



## **Demographic Effects on Student-Reported Satisfaction with Teams and Team-mates in a First-Year, Team-Based, Problem-Based Course**

**Dr. Robin Fowler, University of Michigan**

Robin Fowler is a lecturer in the Program in Technical Communication at the University of Michigan.

## **Demographic Effects on Student-Reported Satisfaction with Teams and Teammates in a First-Year Team-Based, Problem-Based Course**

### **Abstract**

This work-in-progress reports the effect of student gender and team gender make-up on team satisfaction and student assessments of team contributions. In first year team-based student design courses, instructors use student self- and peer-assessment information to gauge team functioning and even to affect student project scores. However, students' identity characteristics, such as their gender and race, may impact the scores they receive from others as well as the scores they assign. The poster will also describe the creation of and results from a learning-analytics style investigation of the researcher's own student team assessment data, and the poster presentation will allow others to query the data set with their own questions. The final data set includes assessment information from 620 first-year engineering students working in 132 teams of 4 or 5 to design, build, test, and report in varying open-ended project contexts (from 11 sections of the course taught between Fall 2011 and Winter 2015). The study is presented as a poster because the researcher is eager to connect with others at ASEE to consider new questions of interest. The data set includes demographic data (gender, race, TOEFL score when applicable, SAT/ACT scores, first year GPA, final course grade), as well as team assessments (student self-assessments and assessments of teammates, team report scores, and team satisfaction ratings).

In this large data set, many comparisons were significant. Findings of interest regarding team satisfaction included a pattern of satisfaction by team gender breakdown. Teams with two or more women were much less happy (mean satisfaction of 3.79 on a 5-pt scale) compared to students on teams of just men (4.25) and teams with one woman (4.32). This finding is in contrast to a recommendation common in the literature to avoid stranding women on teams. Teams that included international students were less happy (mean satisfaction of 3.68) compared to teams that did not (4.24). Both of these patterns had medium-sized effect sizes (Cohen's  $d$  was 0.67 for the finding by gender and 0.74 for the finding with international students.) There was no difference in the mean satisfaction of students on teams including or not including under-represented minority students (4.16 for teams with URM students, 4.17 for teams without).

Student ratings of peers were examined by gender. CATME asks students to rate each other on five scales for behaviors related to teamwork. The largest effect was found for "having related knowledge, skills, and abilities." Women were rated lower on this category (by both men and women), and men rated women lower on this category than women rated women (Cohen's  $d$  was 0.69 here). Women were rated higher than men on all other categories (by both men and women), but with much smaller effect sizes (Cohen's  $d$  ranging from 0.12 to 0.32).

This project is proposed as a work-in-progress because there are many limitations related to using my own data for this study. This exploratory study of existing data has uncovered interesting patterns in students' assessments of team satisfaction and contributions that should be investigated further, perhaps through follow-up interviews or through an experimental design. I look forward to talking with conference attendees and getting input before moving forward with such a study. I am also eager to have conversations about what the findings might suggest for how teams are formed and, more importantly, how they are supported and assessed.

## Introduction and Motivation

Engineering classes make use of team-based, project-based pedagogies throughout the curriculum, and such projects are particularly common in first year courses<sup>1</sup>. Learning goals for such projects include teamwork, design, and communication<sup>1</sup>.

Instructors using this pedagogy want the project to go smoothly and efficiently, with teams working well together and all students contributing their fair share, a concept known in the psychology literature as “team effectiveness<sup>1</sup>.” We don’t always know how to make effective teams happen, though.

Instructors work to increase team effectiveness in their team-based, problem-based design courses using many pedagogies. One strategy that is the focus of many studies is on how teams are assigned. Suggestions common in the literature, which sometimes are at odds, include making heterogeneous teams (by gender, by skill level, by race), making homogenous teams (by gender, by skill level, by goals for the project), and avoiding stranding minority students (gender and race, specifically)<sup>1,2,3</sup>.

Another concern of instructors is social loafing; we don’t want students to shirk their project responsibilities yet earn the same scores as their hard-working peers. Social loafing is a particular problem in first year projects, when each students’ contribution is generally not unique because the students typically don’t have the backgrounds to allow for interdisciplinarity<sup>1</sup>. The tasks are not differentiated for students because they are all assumed to have the same set of skills.

It is hard to measure effective teamwork<sup>4</sup>, but instructors adopt many different methods of assessment to try to capture this information. Many of these methods rely on student assessments of their own and peer contributions. In general, when peer assessments might identify social loafers, social loafing decreases or even disappears<sup>1</sup>.

In an attempt to preempt social loafing, instructors conduct student assessments and sometimes even use this information to affect student grades<sup>3</sup>. I am guilty of such a practice in my classrooms. However, studies of student self- and peer- assessments often find systematic over-valuing of one’s own work and systematic undervaluing of the documentation steps in an engineering project<sup>2</sup>. More importantly, the use of such assessments with real-world (grade) consequences also makes it critical that we consider whether students’ identity characteristics are affecting their self-reports or assessments of and from teammates.

There is precedent to believe that self-assessments might be affected by student identity characteristics such as gender. Women in first-year engineering courses report lower engineering self-efficacy on a variety of instruments<sup>5,6</sup>, but it is not known whether such responses will transfer to a team assessment report like is conducted in team-based learning contexts. Some studies have found gender effects in team assessment (and team grades), including that women on engineering teams are more critical in their assessments of other women<sup>7</sup>. Okudan and colleagues also conclude that homogeneous teams (by gender) earn higher project scores than heterogeneous teams, but the result they report is not statistically significant<sup>7</sup>.

Using data similar to the data reported here, Van Tyne, Van Tyne, and Van Tyne<sup>8</sup> investigated actual student team assessments (self and peer) from students who completed a first-year engineering design course. They found no significant differences by gender in either self-

assessment or peer-awarded assessment, and attributed this finding to the support available at a small school focused only on engineering. This study adds to this result with a comparison group at a large public institution with a large proportion of non-engineering students.

## Methods

This data set was assembled from demographic information and team assessments in a first-year engineering and communications course at a highly selective large public university in the Midwest. The teams investigated were teams of four or five students working on open-ended design-build-test-report projects in 11 classes over 8 semesters of 4 months each (2011 to 2015). The first-year course at this university offers varying “flavors” (biomedical, aerospace, etc.) with differing final projects (Meadows, Fowler, and Hildinger<sup>9</sup> provide a more detailed description of the philosophy of the course, as well as historical enrollment information). The particular final projects in this data set ranged from seven to ten weeks within the semester and included:

- Battery-Powered Submarine. Designing and building a battery-powered submarine (using provided materials) operated from land via tether, competing in a series of speed and maneuverability challenges with the submarine, and reporting about the design and its performance via oral and written reports.
- Wind Turbine. Designing and building a small-scale wind turbine with a \$200 budget and a given motor, powering a cell phone with the turbine, and reporting about the design and its scaled-up performance via oral and written reports.
- Bioreactor or anaerobic digester. Designing and building a bioreactor or anaerobic digester with provided materials and a small budget, collecting performance information, and presenting the resulting design idea, orally and in written form, to a fictitious client who is building a sustainable apartment complex.
- Catchment and filtration system. Designing and building a graywater catchment and filtration system, testing it to determine its filtration performance, and reporting about its performance via oral and written reports.

In all classes, instructors attempted to design a project scope appropriate for the team size (four or five students). In all classes, teams physically built their products and systems using various machines (power saws, sanders, etc.). In most classes, students created 3D models of the design idea and updated the model after the item was built. In all classes, grades were awarded primarily on the resulting communications created rather than on design performance. However, it is difficult to separate these ideas, as it is easier to write a compelling report about a design that was well-conceived and that performed well.

### *Team Formation*

Teams were formed by instructors, based when possible on suggestions common in the literature (avoid isolating females and under-represented minority students; distribute high-achieving students when this information is available; consider convenience factors such as student availability and dormitory location<sup>4,10</sup>). Importantly, because of the small size of the laboratory sections from which the teams are formed (capped at 20 students), these competing suggestions can never all be satisfied. Sometimes women and URM students are isolated on teams even though it is not considered best practice.

## Team Assessments

At the end of the seven- to ten-week project, students completed teammate ratings via the Comprehensive Assessment of Team Member Effectiveness (CATME) system<sup>11</sup>. (In the middle of the project, they completed a similar assessment.) They rated themselves as well as each teammate on five behaviorally-anchored items:

1. contributing to the team's work,
2. interacting with teammates,
3. keeping the team on track,
4. expecting quality, and
5. having related knowledge, skills, and abilities.

Figure 1 presents a screenshot of one of the five behaviorally-anchored scales.

|                       |                       |                       |                       | Description of Rating  |
|-----------------------|-----------------------|-----------------------|-----------------------|--|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <ul style="list-style-type: none"><li>• Demonstrates the knowledge, skills, and abilities to do excellent work.</li><li>• Acquires new knowledge or skills to improve the team's performance.</li><li>• Able to perform the role of any team member if necessary.</li></ul>                          |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Demonstrates behaviors described immediately above and below.  |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <ul style="list-style-type: none"><li>• Demonstrates sufficient knowledge, skills, and abilities to contribute to the team's work.</li><li>• Acquires knowledge or skills as needed to meet requirements.</li><li>• Able to perform some of the tasks normally done by other team members.</li></ul> |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Demonstrates behaviors described immediately above and below.  |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <ul style="list-style-type: none"><li>• Missing basic qualifications needed to be a member of the team.</li><li>• Unable or unwilling to develop knowledge or skills to contribute to the team.</li><li>• Unable to perform any of the duties of other team members.</li></ul>                       |

Figure 1. Example rating page for sample team on the characteristic “having related knowledge, skills, and abilities.” Students rated themselves and their teammates on this scale.

Note that students are also asked to rate themselves. The scale gives descriptors so that students aren't just assigning a value, as the temptation might be to assign all students “perfect” scores.

When students see the assessment information after it is completed, they see aggregate information. For each of the five behaviors, they can compare their own self-ratings to the average score they received from their teammates, as well as to the average score on their team. See Figure 2 for a screenshot of the results as they are presented to students.

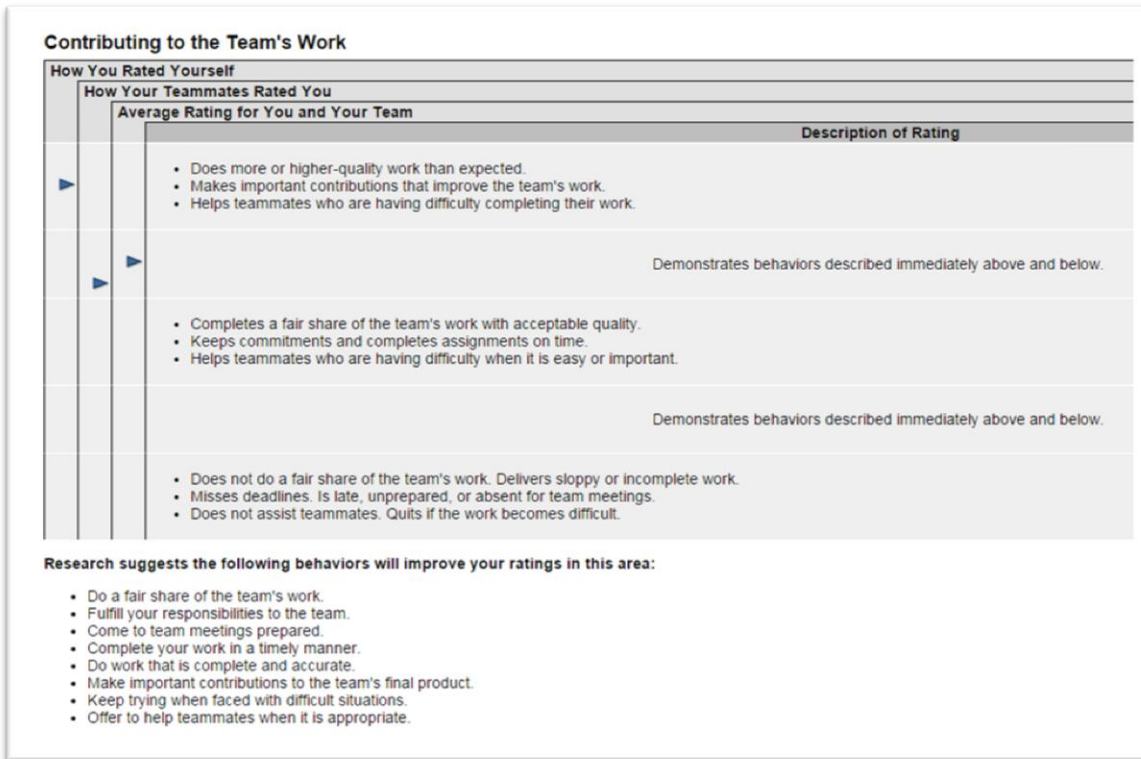


Figure 2. Screenshot of student results page from CATME, which shows student's self-rating, average rating received from peers, and the team's average, for the "contributing to the team's work" behavioral scale.

Finally, the instrument also asks students to rate three "team satisfaction" statements on a Likert scale (1 = not at all true and 5 = very true):

- I am satisfied with my present teammates.
- I am pleased with the way my teammates and I work together.
- I am very satisfied with working with this team.

### *Creation of the Data Set*

The data set includes information from 620 students (514 men and 106 women) on 132 teams of 4 or 5 students from 11 sections of an introductory engineering course taught between Fall 2011 and Winter 2015. Student assessment scores from CATME (self-assessments and peer assessments, from male teammates and from female teammates) and team satisfaction scores from CATME were consolidated with Registrar-provided demographic information (race/ethnicity, TOEFL score, gender, final course grade, GPA in first-year coursework, SAT/ACT score), and the resulting spreadsheet was anonymized (with a separate key).

### **Analysis**

**Peer ratings by gender.** Peer ratings were compared by gender. Figures 3 through 7 show scores on CATME peer assessments, by gender. The error bars shown are the 95% confidence intervals, using a students' t-distribution because the population parameters are unknown. As can be seen in the figures, the gender of the person being scored and the gender of the person doing the scoring make a difference.

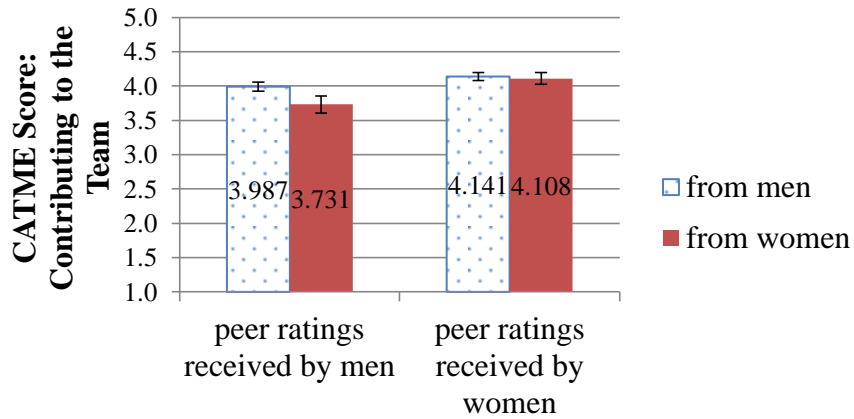


Figure 3. “Contributing to the team” scores, by gender (of receiver and scorer)

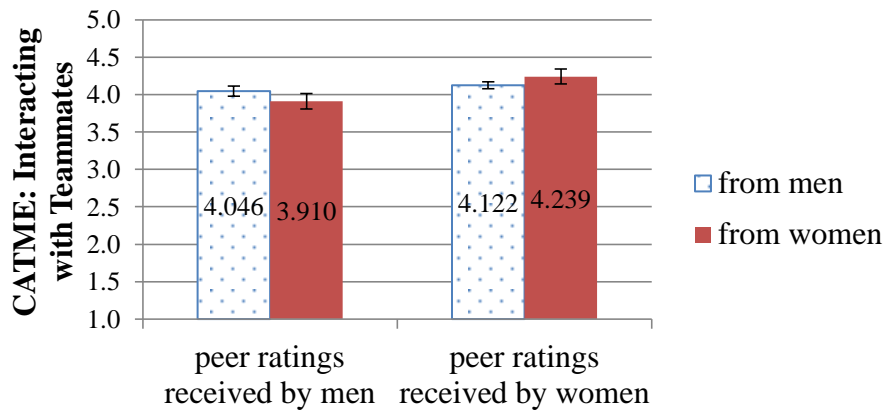


Figure 4. “Interacting with Teammates” scores, by gender (of receiver and scorer)

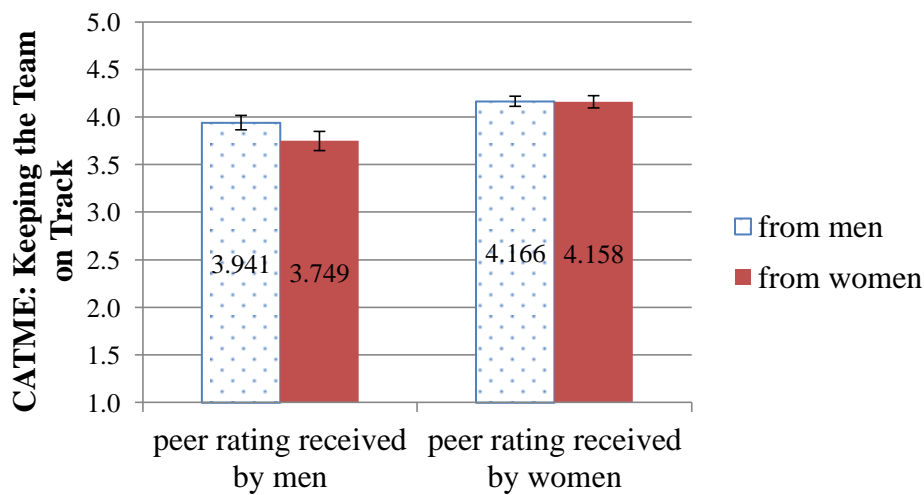


Figure 5. “Keeping the team on track” scores, by gender (of receiver and scorer)

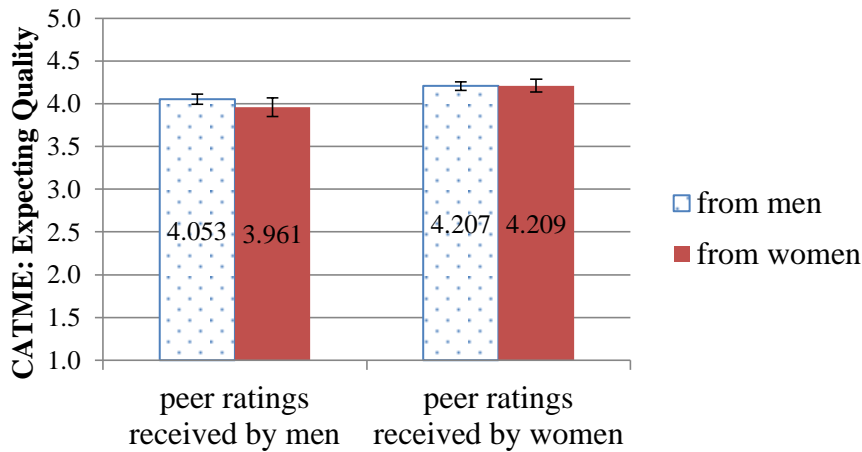


Figure 6. “Expecting quality” scores, by gender (of receiver and scorer)

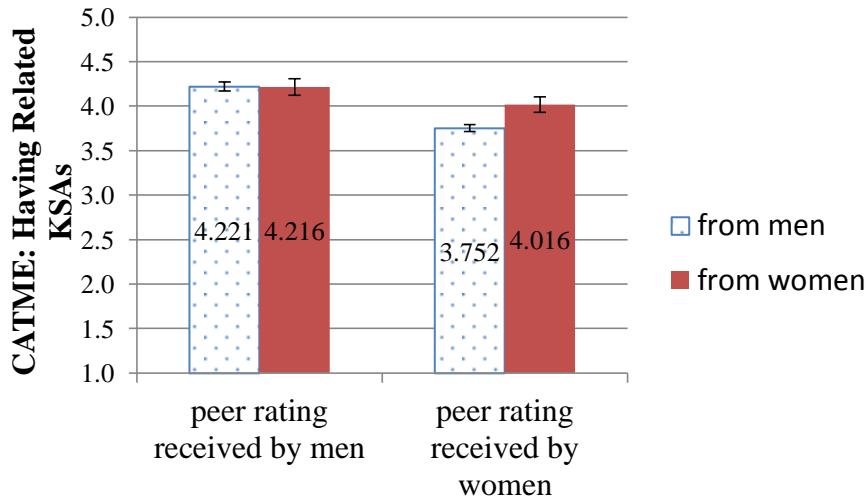


Figure 7. “Having related knowledge, skills, and abilities” scores, by gender (of receiver and scorer)

On most CATME constructs (contributing to the team, interacting with teammates, keeping the team on track, and expecting quality), women are scored higher than men, and women score men lower than men score men. The effect sizes for these findings are mostly small, with Cohen’s *d* values ranging from 0.12 to 0.32. These values mean that the gender difference in average received values range from 0.12 to 0.32 standard deviations.

The finding with the largest effect size is that men are scored higher for “having related knowledge, skills, and abilities,” especially by men. The effect size for this finding is medium-large, with a Cohen’s *d* of 0.69, meaning the average score men receive is 0.69 standard deviations higher than the average score women receive.



**Team satisfaction.** Gender breakdown, team inclusion of URM students and international students, and performance metrics (team report grade and student first-year GPA) were investigated to see relationships between these factors and team satisfaction. Figures 8-10 present these results. Error bars are 95% confidence intervals, again using a students' t-distribution because population parameters are unknown. Figure 8 presents team satisfaction values for teams with differing gender breakdowns.

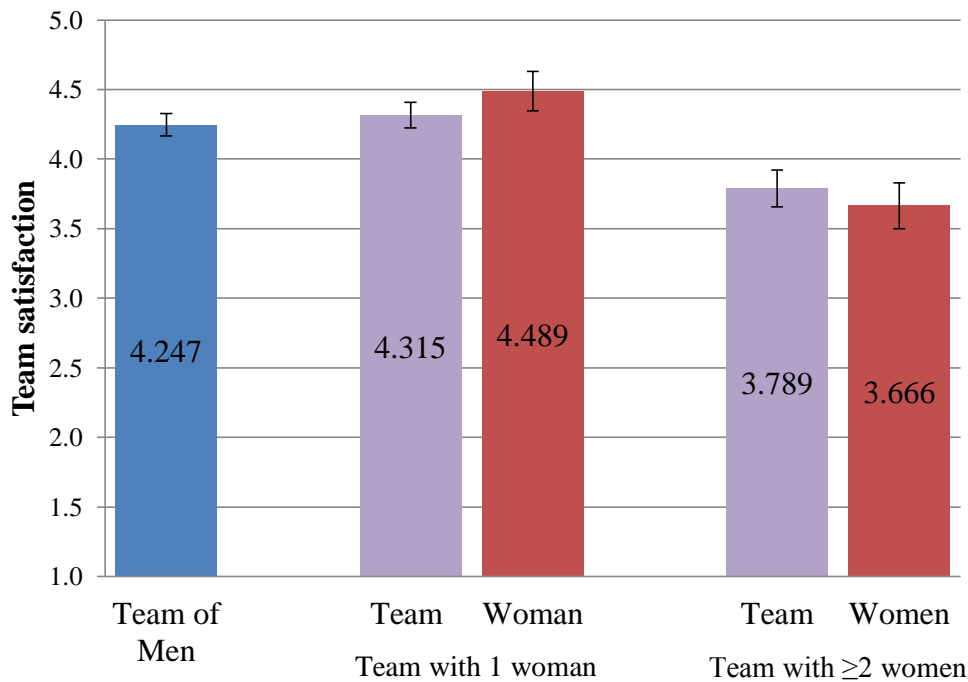


Figure 8. Team satisfaction for teams of various gender breakdowns. The grand mean is 4.167.

As can be seen in Figure 8, teams with two or more women are less happy than teams of all men or teams with a single woman. Women's satisfaction levels seem to follow the team satisfaction patterns. The grand mean was 4.167. Teams of men are not significantly different from this value. Teams with a single woman report more satisfaction than this average, and teams with two or more women report less satisfaction. The effect size here is medium-large: Cohen's  $d$  is 0.67.

Figure 9 shows team satisfaction on teams with under-represented minority students and international students. Teams with international students were split into teams with a single international student and teams with 2 or more international students to explore whether the practice of avoiding isolating students was related to team satisfaction. The splitting of teams with URM students into teams with a single URM student and teams with multiple URM students was not done because of small sample size concerns.

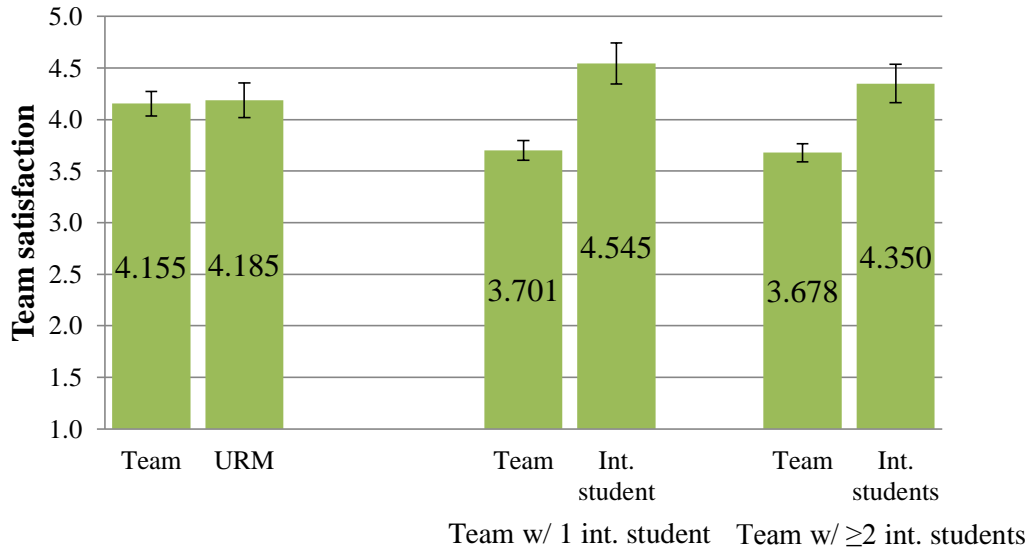


Figure 9. Team satisfaction for teams with URM students and international students. The grand mean is 4.167.

The grand mean (overall average satisfaction) was 4.167. Teams with under-represented minority students, and the URM students themselves, are not significantly different from this. Teams with international students are significantly less satisfied than their peers, and the international students on these teams are significantly more satisfied. This pattern held whether the international students were stranded on teams or were paired on teams. The effect size here was medium-large: Cohen's  $d$  is 0.74.

Finally, academic performance characteristics were considered for their relationship to team satisfaction. One team performance metric that was readily accessible was the combined score from final written and oral reports. Teams were split into bottom, middle, and top thirds on this metric, and bottom and top thirds were compared. One student performance metric that was both readily accessible and also not conflated with team performance was GPA in the first year (omitting the class this data comes from). Students were split into bottom, middle, and top thirds on this metric, and bottom and top thirds were compared. The comparisons for both performance measures are shown in Figure 10.

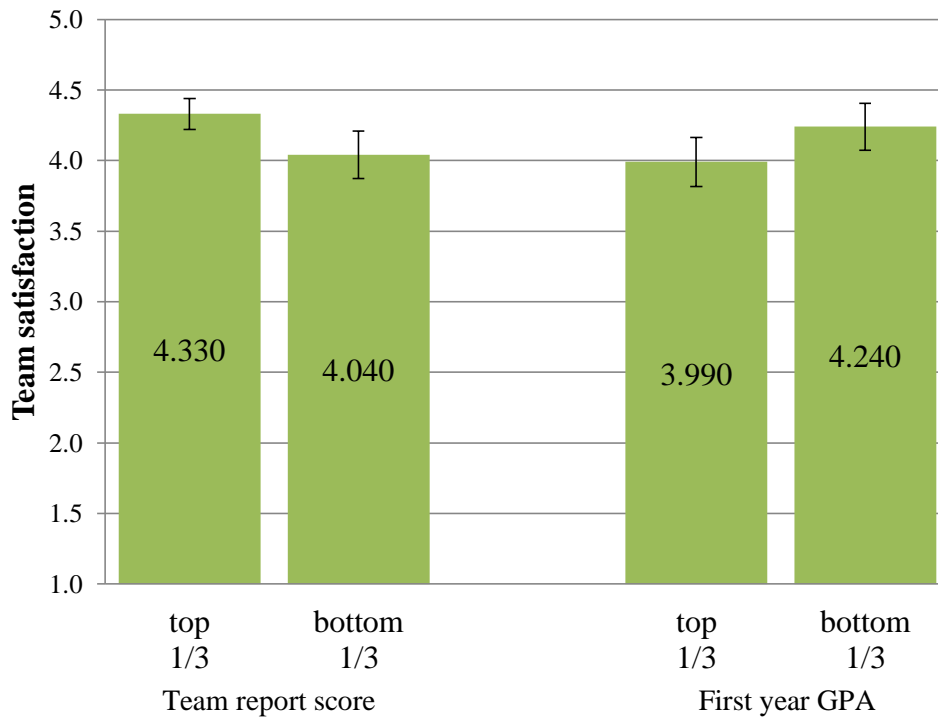


Figure 10. Team satisfaction, by team performance (report score) and individual performance (First year GPA).

Unsurprisingly, teams that score higher are more satisfied with the team than teams that score lower, but the differences are not significant. Perhaps also unsurprisingly, students who individually score in the upper third for first-year GPA are less happy than students who score in the bottom third. Students with high GPAs may find teamwork more stressful as they are concerned that poor group performance may hurt their course grade (anecdotal evidence suggests this). This pattern does appear in the data, but again, the difference is not significant.

### Significance

This proposed study is only an exploratory study. The findings can inform my teaching strategies and directions for future studies, but they should not be over-interpreted. There was no attempt initially to create random or counter-balanced teams; in fact, students were assigned with instructor knowledge that should influence findings from the data set. For example, when instructors were deciding whether to prioritize “don’t strand a woman” or “put students together who live near each other,” we sometimes used our own perceptions of the particular female student’s self-confidence in making this decision. Therefore, any findings should be interpreted with caution.

Three significant findings warrant discussion here, though. The first is the different scoring by gender. Women earn higher scores on most of the CATME contribution measures than men. In contrast, women earn lower scores for “having related knowledge, skills, and abilities” than men. These differences in received means interact with the gender of the scorer: women score men lower on most of the contribution categories than men score other men. Men score women lower on “having related knowledge, skills, and abilities” than women score other women. Importantly,

the existence of mean differences in ratings by student gender or other identity factors does not necessarily mean the ratings are biased. I have no measure to show that gender and other demographic factors are unrelated to performance in the various CATME categories. Further research should look at ways of controlling for or measuring different contributions and background knowledge so that differences in mean peer ratings can be better interpreted. Follow-up interviews could investigate how students are interpreting the CATME language to see whether the gender differences could be understood through this lens as well.

The finding of differences in CATME scores based on gender is not surprising, given the findings by other studies of differences in terms of peer ratings of classmate knowledge in engineering and broader STEM contexts<sup>12</sup>. Given this knowledge, and my finding that there are differences in my own students' scores, I will reconsider my use of peer feedback to scale project scores. I have always thought that I am rewarding students for good team contributions with this practice, but I need to be careful that I am not allowing students' identity characteristics to affect their grades.

A second important finding is the difference in team satisfaction by the gender breakdown of the team. I have followed advice<sup>4</sup> to avoid stranding women and other minority groups when possible on my teams, which means I attempt to form teams that are all men or that include two or more women. I am surprised to see that the teams with two or more women are less happy than the teams with only a single woman. Team satisfaction is not the only or even the primary goal, but this difference in team satisfaction is a concern for me. One possibility is that stranded women are taking particular roles on teams, maybe even ones they do not want, in ways that improve team performance and harmony (and therefore satisfaction). Teams where women are not stranded may be dividing work in other ways. I want to follow-up on this result of a difference in team satisfaction with interviews or focus groups to see what might be going on, so that I can decide how I want to address this difference in my class.

Finally, I am concerned about the lower team satisfaction on teams with international students. Again, I want to follow up on this result with focus groups or interviews, particularly with students who have been on teams with international students. It is possible that international students excel in particular team roles but do not contribute to all parts of a project equally, or perhaps cultural differences in team expectations or communication result in lowered satisfaction for their team. If I understand the reasons for this pattern of lower team satisfaction, I can intervene by better preparing my students for cross-cultural team work.

The real contribution of this poster to the ASEE community is the sharing of the process. This is an example of learning analytics and of scholarship of teaching and learning used in my own classroom. I hope that the ASEE community at the poster session will pose interesting questions, and that I might even find potential collaborators who would like to explore these questions with me further. I look forward to talking with conference attendees and getting input before moving forward with a follow-up study. I am also eager to have conversations about what these findings suggest for how teams are formed and, more importantly, how they are supported.

1. Borrego, M., Karlin, J., McNair, L. D., & Beddoes, K. (2013). Team effectiveness theory from Industrial and Organizational Psychology applied to engineering student project teams: A research review. *Journal of Engineering Education*, 102, 472-512.
2. Davis, D., Trevisan, M., Gerlick, R., Davis, H., McCormack, J., Beyerlein, S., ... Brackin, P. (2010). Assessing team member citizenship in capstone engineering design courses. *International Journal of Engineering Education*, 26(4), 771-783.
3. Schultz, T. (1998). Practical problems in organizing students into groups. In *Proceedings of Frontiers in Education Conference*. Tempe, AZ, USA.
4. Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, 2(1), 9-34.
5. Marra, R.M., Rodgers, K.A., Shen, D., Bogue, B. (2009). Women engineering students and self-efficacy: A multi-year, multi-institution study of women engineering student self-efficacy. *Journal of Engineering Education*, 98, 27-38.
6. Hutchison, M. A., Follman, D. K., Sumpter, M., Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, 95, 39-47.
7. Okudan, G.E., Horner, D., Bogue, B., & Devon, D. (2002). An investigation of gender composition on integrated project team performance: Part III. In *Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition*.
8. Van Tyne, N. C., Van Tyne, C. J., & Van Tyne, K. (2011). Gender differences in individual and teammate performance evaluations by students on engineering design teams. In *Proceedings of the 2011 American Society for Engineering Education Annual Conference and Exposition*.
9. Meadows, L. A., Fowler, R., & Hildinger, E. S. (2012, June). *Empowering Students with Choice in the First Year*. Paper presented at 2012 ASEE Annual Conference, San Antonio, Texas. <https://peer.asee.org/21282>.
10. Rosser, S. V. (1998). Group work in science, engineering, and mathematics: Consequences of ignoring gender and race. *College Teaching*, 46, 82-88.
11. Ohland, M. W., Loughry, M. L., Woehr, D. J., Finelli, C. J., Bullard, L. G., Felder, R. M., ... Schmucker, D. G. (2012). The Comprehensive Assessment of Team Member Effectiveness: Development of a behaviorally anchored rating scale for self and peer evaluation. *Academy of Management Learning & Education*, 11, 609-630.
12. Grunpan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE*, 11(2), e0148405. doi:10.1371/journal.pone.0148405.