only part of the semester. In using the End of Term Survey, each student identifies the role(s) they believe they played on the team over the course of the full semester.

Our hypotheses are the following:

1. Men will be rated higher than women on "enacting ideas" using peer ratings.
2. Women will be rated higher than men on "listening" using peer ratings.
3. Women will be rated higher than men on "reliability" using peer ratings.
4. Men will be rated higher than women on "valuable ideas" using peer rankings.

In a study on the influence of gender on roles in teams, men were shown to present a larger ratio of technical slides in a presentation when compared to women [2]. In this same study, a solo woman on a team of men commented that she acted "sort of like a mom" [2]. Because these findings suggest that men tend to assume different roles than women on teams, the hypotheses above are predictions based on connections between the characteristics from these roles; men tend to assume more technical roles associated with "ideas", and women tend to assume more non-technical roles associated with "listening" and "reliability".

Using SPSS, responses that men and women received were compared. There are two between subject measures (gender: men and women). Gender data is pulled from student's official university record; no students identified as non-cisgender in this data set. We also performed an ANOVA with follow-up t-tests to determine the statistical significance of the results.

Results

The End of Term Survey was completed by 118 of 119 total students from Fall and Winter 2019 semesters for a response rate of 99.2%. Of these 118 students, 33 identified as women and 85 identified as men. The summary statistics for the responses are shown in Table 2. As described in the Methods section above, ratings were given on a slider with nine stops. The stops correspond to integers, with the lowest stop corresponding to one and the highest stop corresponding to nine. Women received higher average ratings than men on all of the behavioral indicators we selected to test.

The results of the ANOVA and two-tailed t-tests are shown in Table 3. The difference in mean ratings between genders were statistically significant for both "This team member was a great listener" and "This team member... was generally reliable" ($p = 0.001$ and $p < 0.0005$ respectively). The average rating for women is greater than men for both parameters, and in a follow up t-test, the confidence intervals for both are greater than zero. This suggests that the true mean rating for women is higher than that for men.

It is worth noting that responses for men had larger variance than responses for women (Table 2). To visualize this, we created a series of histograms, overlaying received ratings by gender.

Table 2: The summary statistics shown below indicate that women scored higher on average than men in all categories.

Parameter	Gender	Mean	Std. Deviation	Std Error Mean	P value
This team member introduced ideas which were enacted	Women	7.23	1.26	0.220	0.057
	Men	6.69	1.60	0.173	
This team member was a great listener and encouraged the team	Women	7.64	1.07	0.187	0.001
	Men	6.59	1.62	0.176	
This team member showed up, responded, and was generally reliable	Women	8.13	0.68	0.118	0.000
	Men	6.99	1.89	0.205	
The skills this team member brought to the team are valuable	Women	7.65	1.14	0.198	0.193
	Men	7.32	1.30	0.141	

Table 3: A two-sided t-test was performed for each parameter.

Behavioral Parameter	t	df	sig (2-tailed)	Mean Difference	Std Error Difference	Lower Bound (95% confidence interval)	Upper Bound (95% confidence interval)
Ideas Enacted	1.94	73.6	0.057	0.54	0.280	-0.015	1.10
Listened Well	3.45	116	0.001	1.05	0.306	0.448	1.66
Was Reliable	4.81	115.6	0.000	1.14	0.236	0.669	1.61
Was Valuable	1.31	116	0.193	0.34	0.258	-0.173	0.85

Because our sample was uneven (with more men than women), we used proportion rather than raw values, as seen in Figures 1 – 4.

Discussion

Women were rated higher than men in all four behavioral indicators. Of the four indicators, we predicted that men would be rated higher than women on enacting ideas and on valuable ideas. On average, men were rated 8% lower for enacting ideas and 5% lower for valuable ideas when compared to women. While men were rated lower for these two indicators, it is important to note that the p-values for "Ideas Enacted" and "Was Valuable" were 0.057 and 0.193 respectively. At 95% confidence, the difference in means was not statistically significant because $p > 0.050$. Therefore, although men received a lower average rating than women, the differences in average ratings for these two attributes are not statistically significant between men and women.

We predicted that women would be rated higher than men on reliability and listening; women received scores 16% higher than men in these two behavioral parameters. The percent difference of women's mean score compared to men's mean score is 16% for listening and is 16% for reliability. The difference in average ratings between men and women was statistically different ($p = 0.001$ for Listened Well and $p < 0.0005$ for was Reliable). We are 95% confident that the true difference in means is within the lower and upper bound listed in Table 3 above. For "Listened Well" and "Was Reliable", this interval is above zero, which supports our hypothesis that women are rated higher for these two attributes.

The mean total behavioral score for women is 11% greater than that of men. The differences in

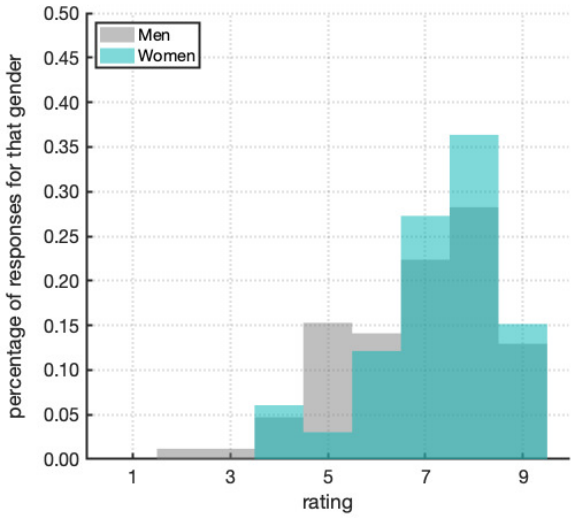

Figure 1: This Team Member's Ideas were Enacted Histogram

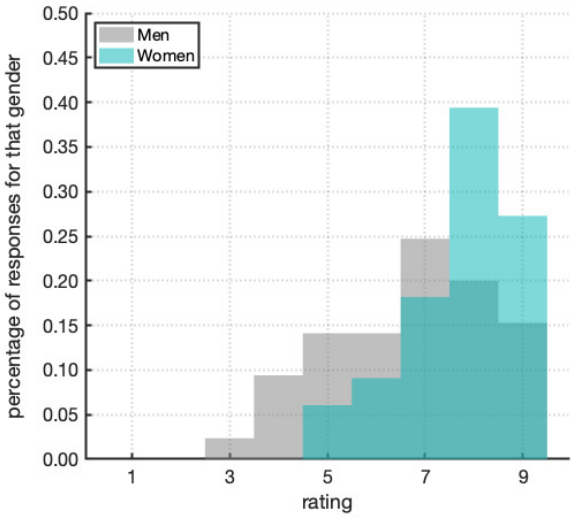

Figure 2: This Team Member Listened Well Histogram

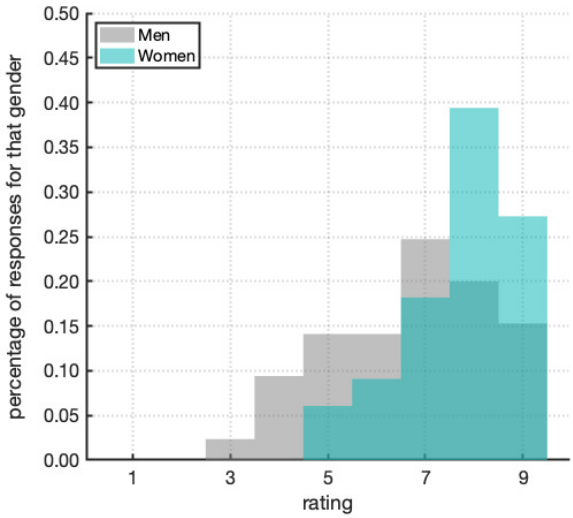

Figure 3: This Team Member was Reliable Histogram

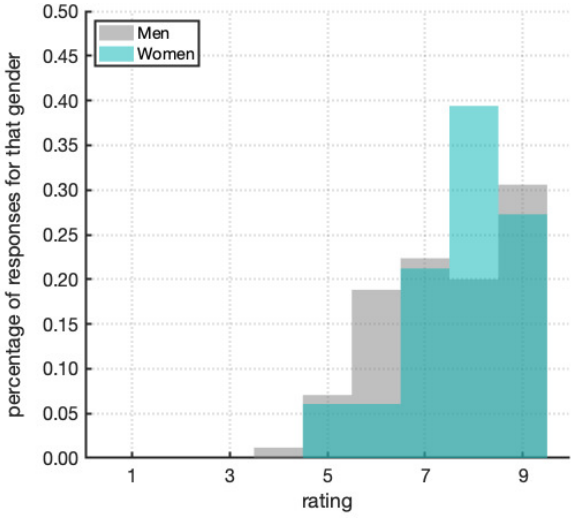

Figure 4: This Team Member Contributed Valuable Ideas Histogram

mean in these values, and in all these values, seem to come from the bottom half of the population. This suggests that a sub-population of men do not expend full effort on teams. While it is not the case that only men could make the decision to slack on a team, it may be the case that women are less likely to, perhaps because of preexisting differences in motivation of men and women enrolling in the College of Engineering, or perhaps because women feel extra pressure as a minority group in the engineering class. Additional research is needed to determine whether this finding holds up statistically and across other contexts, as well as to differentiate other competing

explanations.

The results seem to suggest that it is the left tail that differs between men and women in our sample, and this tail is evident even for behavioral indicators that weren't deemed significant by a conventional test for mean differences. We interpret this finding as suggesting that a small subset of the men in our sample are rated lower than classmates on idea enaction, listening, and reliability; these are perhaps the "social loafers" addressed in teamwork literature. There is not a corresponding group of social-loafing women in our particular sample.

Limitations

A number of factors were beyond our control in this study: an important issue is whether or not students completed the surveys honestly. While the survey is confidential, this is a class for students who are surrounded by their peers and friends. However, lack of honesty might be assumed to affect ratings of men and women similarly. It is also important to consider that there may be bias in ratings by gender. There may be a difference between men rating men, men rating women, women rating men, and women rating women. The size of this sample did not allow us to investigate this further, but it would be an interesting project for future work. This study also does not take personal biases into account. Some students may have a strained relationship, while others may be best friends; these personal relationships are outside the scope of the survey but may influence ratings. Finally, and probably most importantly, this study was conducted in a single class over two semesters. The sample size is relatively small, and students who enroll in this class and at this university cannot be expected to represent engineering students across the US or across the world. Care should be taken when applying these results to other settings.

Conclusions and Recommendations

Peer assessments have been implemented within our course in order to combat social loafing. Student's ratings of others may be impacted by their teammates' identity characteristics [8, 9, 13, 14, 15, 16, 17]. In our study, we examined the role gender plays on peer-reported teamwork behaviors in the class assessed end of term survey. We focused on the individual's contribution of enacted ideas, contribution of valuable ideas, reliability, and listening skills.

The data set is from two semesters of a team-based, project-based introductory engineering course with a total of 118 student responses analyzed (N = 118). There were 33 women and 85 men who completed the survey. Women were rated higher than men on all of the behavior indicators we selected to test. The average rating of these four parameters was 11% greater for women than for men. We predicted that men would be rated higher than women on enacted ideas and valuable ideas. Women were rated 8% and 5% higher respectively, although these differences in average scores were statistically insignificant. We predicted that women would be rated higher than men on reliability and listening; women received scores more than 15% higher than men in these two behavioral parameters.

These findings lead to several new potential research questions for the future:

- Do these findings differ, if at all, when disaggregated by multiple social groups (race/ethnicity, gender, sexual orientation, etc.)? Answering this question would likely

require a true longitudinal study and/or investigation in other contexts to produce enough data for each of these sub-groups.

- Do these findings differ, if at all, when compared across other contexts, such as amongst students completing a similar project at a different university, or amongst students completing a project in a different discipline, such as in an introductory nursing course?
- Is there a significant difference between ratings of the same gender? For example, do men rate other men differently than they rate women?

Acknowledgement

This research has been determined exempt from human subjects control under exemption #1 of the 45 CFR 46.101.(b) by the U-M Institutional Research Board (HUM00135376).

References

[1] B. Linder, M. Somerville, Ö. Eris, and N. Tatar. Work in progress—taking one for the team: Goal orientation and gender-correlated task division. *IEEE Frontiers in Education Conference*, pages F4H–1, Oct. 2010.

[2] L. A. Meadows and D. Sekaquaptewa. The influence of gender stereotypes on role adoption in student teams. *Proc. 120th ASEE Annual Conf. Exposition*, pages 1–16, Oct. 2013.

[3] E. A. Strehl and R. Fowler. Experimental evidence regarding gendered task allocation on teams. *Proc. 126th ASEE Annual Conf. Exposition*, pages 1–14, Oct. 2019.

[4] E. Scanlon. How gender influences learners working collaboratively with science simulations. *Learning and Instruction*, 10(6):463–481, 2000.

[5] R. Eggert, A. Joshi, S. Mehrotra, Y. V. Zastavker, and V. Darer. Using discourse analysis to understand failure modes of undergraduate engineering teams. *IEEE Frontiers in Education Conference Proceedings*, pages 1–5, Oct. 2014.

[6] I. Bohnet, A. Van Geen, and M. Bazerman. When performance trumps gender bias: Joint vs. separate evaluation. management. *Management Science*, 62(5):1225–1234, 2015.

[7] J. Stonewall, M. C. Dorneich, J. Rongerude, and C. Dorius. A review of bias in peer assessment. *In Collaborative Network for Engineering and Computing Diversity (CoNECD) Conference*, Apr. 2018.

[8] J. B. Main and M. Sanchez-Pena. Student evaluations of team members: Is there gender bias? *IEEE Frontiers in Education Conference Proceedings*, pages 1–6, Oct. 2015.

[9] P. M. Ostafichuk, A. d'Entremont, N. Shirzad, and J. Sibley. Gender and personality type influence in peer evaluation. *Proc. of ASEE 2015 Conference*, pages 15–17, 2015.

[10] Borrego M., Karlin J., McNair L.D., and Beddoes K. Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams: A research review. *Journal of Engineering Education*, Nov. 2013.

[11] B. Oakley, R. M. Felder, R. Brent, and I. Elhajj. Turning student groups into effective teams. *Journal of Student Centered Learning*, 2(1):9–34, Jan. 2004.

[12] D. Davis, M. S. Trevisan, H. Davis, R. Gerlick, J. Mccormack, Thompson P. Beyerlein, S., S. Howe, P. Leiffer, and P. Brackin. Assessing team member citizenship in capstone engineering design courses. *International Journal of Engineering Education*, 26(4):771–783, 2010.

[13] R. M. Marra, K. A. Rodgers, D. Shen, and B. Bogue. Women engineering students and self efficacy: A multi-year, multi-institution study of women engineering self-efficacy. *Journal of Engineering Education*, pages 27–38, Jan. 2009.

[14] M. A. Hutchison, D. K. Follman, M. Sumpter, and G. M. Bodner. Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, pages 39–47, Jan. 2006.

[15] G. E. Okudan, D. Horner, B. Bogue, and R. Devon. An investigation of gender composition on integrated project team performance: Part iii. *Proc. of the 2002 American Society for Engineering Education Annual Conference & Exposition*, 2002.

[16] S. Torres-Guijarro and M. Bengoechea. Gender differential in self-assessment: a fact neglected in higher education peer and self-assessment techniques. *Higher Education Research & Development*, pages 1072–1084, Dec. 2016.

[17] R. Fowler. Demographic effects on student-reported satisfaction with teams and teammates in a first-year, team-based, problem-based course. *Proc. of ASEE 2016 Conference*, June 2016.

[18] N.C.T. Van Tyne, C.J. Van Tyne, and K.C.T. Van Tyne. Gender differences in individual and teammate performance evaluations by students on engineering design teams. *American Society of Engineering Education Annual Conference*, Jun. 2011.

[19] R. Tucker. The architecture of peer assessment: do academically successful students make good teammates in design assignments? *Assessment & Evaluation in Higher Education*, 38, Aug. 2013.

[20] J.A. Coller, M.P. Su, L.K. Alford, S. Sheffield, and R. Fowler. Assessment of peer mentoring of teams in a first-year design-build-test-communicate class. In *2017 ASEE Annual Conference & Exposition*, Columbus, Ohio, June 2017. ASEE Conferences. https://peer.asee.org/27636.