Teammates Stabilize over Time in How They Evaluate Their Team Experiences

Rebecca L. Matz
University of Michigan
Ann Arbor, Michigan, USA
rlmatz@umich.edu

Albert M. Lee
University of Michigan
Ann Arbor, Michigan, USA
cholma@umich.edu

Robin R. Fowler
University of Michigan
Ann Arbor, Michigan, USA
rootsr@umich.edu

Caitlin Hayward
University of Michigan
Ann Arbor, Michigan, USA
cholma@umich.edu

ABSTRACT

It is difficult for instructors, and even students themselves, to become aware in real-time of inequitable behaviors occurring on student teams. Here, we explored a potential measure for inequitable teamwork drawing on data from a digital pedagogical tool designed to surface and disrupt such team behaviors. Students in a large, undergraduate business course completed seven surveys about team health (called team checks) at regular intervals throughout the term, providing information about team dynamics, contributions, and processes. The ways in which changes in students' scores from team check to team check compared to the median changes for their team were used to identify the proportions of teams with outlier student scores. The results show that for every team size and team check item, the proportion of teams with outliers at the end of the term was smaller than at the beginning of the semester, indicating stabilization in how teammates evaluated their team experiences. In all but two cases, outlying students were not disproportionately likely to identify with historically marginalized groups based on gender or race/ethnicity. Thus, we did not broadly identify teamwork inequities in this specific context, but the method provides a basis for future studies about inequitable team behavior.

CCS CONCEPTS

• **Applied computing** → **Collaborative learning**; **Interactive learning environments**.

KEYWORDS

Educational technology, equity, gender, groupwork, higher education, learning analytics, peer evaluation, race/ethnicity, teamwork, undergraduate education

ACM Reference Format:
Rebecca L. Matz, Albert M. Lee, Robin R. Fowler, and Caitlin Hayward. 2022. Teammates Stabilize over Time in How They Evaluate Their Team Experiences. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3506860.3506891

LAK22, March 21–25, 2022, Online, USA

ACM ISBN 978-1-4503-9573-1/22/03…$15.00
https://doi.org/10.1145/3506860.3506891

1 INTRODUCTION

Learning to work effectively in a team is important for preparing undergraduate students for graduate work and their careers [4, 14, 19, 21]. Regularly cited as key by employers [10], teamwork is a critical skill in professional environments as well as a useful pedagogy for teaching technical and other professional content [15, 22]. While teamwork is a common pedagogy, supporting inclusive teamwork remains challenging. Marginalization on teams has been linked to aspects of students' social identity, such as identifying as a woman or person of color [13, 18], and can take a variety of forms, such as certain students' ideas going unheard or certain students disproportionately completing particular tasks [8].

Particularly in introductory courses and at large colleges and universities, instructors often face the additional hurdle of having many students and student teams. With these compounding factors, it is difficult for instructors, and even students themselves, to become aware in real-time of inequitable behaviors occurring on student teams. It is more difficult still to know how to address such inequitable behaviors. The consequences of inequitable and ineffective teams are serious, leading to differences in mastery experiences and ultimately contributing to students leaving degree programs [12, 16, 24]. Lack of retention renders lost educational and career opportunities for individual students and results in a significant loss of diversity and creativity for the disciplines themselves [1, 23].

Much prior work has investigated gendered and racialized differences between students on project teams with data sources ranging from self-reported surveys and peer evaluations to activity logs, observations, and interviews. We draw especially from engineering education studies here because team- and project-based courses are typical throughout engineering curricula. In a recent review, for example, Beigpourian and Ohland [3] summarized gendered and racialized patterns with respect to different aspects of teamwork such as collaboration, communication, and leadership. In our own institutional context, we have studied the relationship between extroversion and willingness to speak up amongst teammates including how this changed over course of a semester [5], finding that both women and international students were likely to report more introversion than their respective comparative groups at the end of the semester and to show lower levels of voice safety and voice enactment.

2 TANDEM: A TOOL FOR IMPROVING TEAMWORK

Here, we investigate a potential measure for equitable teamwork relying on Tandem, a digital pedagogical tool designed to disrupt inequitable team behavior primarily through providing feedback to students and instructors about how teams are doing [7]. Developed at the University of Michigan (U-M), a large, primarily residential, and research-intensive university, the underlying framework of Tandem draws on the Michigan Tailoring System, an existing U-M software application first designed to support tailored content in health communications and now in use across multiple educational technologies [17]. Tandem is also connected to Canvas, the learning management system, for the purposes of pulling enrollment data and reporting back graded assessments.

Tandem has several functions. In particular, Tandem surveys students about their individual approaches to teamwork, provides personalized feedback for individuals and teams with prompts for reflecting on how to improve and targeted lessons based on team performance, and gives instructors real-time reports to monitor team health. The lessons provide content on best practices in collaboration and information about common issues teams face paired with reflection questions designed to help students situate the knowledge in their own contexts. For example, lessons address why diverse teams face challenges but ultimately come up with better solutions than more homogeneous teams, how to recognize both good and bad team conflict, and how to communicate equitably. Importantly, the development team was intent on a tool that could recognize and address teamwork challenges that disproportionately affect traditionally marginalized student groups (e.g., through gendered division of team roles).

Tandem integrates information from several types of surveys that students complete throughout the term:

(1) The baseline beginning-of-term survey is given before the team experience begins. It asks students about their prior exposure to disciplinary content and skills, academic orientation, and concerns regarding working in teams, all of which are related to self-efficacy.
(2) The early group communication check is given after the first week of teamwork and asks students if they have been able to speak up, if they feel they are being heard, what they sense about each teammates' contributions thus far, and about barriers to equitable group decision making.
(3) Team checks (Table 1) for team health are short (5-question, 30-second, closed-ended) surveys completed frequently, usually on a weekly or biweekly basis. Students rate their team, not themselves or specific teammates, on equitable distribution of work, team logistics, equity of ideas shared, confidence of team success, and how well the team is working together in general.
(4) A peer review survey asks students to evaluate the extent to which they and each teammate have exhibited good teamwork, both in tasks (like how they contributed to specific project elements) and in behavior (like how well they listened to their teammates). This survey is often used at the middle and end of team projects.

3 RESEARCH QUESTIONS

Teams can be power-laden course structures that, in their worst forms, perpetuate inequity. Thus, identifying if and how teams are marginalizing specific students, particularly at scale using quantitative data, is an important goal [6]. Under the assumption that teammates are more likely to have conflicting perspectives on team effectiveness when some are having negative team experiences (e.g., experiencing microaggressions), here we investigate if individual perceptions being different from group perceptions can be a marker of inequitable team behaviors. We define such inequitable team behaviors to include disproportionate gendered and racialized differences in terms of who gets listened to, whose ideas are discussed and pursued, how evenly work is distributed, and who is seen as an expert.

Specifically, we explore divergence in scores on the team check surveys which provide information about team dynamics, contributions, and processes. The five team check items are each presented to students on an end-anchored nine-point numeric scale (Table 1), and visual summaries of the team averages are shown to instructors and individuals throughout the term so long as three or more teammates have responded. The team check data potentially support a quantitative assessment of the extent to which teamwork behavior is equitable; thus, with the ultimate goal of investigating whether marginalization of individuals on teams is detectable in the team check data, here we pursued the following research questions: RQ1) How does the variability in student evaluations of their team change over time? and RQ2) How do the demographics of students who report outlier responses compare to their peers?

4 METHODS

4.1 Data Collection

Data were collected from the 624 undergraduate students who earned a course grade in a lower-division business course (referred to herein as BUS 201) about leadership, creating value, and the role of business in society. Gender and race/ethnicity data were self-reported by the students on the Tandem beginning-of-term survey (Table 2); that is, the categories reflect a wider range of identities than those available through U-M's student information system. BUS 201 typically enrolls about 80% sophomore-level students, 20% junior-level students, and very few senior and first-year students. This research was determined by the Institutional Review Board to be exempt from ongoing review.

BUS 201 is typically taught in person; however, during this semester studied the course was offered both in hybrid and fully online formats due to the COVID-19 pandemic. Students attended two class meetings per week in sections of about 20 students each. One in-person or online class meeting was instructor-led and the other online-only discussion-like session was facilitated by teaching assistants and largely devoted to team-based activities. Students were grouped by instructors into 128 teams with between four and six students per team (M = 4.9). The targeted group size was five students, and efforts were made to not strand students identifying with

Table 1: The five Tandem team check items to which students individually respond.

Team check item	Lower anchor (1)	Upper anchor (9)
Equal workloads	The workload is not distributed evenly.	Everyone is pulling their own weight.
Logistics	We often face logistical barriers (e.g., we cannot find convenient meeting times).	We have no problems with logistics (e.g., we all stay in touch about the project).
Sharing ideas	Only one or two people contribute ideas for our project.	Everyone evenly contributes ideas for our project.
Team confidence	I worry we won't do well on this project.	We're definitely going to do well on this project.
Working well	We often have problems working together.	We work really well together.

Table 2: Summary of selected demographic information for the 624 students in the course of interest. Students were able to select more than one race/ethnicity so these values sum to more than 100%.

Characteristic	*n*	%
Gender		
Female	255	41
Male	363	58
Non-binary	0	0
Self-described	1	<1
Prefer not to disclose	4	1
Missing	1	<1
Race/ethnicity		
American Indian or Alaskan Native	5	1
Asian	215	34
Black or African American	24	4
Hispanic or Latinx	48	8
Middle Eastern or North African	20	3
Native Hawaiian or Other Pacific Islander	2	<1
White	340	54
Self-described	5	1
Prefer not to disclose	11	2

historically marginalized populations (by gender and race/ethnicity [20]) alone on a team.

Each team worked together for the duration of the semester, producing a customer insight report near the middle of the term as well as a cumulative final video project. Students were expected to complete the Tandem team checks (Table 1) in seven different weeks during the 15-week semester; example team check data is shown in Table 3. Students also received a different, personalized Tandem lesson about teamwork each week. Some lessons were timed to be distributed to the whole course during the same week, but during other weeks, each team received lessons that were algorithmically selected based on how they had responded to the team checks. Responses to these team checks and lesson activities together constituted 6% of students' final course grades. Importantly, teamwork skills such as learning how to provide constructive feedback and resolve team conflict were explicit course learning goals.

4.2 Data Analysis

For each team check item (e.g., logistics), the change in each student's score from week to week was calculated. This transformation was employed to help avoid issues resulting from students potentially using the team check scale in different normative ways; for example, some students might default to using most of the scale while others are more reluctant to give any low or neutral ratings. We hypothesized that differences in a student's scores from week to week might be more meaningful than the absolute scores they reported. Because there were seven team checks, this manipulation resulted in a vector of six scores per student per team check item. These "change" scores were calculated for each team member; for instance, in the provided example (Table 4), between the first and second team checks (TC2-TC1), team member 1 did not change their logistics score while team member 3 decreased their logistics score by three units. Students' change scores were individually compared against the team's median change score to identify outliers, defined herein as those greater than one unit away from the team median for that particular change. We refer to these responses as outliers to signify the difference from their teammates' reported perspectives, but highlight that this is not the typical usage of outlier from a statistical lens. We considered this definition reasonable because it resulted in a low number of change scores being considered outliers overall (1,357 outliers / 11,820 total change scores = 11%). All data transformations and analyses were conducted using Python 3.

4.3 Limitations

This study limits the metric for inequitable behavior to regularly-collected but still self-reported survey items. We assume that these data actually correlate with real positive and negative team experiences but acknowledge that it is important to validate this assumption by triangulating with other data sources such as class meeting observations, interviews with both students and instructors, and assessments of who contributed and how often to team products.

Additionally, relying on the changes in students' team check scores from week to week made it challenging to deal with missing data because most missing team checks automatically impacted two change scores (this is the only way in which missing data was generated because students were required to respond to all five team check questions in order to submit the survey). We purposefully

Table 3: Example raw data from one of the seven team checks for one four-person team. Items were scored on a 1 (low) to 9 (high) scale; refer to Table 1 for scale anchors.

Team member	Equal workloads	Logistics	Sharing ideas	Team confidence	Working well
1	6	8	7	7	8
2	5	5	5	4	5
3	6	6	6	6	6
4	8	8	8	7	8

Table 4: Example logistics team check (TC) change scores and outliers (*) for one four-person team. As an example, TC2-TC1 shows the difference between team check 2 and team check 1. * indicates scores that were considered outliers for the given change.

Team member	TC2-TC1[a]	TC3-TC2	TC4-TC3	TC5-TC4	TC6-TC5	TC7-TC6
1	0	0	1	0	0	0
2	-1	-1*	2*	0	1	0
3	-3*	4*	-1*	2*	-1	1
4	3*	1	0	0	0	0
Mdn	-0.5	0.5	0.5	0	0	0

deleted the missing data listwise, removing all students (226 / 624 = 36%) who were missing one or more team checks, affecting almost all (116 / 128 = 91%) of the teams in the course. This heavy-handed approach rendered a relatively large pool of missing data, indicating that the results described herein are not representative of the full BUS 201 population, though we found no statistical gendered (χ^2 (2, N = 639) = 1.05, p = .59) or race-based (χ^2 (6, N = 621) = 8.89, p = .18) patterns in the groups of students who completed all team checks versus those who missed one or more. In any case, this limitation was acceptable for this exploratory study because we were more intent on investigating the behavior of students in rating their teams than the behavior of students particularly in the BUS 201 course context. Future iterations of this work could take a pairwise deletion approach, removing only the affected change scores rather than all scores for a student missing one or more team checks.

5 RESULTS

5.1 RQ1) How does the variability in student evaluations of their team change over time?

The real team sizes ranged from four to six students. Because students were removed when they were missing any team check, however, the number of students per team included in the analyses ranges from one to five (every team of six students had at least one student miss at least one team check). For simplicity, we continue to refer to these groups of students as teams recognizing that, in reality, the full teams include the students who were eliminated for having missing data. Overall, we found that the average team check scores vary minimally from survey to survey (Table 5) but generally increase for each team check item by about a half point over the term. The proportion of teams with one or more outliers for each change between team checks is shown in Figure 1. For every team size and team check item, the proportion of outliers at the end of the semester is smaller than at the beginning; all linear trendlines show a negative slope, ranging from -0.01 to -0.12. In other words, students evaluate changes in their team experience (or lack thereof) more similarly to their teammates over time; the variability between teammates' change in scores decreases. For the logistics, sharing ideas, and working well items, the decrease in outlier change scores over the semester is relatively smooth. The equal workloads and team confidence items show somewhat more variable changes, perhaps reflecting an uneven distribution of teamwork required throughout the semester and therefore less opportunity for teams to come to a shared understanding of team norms (e.g., final project work is likely most intense towards the end of the semester).

5.2 RQ2) How do the demographics of students who report outlier responses compare to their peers?

The ultimate goal of this study was to understand if the students who exhibited outlying behavior (reported a team change differently than their teammates) differed systematically from those who were never outliers by two demographic variables: gender and race/ethnicity. If outlier status regarding team change scores predicts awareness of marginalization on a team, we predicted women and historically marginalized students would be over-represented in these groups. In these analyses, the race/ethnicity categories were grouped as either 1) persons identifying with historically marginalized populations—American Indian or Alaskan Native, Black or African American, Hispanic or Latinx, Middle Eastern or North African, and Native Hawaiian or Other Pacific Islander (if students identified with any of these populations, they were included here; for example a student who selected both Black or African American and White was categorized with this group), or 2) Asian and White students, considered together because they tend to earn bachelor's degrees at similar rates [26]. We note that the results were the same when this group was limited to White students alone.

Table 5: Average scores ($M \pm SD$) on a nine-point scale (subject to ceiling effects) for all students, regardless of team size, who completed all seven team check (TC) surveys (N = 398).

TC item	TC1	TC2	TC3	TC4	TC5	TC6	TC7
Equal workload	7.42 ± 1.49	7.41 ± 1.58	7.62 ± 1.48	7.69 ± 1.52	7.56 ± 1.65	7.68 ± 1.51	7.83 ± 1.29
Logistics	7.30 ± 1.65	7.59 ± 1.46	7.78 ± 1.26	7.77 ± 1.45	7.83 ± 1.44	7.95 ± 1.26	8.05 ± 1.17
Sharing ideas	7.32 ± 1.57	7.29 ± 1.72	7.52 ± 1.56	7.81 ± 1.21	7.58 ± 1.58	7.71 ± 1.42	7.82 ± 1.31
Team confidence	7.43 ± 1.45	7.55 ± 1.40	7.77 ± 1.29	7.73 ± 1.32	7.82 ± 1.30	7.76 ± 1.48	7.74 ± 1.44
Working well	7.63 ± 1.28	7.75 ± 1.29	7.92 ± 1.19	8.13 ± 1.09	8.05 ± 1.21	8.10 ± 1.09	8.08 ± 1.05

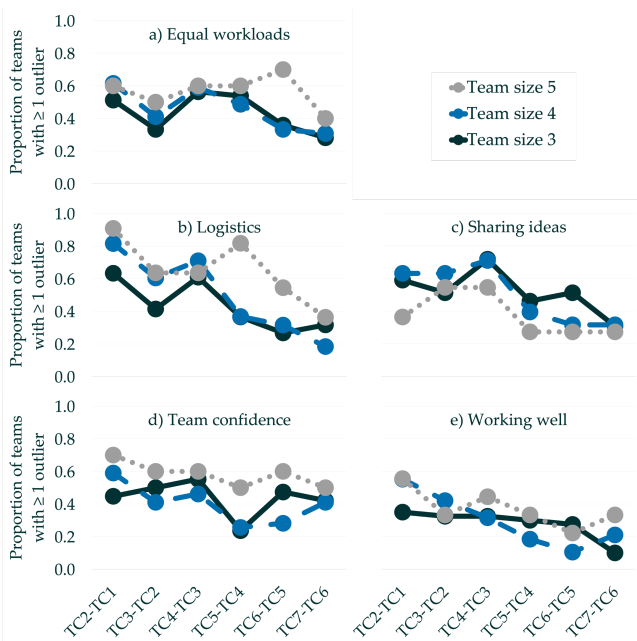

Figure 1: The proportions of teams with one or more outliers from the team median change for the a) equal workloads, b) logistics, c) sharing ideas, d) team confidence, and e) working well team check (TC) item scores over time. Team sizes of three, four, and five students are indicated by the solid, medium dashed, and small dotted lines, respectively. Note that the results shown in this figure are limited to only those teams (N = 90) where three, four, or five students (N = 328) responded to all the team checks because of the inherent variability and lack of ability to calculate medians introduced with teams where two or fewer students responded.

For each team check item, a Pearson's χ^2 test was used to assess correlation between the demographic variable and whether or not students had ever been an outlier, including only students who responded to all seven team checks. Of these 10 tests for correlation, two showed a significant relationship (Table 6). Specifically, female students were overall more likely to show deviations from the team median change on the logistics (57% (N = 93) of female students compared to 47% (N = 109) of male students) and working well items (38% (N = 61) of female students compared to 28% (N = 65) of male students). In terms of the odds ratio, female students were 1.5 times more likely to be an outlier than male students for both items. These differences were apparent for five out of the six change scores (all except TC7-TC6) and were always in the positive direction compared to the median. In other words, female students were more likely than male students to have positive changes in their logistics and working well scores from week to week. While these specific findings may be subject to future investigation, overall these analyses indicate few gendered or racialized patterns in week-to-week changes in response to team check items.

6 DISCUSSION

Here, we sought to understand if and how changes in student ratings of their team compare across teammates. In general, student teams appear to stabilize in terms of how synchronized they are in score changes over time, regardless of team size. This pattern suggests that teams establish norms and expectations over the duration of this course, and that towards the end, students on a team report changes in team health more similarly than they do at the beginning of the term. In studies of collaborative and project-based learning, then, researchers should be highly conscientious about when team survey or evaluation data is collected. Further, in contrast to our presuppositions, students who had markedly different changes in their ratings from week to week as compared to their teammates were, in all but one case, not more likely to identify with a particular gender or race/ethnicity. That is, while a number of individuals report team health changes in a different direction or of a different magnitude than their teammates, these students are not disproportionately of groups observed in prior studies to be more likely to experience team marginalization [13, 18].

Overall, while this measure has broadly not revealed teamwork inequities in this context, the method provides a basis for future research. In courses where deviations among teammates in team check score changes remain consistent throughout the term, further investigation of team data (such as studies of the more direct peer evaluations on the Tandem mid- or end-of-term surveys) may reveal if teammates' different change scores are indicative of inequitable or otherwise poor team behaviors. It is worth reminding the reader that here, students missing any team check data were not included; students with especially strong or negative experiences may have been less likely to complete all team checks, perhaps to avoid any indication of team conflict.

This exploratory work exposes additional research questions, two of which we describe here. First, the stabilization of changes to team check scores could reasonably indicate at least two different scenarios. Teammates may indeed learn to each see their

Table 6: Pearson's χ^2 test statistics (df = 1; * p < .05) for correlations between the gender (N = 393) or race/ethnicity (N = 398) categories and whether or not the student was ever an outlier. Gender categories besides female and male were not included here due to low numbers of students.

	Equal workloads	Logistics	Sharing ideas	Team confidence	Working well
Gender	0.06	3.98*	<0.01	0.61	3.96*
Race/ethnicity	1.03	0.49	0.18	0.26	0.02

team through similar lenses and interpret their team experiences in a common way. In contrast, teammates with dissenting perspectives may self-regulate and begin to cover these opinions when responding to the team checks, essentially burying real variability. A mixed methods study following teams that show different team check patterns with interviews for how and why students change their scores between surveys would be useful for building better understanding of the meanings behind changes in the data.

Second, regarding RQ2, here we have bluntly treated all students who ever exhibited an outlier change score together as one group, but the distribution of who is an outlier may be different depending on the specific time point, the frequency with which the student is an outlier, and the direction in which the student is an outlier. Understanding if some groups of students are more or less prone to being an outlier on a team, particularly in the more variable early parts of the semester, could reveal an interaction between inequitable team behavior and team maturity. The "direction" (positive or negative) in which a student is an outlier from the median team change should, in particular, reveal important information about students' team experiences.

7 CONCLUSION

Clearly, building students' teamwork skills is a broad goal across undergraduate curricula. However, when teams behave inequitably, the consequences extend beyond the exclusionary behavior itself, leading to a lack of learning opportunities for students who are marginalized. Women have been known to take on or be assigned more culturally normative tasks of project management and communication, in contrast to technical tasks like designing and building [11, 18, 25], making possible a perpetual cycle in which women both embrace and are offered fewer technical roles even as they advance in their coursework. Marginalization based on race/ethnicity in project teams has also been reported [9, 27]. Unfortunately, instructors do not always know how to teach for inclusive behavior in teams, and instructors have themselves called for tools and resources to help implement inclusive pedagogical practices [2]. Here, we have used Tandem team check data to develop a measure showing that the changes in students' scores tend to stabilize over the course of the semester when compared with their teammates. While gendered and racialized patterns were largely not identified in this context, the measure will be useful in future investigations of inequitable team behavior.

ACKNOWLEDGMENTS

This work was supported by Catherine Shakespeare, our key course partner in the Ross School of Business. Robin R. Fowler, Laura K. Alford, and Stephanie A. Sheffield are the faculty leads responsible for the conceptual foundation and inspiration of Tandem, while design and development have been led by the Center for Academic Innovation. Funding was provided by the College of Engineering Educational Innovation Accelerator program and the National Science Foundation (DUE 2120252). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] David J. Asai. 2020. Race Matters. *Cell* 181, 4 (May 2020), 754–757. https://doi.org/10.1016/j.cell.2020.03.044

[2] Kacey Beddoes and Grace Panther. 2018. Gender and teamwork: an analysis of professors' perspectives and practices. *European Journal of Engineering Education* 43, 3 (May 2018), 330–343. https://doi.org/10.1080/03043797.2017.1367759

[3] Behzad Beigpourian and Matthew W Ohland. 2019. A systematized review: Gender and race in teamwork in undergraduate engineering classrooms. In *2019 ASEE Annual Conference & Exposition*. ASEE Conferences, Tampa, Florida. https://peer.asee.org/32011

[4] Maura Borrego, Jennifer Karlin, Lisa D. McNair, and Kacey Beddoes. 2013. Team Effectiveness Theory from Industrial and Organizational Psychology Applied to Engineering Student Project Teams: A Research Review. *Journal of Engineering Education* 102, 4 (Oct. 2013), 472–512. https://doi.org/10.1002/jee.20023

[5] James A Coller, Robin R Fowler, and Laura K Alford. 2020. Investigating the Impact and Pedagogical Implications of Extroversion on Team Dynamics in Post-Secondary Student Teams. virtual, 9. https://doi.org/10.1109/FIE44824.2020.9273825

[6] Darryl A. Dickerson, Stephanie Masta, Matthew W. Ohland, and Alice L. Pawley. 2021. How Can We Identify Teams at Risk of Marginalizing Minoritized Students, at Scale?. In *2021 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, virtual. http://peer.asee.org/how-can-we-identify-teams-at-risk-of-marginalizing-minoritized-students-at-scale

[7] Robin Fowler, Laura K Alford, Stephanie Sheffield, Caitlin Hayward, Trevion S Henderson, and Rebecca L Matz. 2021. Supporting Equitable Team Experiences Using Tandem, an Online Assessment and Learning Tool. In *2021 ASEE Virtual Annual Conference Content Access*. ASEE Conferences, Virtual Conference. https://peer.asee.org/37787

[8] Robin R Fowler and Magel P Su. 2018. Gendered risks of team-based learning: A model of inequitable task allocation in project-based learning. *IEEE Transactions on Education* 61, 4 (2018), 312–318. https://doi.org/10.1109/TE.2018.2816010

[9] Patton O. Garriott, Rachel L. Navarro, Lisa Y. Flores, Hang-Shim Lee, Ayli Carrero Pinedo, Diana Slivensky, Melissa Muñoz, Ruben Atilano, Ching-Lan Lin, Rebecca Gonzalez, Laura Luna, and Bo Hyun Lee. 2019. Surviving and thriving: Voices of Latina/o engineering students at a Hispanic serving institution. *Journal of Counseling Psychology* 66, 4 (2019), 437–448. https://doi.org/10.1037/cou0000351

[10] Hart Research Associates. 2015. *Falling Short? College Learning and Career Success*. Technical Report. Association of American Colleges and Universities, Washington, DC. https://www.aacu.org/leap/public-opinion-research/2015-survey-falling-short

[11] Laura Hirshfield and Debbie Chachra. 2015. Task choice, group dynamics and learning goals: Understanding student activities in teams. In *2015 IEEE Frontiers in Education Conference (FIE)*. IEEE Computer Society, 1–5. https://doi.org/10.1109/FIE.2015.7344043

[12] Laura Hirshfield and Debbie Chachra. 2019. Experience is not mastery: Unexpected interactions between project task choice and measures of academic confidence and self-efficacy in first-year engineering students. *The International Journal of Engineering Education* 35, 3 (2019), 806–823. https://dialnet.unirioja.es/servlet/articulo?codigo=6889297

[13] Laura J. Hirshfield and Debbie Chachra. 2019. Comparing the Impact of Project Experiences across the Engineering Curriculum. *International Journal of Research in Education and Science* 5, 2 (2019), 468–487. http://eric.ed.gov/?id=EJ1203804

[14] Mona Itani and Issam Srour. 2016. Engineering Students' Perceptions of Soft Skills, Industry Expectations, and Career Aspirations. *Journal of Professional Issues in Engineering Education and Practice* 142, 1 (Jan. 2016), 04015005. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000247

[15] David W. Johnson, Roger T. Johnson, and Karl A. Smith. 1998. Cooperative Learning Returns to College: What Evidence Is There That It Works? *Change* 30, 4 (1998), 26–35. https://doi.org/10.1080/00091389809602629

[16] Rose M. Marra, Kelly A. Rodgers, Demei Shen, and Barbara Bogue. 2009. Women Engineering Students and Self-Efficacy: A Multi-Year, Multi-Institution Study of Women Engineering Student Self-Efficacy. *Journal of Engineering Education* 98, 1 (2009), 27–38. https://doi.org/10.1002/j.2168-9830.2009.tb01003.x

[17] Rebecca L. Matz, Kyle W. Schulz, Elizabeth N. Hanley, Holly A. Derry, Benjamin T. Hayward, Benjamin P. Koester, Caitlin Hayward, and Timothy A. McKay. 2021. Analyzing the Efficacy of ECoach in Supporting Gateway Course Success Through Tailored Support. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge (LAK '21)*. Association for Computing Machinery, Irvine, CA, 216–225. https://doi.org/10.1145/3448139.3448160

[18] Lorelle Meadows and Denise Sekaquaptewa. 2013. The Influence of Gender Stereotypes on Role Adoption in Student Teams. In *2013 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, Atlanta, Georgia, 23.1217.1–23.1217.16. https://doi.org/10.18260/1-2--22602

[19] Donald Nordberg. 2008. Group projects: more learning? Less fair? A conundrum in assessing postgraduate business education. *Assessment & Evaluation in Higher Education* 33, 5 (Oct. 2008), 481–492. https://doi.org/10.1080/02602930701698835

[20] Barbara Oakley, Richard M. Felder, Rebecca Brent, and Imad Elhajj. 2004. Turning Student Groups into Effective Teams. *Journal of Student Centered Learning* 2, 1 (2004), 9–34.

[21] Honor J. Passow and Christian H. Passow. 2017. What Competencies Should Undergraduate Engineering Programs Emphasize? A Systematic Review. *Journal of Engineering Education* 106, 3 (2017), 475–526. https://doi.org/10.1002/jee.20171

[22] Linda Riebe, Antonia Girardi, and Craig Whitsed. 2016. A Systematic Literature Review of Teamwork Pedagogy in Higher Education. *Small Group Research* 47, 6 (Dec. 2016), 619–664. https://doi.org/10.1177/1046496416665221

[23] Sue V. Rosser. 1998. Group Work in Science, Engineering, and Mathematics: Consequences of Ignoring Gender and Race. *College Teaching* 46, 3 (1998), 82–88. http://www.jstor.org/stable/27558891

[24] Elaine Seymour and Anne-Barrie Hunter (Eds.). 2019. *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*. Springer International Publishing. https://doi.org/10.1007/978-3-030-25304-2

[25] Elizabeth Strehl and Robin Fowler. 2019. Experimental Evidence Regarding Gendered Task Allocation on Teams. In *2019 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, Tampa, Florida, 32797. https://doi.org/10.18260/1-2--32797

[26] US Department of Education, Office of Planning Evaluation and Policy Development, and Office of the Under Secretary. 2016. Advancing diversity and inclusion in higher education: key data highlights focusing on race and ethnicity and promising practices. (2016). http://www2.ed.gov/rschstat/research/pubs/advancing-diversity-inclusion.pdf

[27] Kerrie G. Wilkins-Yel, Jacqueline Hyman, and Nelson O. O. Zounlome. 2019. Linking intersectional invisibility and hypervisibility to experiences of microaggressions among graduate women of color in STEM. *Journal of Vocational Behavior* 113 (Aug. 2019), 51–61. https://doi.org/10.1016/j.jvb.2018.10.018