How Leaders Emerge: Gender Composition, Leader Selection, and Team Performance in India

Ashutosh Bhuradia[†]

Job Market Paper
December 2025

Updated frequently. Click here for the latest version

Abstract

College students entering the workforce are increasingly expected to collaborate and lead mixed-gender teams. Yet we know little about the interplay of gender, teamwork, and leadership especially in settings that are traditionally gender segregated. This paper examines this interplay through a 2x2 randomized field experiment involving 203 mixed gender teams in a project-based competition at an engineering college in rural India. Students are first randomly assigned to male-majority or female-majority teams and further into one of two leadership conditions: leaders assigned based on a baseline measure of emotional intelligence or chosen by their own teammates. I find that female-majority teams that choose their own leaders outperform other groups by 0.38-0.51σ, driven by greater teamwork and more effective leadership. In contrast, male-majority teams that choose their own leaders have the lowest performance—driven by free-riding, coordination failures, and ineffective leadership—while teams with leaders assigned based on emotional intelligence, regardless of gender composition, fall somewhere in between. These results imply that leadership development and team performance must account for the differing dynamics across gender groups in contexts where gender norms remain strong.

JEL Classification: C93, J16, M54
Keywords: Field Experiments, Teamwork, Leadership, Economics of Gender

I am extremely grateful to Peter Blair, Michela Carlana, Susan Dynarski, Nishith Prakash, Alejandro Ganimian, Saloni Gupta, Alex Bolves, and the participants in the Economics of Education Seminar at Harvard for their feedback and helpful comments. All errors are my own. Harvard IRB approval IRB25-0406. This study is registered in the AEA RCT Registry (AEARCTR-0016103).
[†] Harvard University, School of Education. Email: ashutosh_bhuradia@harvard.edu.

1 Introduction

Jobs that involve non-routine analytical team tasks have expanded rapidly over the past two decades. This shift has raised the value of non-cognitive skills such as teamwork and leadership (Deming, 2017; Edin et al., 2022). As organizations increasingly rely on teams to perform complex work, individuals are expected to collaborate and lead across gender lines. Yet, in many developing countries, gender norms and accompanying social segregation persist, creating tension between the collaborative demands of modern workplaces and the realities of gendered social and educational environments. Understanding how men and women collaborate and lead in such environments is thus central to preparing millions of students for the labor market.

A key question that emerges is how team composition and leadership selection shape collaboration across gender lines in these settings. For instance, team gender composition and the process of leadership selection can influence communication, coordination, the sharing of diverse perspectives, and conflict resolution, all of which affect team performance. However, evidence on when team gender composition matters, and how it interacts with leadership selection, remains unclear. In male-dominated STEM fields, these compositional dynamics may be especially consequential for women: female-majority settings can enhance psychological safety, whereas those with a male-majority may marginalize women. Similarly, the process of leadership selection can either legitimize women's authority or activate cultural biases that undermine it. Understanding how these forces jointly operate is essential for informing how firms organize teams and how higher education programs cultivate teamwork and leadership skills in an equitable manner.

I conduct a study to test the effect of team gender composition and leadership selection on the performance of teams in a project-based competition. Additionally, I examine the extent to which gender composition and leadership selection affects teamwork, leadership, and the role of women leaders. The study involves 203 student teams at an engineering college in rural India. I partner with the college's Training and Placement Office (TPO), which is responsible for students' job-skills training.

The context of rural India is important for several reasons. Gender shapes socialization and work in rural India, through occupational and social segregation (Montes et al., 2018). This mirrors aspects of college life. For instance, there are separate hostels (dorms) for men and women, and entry into hostel buildings is restricted to students of the same gender. In classrooms, women and men are discouraged to sit with each other. While social norms and the organization of the college itself limit mixed-gender interactions, students aspire to jobs that frequently require collaboration between men and women. These conditions provide a natural setting for understanding how gender norms influence teamwork in contexts that

resemble many higher education environments across developing countries.

At the start of the study, students are administered a baseline survey after which they are first randomized into mostly 3-member mixed-gender teams, which are either male or female majority. Teams are then randomized into one of two leadership selection conditions: teams are either assigned a leader based on emotional intelligence measured at the baseline or teams nominate a team member as their leader. The choice of emotional intelligence for leadership assignment is based on prior research that links emotional intelligence to team and leader performance (Weidmann and Deming, 2021; Weidmann et al., 2024).

Teams, with their externally assigned or peer-nominated leaders, then work together on a project for two weeks. The project is part of a competition with cash prizes and involves designing a Mobile App idea that addresses challenges in rural India. The competition's cash prizes are sufficiently incentivized for this sample. For instance, 66% of participants are from families with a self-reported income of less than $1,140 per year. Top performing teams can win $140 on average, which is more than a month's family income for a majority of participants. Teams in the top decile are guaranteed to win $34, which is still more than a week's family income for a majority of participants.

Submitted projects are evaluated and scored by two trained raters who are blind to the treatment conditions. Before the results from this evaluation are announced though, each participant is asked to fill out an endline survey. In the endline survey, participants rate the effectiveness of their team and team leader. They also report on their and other team members' contribution to the project. Top-20 scoring teams (the finalists) are then selected to participate in a in-person competition at the college.

I find that female-majority teams that nominate their own leaders outperform other teams. Specifically, female-majority teams that nominate their own leader score 0.51σ higher on their projects than male majority teams that nominate their own leader, 0.47σ higher than male majority teams that have leaders assigned to them, and 0.38σ higher than female-majority teams who have leaders assigned to them. Moreover, female-majority teams that nominate their own leaders are also more successful in advancing to the final round–13% of their teams are finalists, marginally (though not statistically) higher than female-majority and male-majority teams with assigned leaders at 11% and 10.5% respectively, and significantly higher than male-majority teams that nominate their own leader of whom only 4% make it to the final round. The results suggests that while female-majority teams tend to perform better on average, their overall performance gains are especially pronounced when they have agency to choose their own leader. In contrast, agency in choosing leaders does not benefit male-majority teams, and even undermines their performance at the top end of the distribution.

To understand factors that might contribute to these differences in performance, I ex-

amine outcomes related to teamwork, leadership, and contribution to project work. I find that female-majority teams consistently report better teamwork than male-majority teams. Across the teamwork measures, their scores are between 0.18σ and 0.34σ higher. As such, female-majority teams (with nominated or assigned leaders) report better teamwork than their male counterparts but there are no differences in teamwork between female-majority teams.

In contrast to teamwork, leadership outcomes paint a more nuanced picture and especially help understand the differences in performance between female-majority teams despite similar levels of teamwork. I capture three leadership outcomes. The first outcome is a leader effectiveness score on a scale of 1-10. The second is leadership quality index, a holistic measure constructed from five questions that capture different dimensions of leadership such as coordination, communication, and conflict resolution. Each team member rates their leader on these measures. Finally, I report how much team leaders are perceived to contribute to their team's work. Each team member reports the percentage of work they and other teams members (included the team leader) contributed to the project.

As before, female-majority teams that choose their own leader score 0.26σ higher on leader effectiveness and 0.42σ higher on leadership quality index compared to male-majority teams. However, unlike before, there are significant differences between female majority teams. Specifically, female majority teams that choose their own leader score 0.43σ higher on leader effectiveness, 0.33σ higher on leadership quality, and 8.34pp higher on leaders' contributions compared to female-majority team have been assigned a leader. There are no such differences between male-majority teams based on leadership structure. This suggests how leaders are selected matters more for female-majority teams than it does for teams with a male-majority. Taken together, gender composition of teams matters more for teamwork, and leadership structure matter more to leaders' effectiveness and their contribution to collaborative work, especially in female-majority teams.

Given the differences in leaders' effectiveness and quality, do the type of leaders differ across conditions? For example, in this gender segregated and patriarchal setting, one might expect women might be less likely to be chosen leaders in male-majority teams (accounting for the proportion of women) and less likely to be leaders if their peers choose leaders versus when leaders are assigned externally. However, I find that leadership selection does not affect the gender representation of leaders, and both processes (leaders externally assigned or nominated by peers) yield nearly identical shares of women leaders. Gender composition, in contrast, mechanically shapes who becomes the leader because teams draw from pools that are themselves more male or more female. For example, among teams that nominate their own leaders, 71.4% of leaders in male-majority teams (with 66.6% men) are men, and 75.5% of leaders in female-majority teams (with 66.6% women) are women. These figures

indicate that once we account for the underlying gender composition of teams, leadership selection shows no statistically significant gender disparity; teams tend to select leaders from the majority gender at rates that closely mirror the pool. As a result, when aggregating across all teams, the overall share of women leaders (53%) approximates their proportion in the sample (50.8%).

What then is the source of variation in the quality and performance of leaders across conditions? I observe three main sources of variation. First is the variation that arises from the experimental design: female-majority teams are much more likely to have women leaders. And if women leaders, and women in general, are systematically different than men in the baseline, then part of the variation can be explained by these differences. I find suggestive evidence of differences by gender. While men and women have similar ability (based on baseline IQ and EQ scores), women are more likely to be low-income by 12.6pp and have a stronger preference for leadership by 0.34σ. Thus, strong pecuniary incentives combined with a higher willingness to lead might drive performance and engagement in teams with more women. However, this variation would not explain the difference in performance between female-majority teams nor the *lack of difference* in performance between male and female-majority teams with assigned leaders.

The second source of variation is emotional intelligence, which is partly mechanical since leaders in the assigned-leader condition are selected based on emotional intelligence scores. Specifically, assigned leaders have 0.47σ higher emotional intelligence compared to nominated leaders–however, leader effectiveness scores and leader contributions for teams with assigned leaders suggest that higher emotional intelligence does not necessarily translate into better leadership. The third source of variation is leaders' willingness to lead. Leaders in female majority teams have a 0.22σ higher preference for leadership compared to leaders in male majority teams, reflecting both women's higher baseline preference for leadership and the greater likelihood of women becoming leaders in these teams. Separately, nominated leaders have a 0.23σ higher preference for leadership than assigned leaders, suggesting that peer deliberation favors individuals who are more motivated to take on the role.

These patterns in higher emotional intelligence among assigned leaders and higher willingness to lead among nominated leaders may reveal an important insight: emotional intelligence does not necessarily translate into a motivation to lead. Consistent with this, emotional intelligence and preference for leadership are only very weakly correlated ($\rho = 0.08$). This suggests that while emotional intelligence may matter, a leader's motivation to lead appears to play an independently important role in driving leadership performance.

Finally, an important source of the variation is the process of leadership selection itself. Leaders elected through a consultative process might have more legitimacy than leaders who are assigned by external actors. Women often seen as less legitimate in leadership

roles, unless they legitimize their roles (Vial et al., 2016). Democratic nomination of leaders could potentially serve to legitimize leaders, especially women, which could contribute to greater effectiveness of leaders in teams that nominate leaders. Whereas leaders in female-majority teams that have been assigned a leader, most of which are women, might been seen to have lower legitimacy.

Why then do male-majority teams that choose their own leaders not reap some of the legitimizing benefits of a democratic nomination process? One reason could be than men might be harsher in dealing with other men, especially when competing for leadership roles. Several studies indicate male to male negotiations can lead to less cooperation and worse outcome in competitive settings (Sutter et al., 2009; Eckel et al., 2021; Castillo et al., 2013) compared to female to female and male to female negotiations. This could be the case in male-majority team that choose their own leader–men negotiate to assume leadership and those who are not nominated as leaders might be reluctant to cooperate and contribute to the project. I observe some evidence of this non-cooperation through free-riding–specifically, more free-riding in male-majority teams that choose their own leaders, driven entirely by men in the team who are not nominated leaders.

This study contributes to several strands of literature. First, it adds to our understanding of teamwork especially in the context of mixed-gender teams, which are increasingly common in educational and workplace settings. Gender composition of teams affects business decisions, scientific output, and performance (Yang et al., 2022; Apesteguia et al., 2012a; Truffa and Wong, 2025). Evidence of the effect of team's gender composition on performance, in particular, is mixed or unclear. While Hoogendoorn et al. (2013), Berge et al. (2016) and Lamiraud and Vranceanu (2018) find that gender-equal or female-majority teams perform better, Apesteguia et al. (2012b), Aparicio Fenoll and Zaccagni (2022), and Hardt et al. (2025) find that male-majority or all-male teams do better. Even less clear is how the evidence would apply to a real-world setting. Lamiraud and Vranceanu (2018) and Berge et al. (2016) conduct their experimental study in a lab, and Hardt et al. (2025) and Aparicio Fenoll and Zaccagni (2022) focus on narrowly defined tasks (e.g., solving math problems in teams) that are less akin to non-routine cognitive team tasks that individuals might collaborate on in a workplace. In Apesteguia et al. (2012b), the teams are endogenously formed, thus limiting causal interpretation. While Hoogendoorn et al. (2013) conduct a field experiment, they are unable to identify mechanisms that drive performance between teams of different gender composition. My field experiment contributes to this literature by incorporating a hands-on project that closely resembles a non-routine cognitive team task and tests a number of mechanisms to better understand the determinants of performance.

Second, I contribute to a large and evolving literature on leadership selection. Experimental studies in both the lab and the field demonstrate that leadership selection matters

to the performance of both leaders and teams teams (Deserranno et al., 2019; Englmaier et al., 2024; Levy et al., 2011; Reuben and Timko, 2018; Brandts et al., 2015; Chemin, 2021; Weidmann et al., 2024). Leaders who are elected by team members or assigned based on observable ability tend to perform better than randomly elected leaders (Levy et al., 2011; Reuben and Timko, 2018; Brandts et al., 2015; Chemin, 2021), although in a lab setting randomly assigned leaders perform better than those who nominate themselves (Weidmann et al., 2024). However, the literature is sparse on the differences in the performance of teams and leaders when one set of leaders are elected, and another are assigned based on ability. Learning more about this distinction is important because leaders are seldom randomly selected in real-world scenarios—they are often selected based on observable ability or through deliberation between stakeholders. My study shed lights on the conditions under which different modes of leadership selection enhances or inhibits performance in the field.

Third, I contribute to the literature on leadership selection and the role of women leaders, especially in mixed-gender teams. Leadership selection processes determine the representation and legitimacy of women leaders in a variety of settings (Eckel et al., 2021; Gangadharan et al., 2016) and gender composition further affects women's willingness to become leaders and their influence within teams (Born et al., 2022; Chen and Houser, 2019; Karpowitz et al., 2024). While women leaders are more effective in certain contexts, gender stereotypes tend to lower evaluations of their leadership (De Paola et al., 2022). I see evidence of this in my study–women leaders contribute more than male leaders but are evaluated similarly or lower than their male leaders. My study thus contributes to understanding how gender composition and leadership structure interact to enhance the role, performance, and evaluation of women leaders. Finally, this paper is among the few to experimentally study teamwork and leadership in a developing-country context. By embedding a randomized design in an engineering college in India, it provides causal evidence on how gender composition and leadership selection shape collaboration and performance in settings where gender norms are still highly salient.

This paper is structured as follows. In Section 2 and Section 3, I describe the research design and the sample, and test for balance. In Section 4, I present the main results and mechanisms. In Section 5, I discuss the implication of my findings and conclude.

2 Research Design

I partnered with an engineering college in a predominantly rural part of Central India. The college has undergraduate programs in engineering and technology such as Computer Science, Data Science, Information Technology, Electrical Engineering, Electronics and Communications Engineering, Agriculture Engineering, etc. Within the college, I partnered with

the Training and Placement Office (TPO). Training and Placement Offices are influential departments within Indian colleges and are primarily responsible for job-skills training for students and establishing partnerships with companies to facilitate the hiring of students from the college. The TPO's interest in this study arose from feedback it received from recruiters about the importance of soft-skills (teamwork, leadership, communication skills, etc.) in hiring and deficiencies in the college coursework in developing these skills.

I enrolled 610 students to participate in the study (380 Year-1, 166 Year-2, and 64 Year-3 students). As part of the recruitment, students were told that they would participate in a team-based competition related to engineering and technology, but they were not told the specifics of the competition–who they will team-up with, what specific activity they would have to perform, and what the prizes were. Students who signed-up to participate in the competition were then asked to complete a baseline survey and assessment which captured students' basic demographic information, experience with teamwork and leadership, preferences for teamwork and leadership, IQ (through a shortened version of the Raven test), and their emotional intelligence. The assessment of emotional intelligence was based on PAGE – Perceived AI-Generated Emotions – originally developed by Weidmann and Xu (2024), which I adapted for the context of India.[1]

After the baseline, I randomized students into within-year teams. I followed a two-step randomization protocol which proceeded as follows: first, I randomly assigned students within the same academic year into teams and varied whether students were assigned to female-majority (2 female, 1 male) or male-majority (2 male, 1 female) teams.[2] In total, students were assigned into 203 teams (96 male-majority and 107 female-majority teams). Next, I randomly assigned teams into one of two leadership selection conditions: in one condition, the team member with the highest PAGE score (proxy for emotional intelligence) within each team was assigned to be the team leader (assigned-leader arm with 101 teams); in the other condition, teams were asked to deliberate among themselves for a day and nominate or choose their own team leader (nominated-leader treatment arm with 102 teams).[3] In Table 1, I show the distribution of team leader selection arms by team gender composition.

After team-formation but before informing teams of their leadership selection treatment assignment, teams were brought together in classrooms for a brief session. With the exception of team leader assignment, the sessions were identical but separate for teams in

[1]Details on test development and the psychometric properties of PAGE are provided in the Appendix, Section C.

[2]Because numbers within a year were not always perfectly divisible by 3, two teams had 4 members each and one team had 2 members – one team with 2 female and 2 male, one team with 3 female and 1 male, and another with 1 female and 1 male member. For consistency, I code both gender balanced teams as female-majority. In the Robustness Checks section, I present results from additional analysis to show that dropping these teams altogether from my analysis does not alter the main findings.

[3]Thirteen teams in the assigned-leader treatment arm had 2 members with the same PAGE score. In that case, one of the team members was randomly assigned as the team leader.

Table 1: Distribution of Leader Assignment by Team Gender Composition

Team Gender Composition	**Assigned Leader**	**Nominated Leader**
Female-Majority	54	53
Male-Majority	47	49

Notes: The table shows the number of teams by gender composition and whether the team leader was assigned or nominated. Leader assignment was stratified by team gender composition.

the leader-assigned and leader-nominated conditions. The teams did not know they were allocated to these two different conditions. During this session, team members met and introduced themselves, and played an ice-breaker game. The teams were then told more details about the competition in which they would participate. The competition involved working in their teams to develop a Mobile App idea to solve a problem in rural India. They were provided a clear criterion on which their projects would be judged (project scoring rubric in Appendix C3). They were told that the competition involved 203 teams and that they had 2 weeks to work on and submit their projects. In addition, they were told that the top 20 projects from a pool of 203 would be shortlisted to compete in an in-person competition at the college. Teams in the top 20 were also guaranteed cash prizes of Rs 3000 ($34.86) per team. In addition, there were separate prizes for the in-person winners. The top ranked team would win Rs 15,000 ($174.32), the second-ranked team would win Rs 12,000 ($139.45), and the third-ranked team would win Rs 10,000 ($116.21). At the end of the session, teams in the assigned leader arm were told the name of the assigned leader, and teams in the nominated-leader arm were told to deliberate and nominate their leader the next day and to simultaneously start working on their projects.

After the session, I shared Google Slides templates with each team. The templates were shared with each individual student's email accounts, which they provided in the baseline. I also collected information on the identity of the team leaders in the nominated-leader treatment arm via a Qualtrics link sent to each team the following day. Student teams worked on their projects for 2 weeks and submitted the project either via the shared Google Slide, or as separate pdf or presentation sent to the research team via email. Of the 203 teams, 156 (76.3%) submitted a project.

After the project submission deadline, I reached out individually to the students to complete an endline survey. The survey elicited students' perception of their teams' and team leaders' effectiveness, perceptions of their own and their teammates' contribution. During the same time, I hired and trained two evaluators to score each project. The raters were blinded to the treatment conditions. Specifically, raters were blinded to the gender composition (names were removed from submissions before the rating process) and the

leadership structure of the teams. After each project had been scored, I ranked the teams based on their total score averaged across two evaluators. The top-20 ranked teams were announced on the WhatsApp communities that consisted of all teams. Each of the top 20 teams was also separately informed and invited to compete in the final round. The final in-person round was held in the college auditorium. Each team presented for 10 minutes in a randomly assigned order. A panel of outside judges with a background in rural development and technology relevant to the competition was invited to assess the presentations. The winners were announced on the same day and the cash prizes were distributed in person to each team (see Appendix B for images from the competition). Figure 1 summarizes the timeline for the implementation and data collection activities described above.

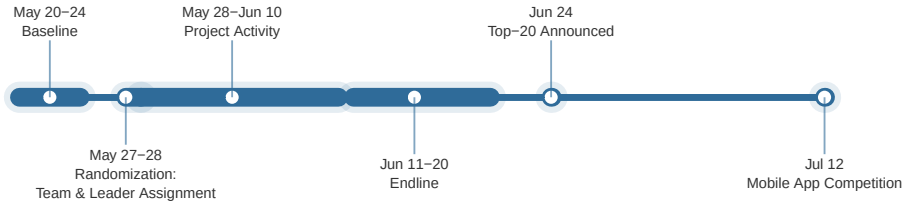


Figure 1: Implementation Timeline, May-July 2025

3 Sample and Balance Test

As reported in Table 2, a majority of the participants are low-income (self-reported annual family income less than Rs 1 lakh or $1140) and lower-caste. Compared to the male participants, female participants comprise 50.8% of the sample, are 12.6 percentage points more likely to be low-income (p-value = .001), and have a 0.34 SD (p-value = .000) higher preference for leadership.[4] However, I do not find differences in ability (IQ as measured by Raven, and emotional intelligence as measured by PAGE) between men and women in the sample.

I further test the correlation between ability, gender, and preference for leadership. I show these in Figure 2. I find a moderate positive correlation (0.37) between Raven and

[4]To measure preference for leadership, I asked students to indicate how much they'd like to be a Team Leader in the upcoming competition on a scale of 1 (I really don't want to be a team leader) to 10 (I really do want to be the team leader). Furthermore, the question said that 'team leaders are responsible for directing the group and making final decisions. The role will involve communicating with your team members, delegating, collating information, and making decisions. Please be as honest as possible.'

Table 2: Differences in Baseline Characteristics by Gender

Variable	**Mean (All)**	**Mean (Male)**	**Mean (Female)**	**Difference**	***p*-value**
Low income	0.667	0.603	0.729	0.126***	0.001
Upper caste	0.284	0.299	0.269	−0.030	0.427
PAGE score (std.)	0	-0.029	0.028	0.057	0.490
Raven score (std.)	0	0.067	-0.065	−0.132	0.103
Team lead pref. (std.)	0	-0.174	0.168	0.342***	0.000
Percent of sample		*49.2%*	*50.8%*		

Notes: The table reports mean values for male and female students and the p-value from a two-sided t-test of the difference in means. Stars denote statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. PAGE score, Raven score, and Team lead preference are standardized to have a mean of 0 and standard deviation of 1.

PAGE (emotional intelligence and IQ), and a positive weak correlation (0.18) between Raven and preference for leadership. Although I find a negligible correlation between gender and ability, I observe a positive, if weak, correlation (0.17) between being female and the preference for leadership. The positive relationship between being female and the preference for leadership is in contrast to Western settings, where the preferences for leadership are positively correlated with being male(Weidmann et al., 2024).

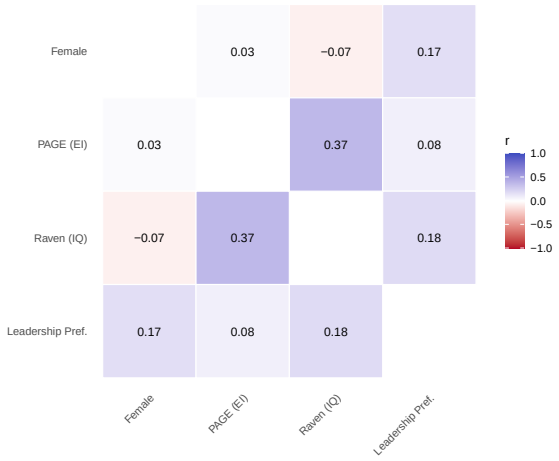


Figure 2: Correlation between Gender, Ability, and Leadership Preference

Notes: Pairwise Pearson correlations among gender, and standardized measures of emotional intelligence (PAGE), IQ (Raven), and leadership preference. Diagonal omitted.

I conduct a two-stage randomization to first randomize individuals within academic years into male and female majority teams, and then the teams into assigned-leader and nominated-leader arms. Thus, I conduct two balance tests. First, following Karpowitz

et al. (2024), I test for within-gender differences between male and female students assigned to teams of different gender composition. I report this balance in Table A1. I find no statistically significant within-gender differences between individuals assigned to male and female-majority teams. Second, I test for balance between teams in the assigned and nominated leadership arms. I report this balance in Table A2. Here, too, I find no statistically significant differences between teams assigned to different leadership arms. However, to demonstrate the robustness of my findings, I report both covariate-adjusted results in section 4.7 (*Robustness Checks*).

4 Results

4.1 Effect of Team Gender Composition and Leadership on Performance

In this section, I report results on team performance, teamwork, and leadership. The main outcome of interest is how well teams perform on their projects. As noted earlier, teams worked for two weeks developing App ideas that address challenges in rural India. The projects focused on challenges related to agriculture or education. The projects were then scored by two trained evaluators. The scoring was based on a pre-defined rubric shared with students at the beginning of the competition. The projects were evaluated according to the following criteria: problem statement, feasibility of the solution, user interface quality, technical innovation, and potential social impact. For each criterion, a team could score a minimum of 0 and a maximum of 4 points. All criteria were equally weighted. The projects that were not submitted received a score of 0. Thus, overall performance is a cumulative measure of whether teams submitted the projects and scores of teams that did.

I use the following regression specification to examine the effect of team gender composition and leadership on performance:

$$\begin{aligned} Y_i = \beta_0 \; &+ \; \beta_1 \, \text{FemaleMajority}_i + \beta_2 \, \text{Assigned}_i \\ &+ \; \beta_3 (\text{FemaleMajority}_i \times \text{Assigned}_i) + \delta_{\text{Year}(i)} + \varepsilon_i \end{aligned} \tag{1}$$

where Y_i denotes the outcome for team i. FemaleMajority_i is an indicator variable equal to one if the team is female-majority and zero if the team is male-majority, Assigned_i is an indicator equal to one if the team leader was assigned based on PAGE scores (emotional intelligence) and zero if the leader was nominated, and $\text{FemaleMajority}_i \times \text{Assigned}_i$ is the interaction between the two indicators. Because individuals are randomized into teams within academic cohorts, I also include year (cohort) fixed-effects, $\delta_{\text{Year}(i)}$. ε_i is the error term at the team level.

However, to aid interpretation, I estimate and present the following pairwise compar-

isons: within each leadership structure (nominated or assigned leader), I show the difference between teams of female and male-majority; then within each gender composition (male or female-majority), I show the differences between teams with a nominated versus those with an assigned leader. These pairwise comparisons correspond directly to the four experimental arms reported in Table 1. I report results for the four main team performance outcomes: (1) project completion; (2) project scores for teams that completed; (3) total project score (assigning zero to non-completers); and (4) the probability of being a finalist.

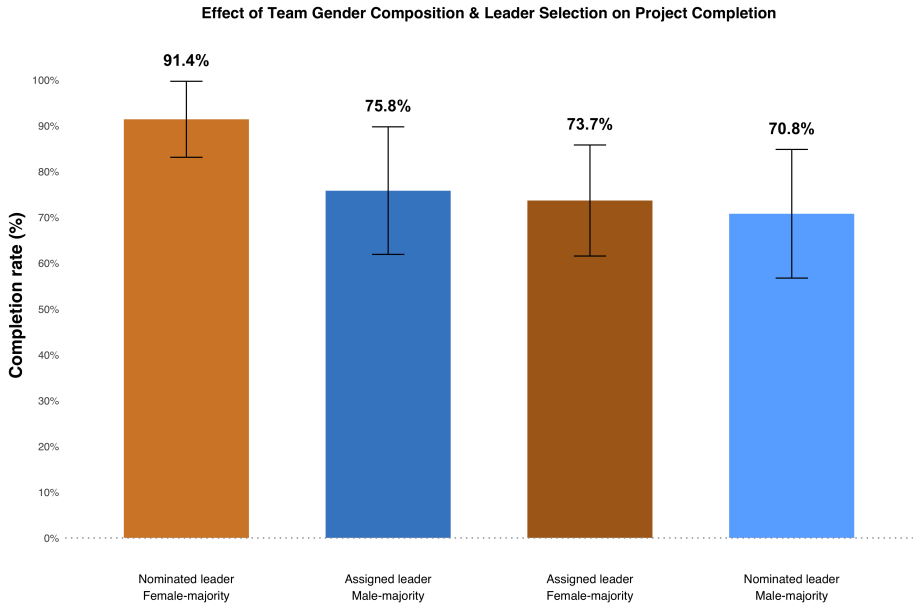


Figure 3: Effect on Gender Composition & Leadership Selection on Completion Rate

Notes: This figure shows project completion rates for each treatment group. The outcome is a binary variable, 'complete', which indicates whether a team completed (and submitted) their project or not. Each bar shows the mean of the outcome for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

In Figure 3, I show completion rates for all four groups. Female-majority teams with *nominated* leaders have the highest completion rate (91.4%), followed by male-majority teams with *assigned* leaders (75.8%), female-majority teams with *assigned* leaders (73.7%) and male-majority teams with *nominated* leaders (70.8%).

Column (1) of Table 2 tests whether these differences in completion rates are statistically significant. First, I compare teams within the same leadership structure (nominated or assigned). Among teams with nominated leaders, female-majority teams are 20.6 percentage points more likely to complete than their male-majority counterparts ($p < 0.01$). Among

teams with assigned leaders, the difference in submission rates between female- and male-majority teams is small and statistically insignificant.

Next, I compare teams with the same gender composition (female or male-majority). Among female-majority teams, assigning (rather than nominating) a leader reduces the completion rate by 17.7 percentage points ($p < 0.05$). Among male-majority teams, the roughly 5 pp higher completion rate under assigned versus nominated leadership (75.8% vs. 70.8%) is not statistically significant. Taken together, the completion rate is highest when female-majority teams nominate their own leader, and external assignment largely eliminates that advantage. However, the process of leadership selection does not affect completion rates for male-majority teams. These results indicate that the combination of female-majority composition and leader nomination yields the highest completion rates, while external assignment dampens that advantage.

Table 2: Performance by Gender Composition and Leadership Structure

	Completed	Project score (z)	Project score (z), 0 for incomplete	Finalists (%)
Panel A: Within Leadership Structure				
Nominated leader: Female-majority (=1)	0.206***	0.171	0.502***	0.123**
	(0.080)	(0.221)	(0.190)	(0.056)
Assigned leader: Female-majority (=1)	-0.022	0.277	0.082	0.036
	(0.089)	(0.261)	(0.212)	(0.065)
Panel B: Within Gender Composition				
Female-majority teams: Assigned (=1)	-0.177**	-0.045	-0.382**	-0.023
	(0.073)	(0.181)	(0.174)	(0.062)
Male-majority teams: Assigned (=1)	0.051	-0.150	0.038	0.065
	(0.092)	(0.278)	(0.216)	(0.053)
Observations (teams)	203	156	203	203

Notes: Each cell reports the coefficient on the listed indicator from a separate OLS regression with year fixed effects. Heteroskedasticity-robust standard errors, clustered at the team level, are in parentheses. Intercepts omitted. "Completed" is an indicator for whether a team completed (and submitted) the project or not. "Project score (z)" are standardized project scores for the subsample of completers. "Project score (z), 0 for incomplete" are standardized scores that include all teams–and teams with incomplete projects are assigned a score of 0. "Finalists" is an indicator for Top-20 teams. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The second outcome is project scores for teams that completed (and subsequently submitted) their projects. Figure 4 shows standardized project scores for the completers. The figure shows that female-majority teams perform slightly better than male-majority teams, although none of these differences are statistically significant. Column (2) of Table 2 tests these differences and confirms the pattern displayed in Figure 4.

I first compare teams within the same leadership structure (nominated or assigned). Among teams with nominated leaders, female-majority teams score about 0.17σ higher than male-majority teams; among teams with assigned leaders, the female–male majority

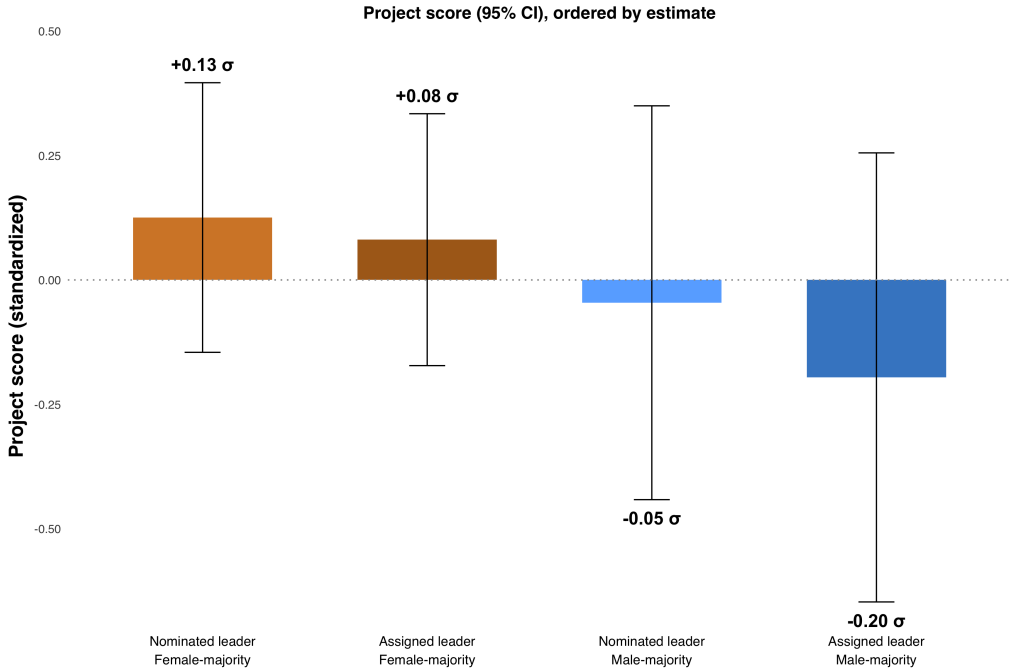


Figure 4: Effect on Gender Composition & Leadership Selection on Project Scores

Notes: This figure shows project scores for each treatment group. Scores are estimated at the team-level. Each team's project is scored by two trained raters. The project scores are then calculated by averaging the two scores and then by standardizing them to have a have mean of zero and a standard deviation of one. Teams that did not complete the project are excluded from this analysis. Thus, this figure compares the subsample of 'completers'. Each bar shows the mean of the outcome. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

difference is approximately 0.27σ—both differences are large but imprecisely estimated. Comparing teams with the same gender composition, nominating versus assigning a leader in female-majority teams leads to a modest but statistically insignificant increase of about 0.05σ; among male-majority teams, the analogous difference is roughly 0.15σ and likewise insignificant. Conditional on completion, project quality appears comparable across all groups.

To capture overall team performance, I assign zero to teams that did not submit and standardize total scores across all teams. This outcome shows the combined effect of completion and project quality. Figure 5 shows that female-majority teams with nominated leaders substantially outperform all other groups, with a mean of roughly $+0.36\sigma$, compared to near-zero or negative scores for others. Column (3) of Table 2 quantifies these patterns.

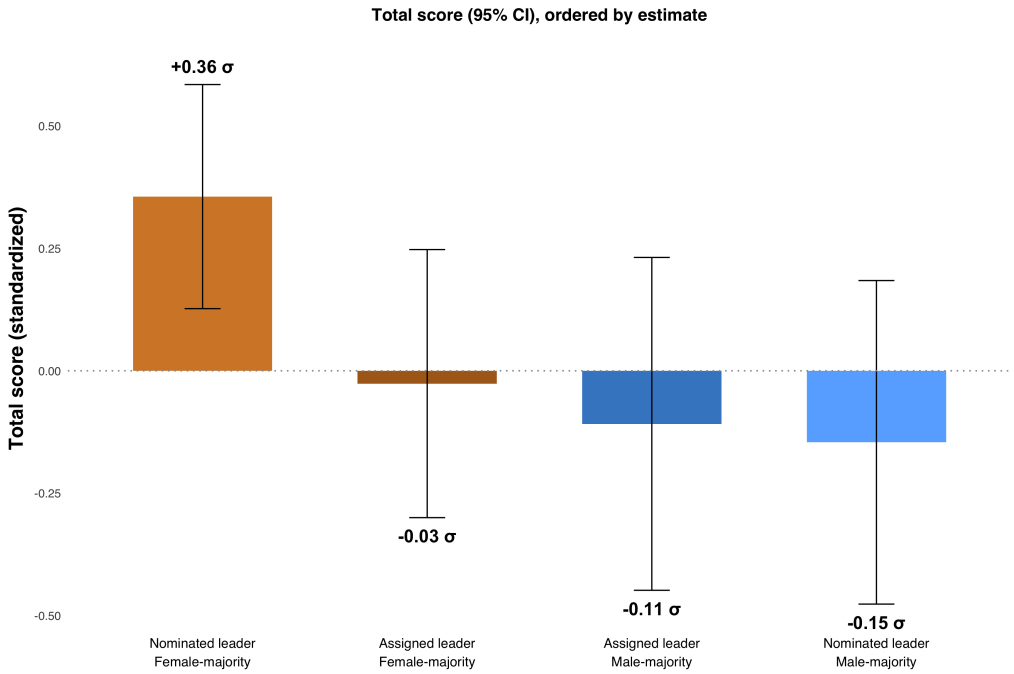


Figure 5: Effect on Gender Composition & Leadership Selection on Total Scores

Notes: This figure shows project scores each treatment group. Outcome is a z-score, standardized to have a mean of 0 and a standard deviation of 1. This includes all teams, with teams that did not complete (and submit) their projects assigned a score of 0. Each bar shows the mean score for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

When comparing teams within the same leadership structure, female-majority teams with nominated leaders outperform their male-majority counterparts by 0.51σ. Among teams with assigned leaders, the gap between female and male-majority teams is small (0.08σ) and not significant. For teams with the same gender composition, female-majority teams that nominate their own leader score 0.38σ ($p < 0.05$) higher than female-majority teams that are assigned a leader. For male-majority teams, leadership assignment makes little difference. Overall, female-majority teams that nominate their own leader outperform all other teams by a significant margin. This pattern suggests that agency in leadership selection benefits female-majority teams, while it does not benefit male-majority teams.

Finally, Figure 6 presents the share of teams that reached the finalist stage. For interpretability, the figure reports raw finalist rates (without year-fixed effects), while Column (4) of Table 2 shows estimates with year fixed effects. Female-majority teams with nominated leaders have the highest finalist representation (13.1%), followed by female-majority teams

with assigned leaders (11.1%), male-majority teams with assigned leaders (10.6%), and male-majority teams with nominated leaders (4.1%).

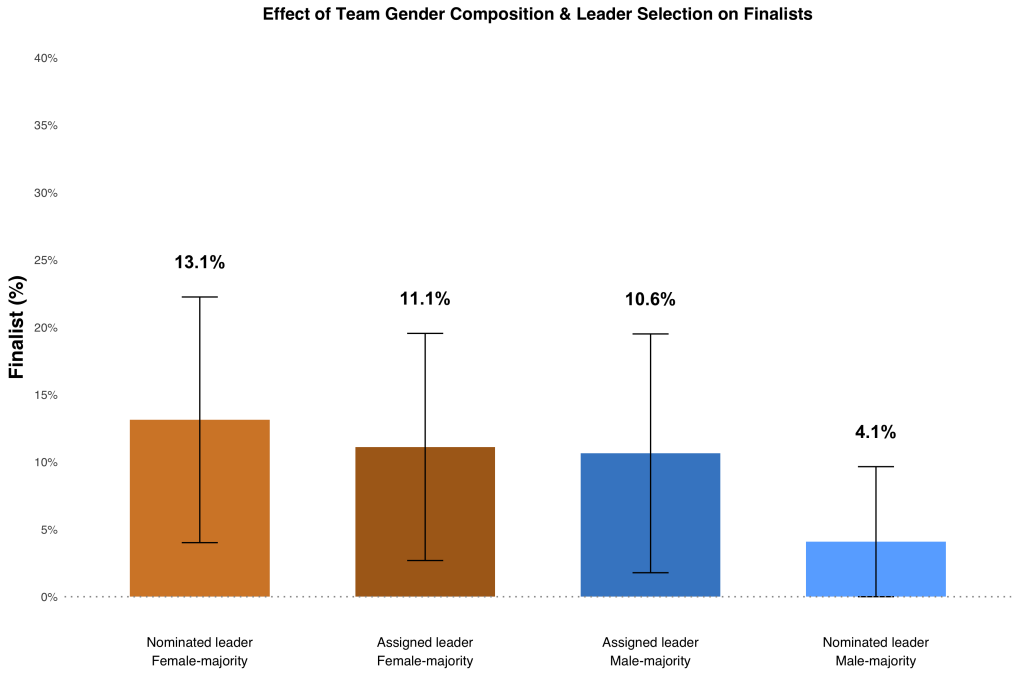


Figure 6: Representation of Teams in the Final Round

Notes: This figure shows the share of teams within each groups that were finalists. Bars represent the raw proportion of teams within each groups that reached the final round, with percentages displayed above each bar. The y-axis is expressed as the percentage of teams from that group that made it to the final round.

Within leadership structure: among teams with nominated leaders, female-majority teams are 12.3 percentage points more likely to reach the finalist stage than their male-majority counterparts ($p < 0.05$). Among teams with assigned leaders, the female–male difference is smaller (about 3.6 pp) and statistically insignificant. *Within gender composition:* for female-majority teams, assigning rather than nominating a leader slightly reduces the finalist rate by about 2.3 pp; for male-majority teams, assigned leaders lead to 6.5 pp higher finalist rate. However, neither of these estimates is statistically significant.

In summary, the performance of female-majority teams that nominate their own leaders seems to be driven by 'getting things done', as reflected by these teams completing their projects and getting them over the finish line. Even when accounting for only completed projects, teams with female-majority that nominate their leader score the highest on average (0.13σ) and reach the final round at a higher rate (13%). Taken together, the results suggest

that performance gains of female-majority teams are amplified when they have the autonomy to choose their own leaders. In contrast, agency in leadership selection does not improve outcomes for male-majority teams and may even dampen their performance at the top of the distribution.

4.2 Effect on Teamwork

To understand why female-majority teams—particularly those with nominated leaders—perform better, I examine several underlying components, including teamwork, and leaders' effectiveness and contributions. I begin by analyzing teamwork, using two complementary measures: one for team effectiveness and another for team equity. I construct these two indices that capture team effectiveness and team equity, based on endline survey items, which are standardized using factor analysis.[5] Both indices load cleanly onto single factors and exhibit high internal consistency (Cronbach's $\alpha = 0.92$ for effectiveness, 0.80 for equity).[6]

Figures 7 and 8 plot standardized team outcomes for each of the treatment groups. Table 3 tests whether differences across groups are statistically significant. I find female-majority teams with nominated leaders report substantially stronger teamwork experiences than their male-majority counterparts: team effectiveness is 0.34σ ($p < 0.05$) and team equity is 0.32σ ($p < 0.05$) higher. When leaders are externally assigned, differences remain positive and meaningful—female-majority teams report 0.18σ deviations higher team effectiveness (though not statistically significant) and 0.23σ higher team equity (p-value < 0.1) compared to male-majority teams.

The bottom panels of Table 3 compare teams with the same gender composition. Among female-majority teams, assigning a leader slightly reduces teamwork: team effectiveness falls by about 0.14σ and team equity by 0.12σ, though neither difference is statistically significant. Among male-majority teams, assigning a leader does not materially change teamwork, with near-zero differences in both indices.

Taken together, Figures 7 and 8 show that gender composition—particularly being on a female-majority team—plays a larger role in shaping teamwork quality than how leaders are selected. Female-majority teams consistently report stronger collaboration and a greater sense of fairness within teams, whether leaders are nominated or assigned.

[5]The Team Effectiveness Index was constructed from five Likert-scale items (e.g., "Our team worked together effectively") that capture overall team functioning and performance. The Team Equity Index reflects students' perceptions of fairness, voice, and inclusion within the team and was derived from six items (e.g., "Team members from all backgrounds were equally respected") All items used a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

[6]A pooled PCA of all 11 items shows the constructs are correlated ($r = 0.59$) but load distinctly on two factors, supporting the decision to treat team effectiveness and equity as related but conceptually separate outcomes.

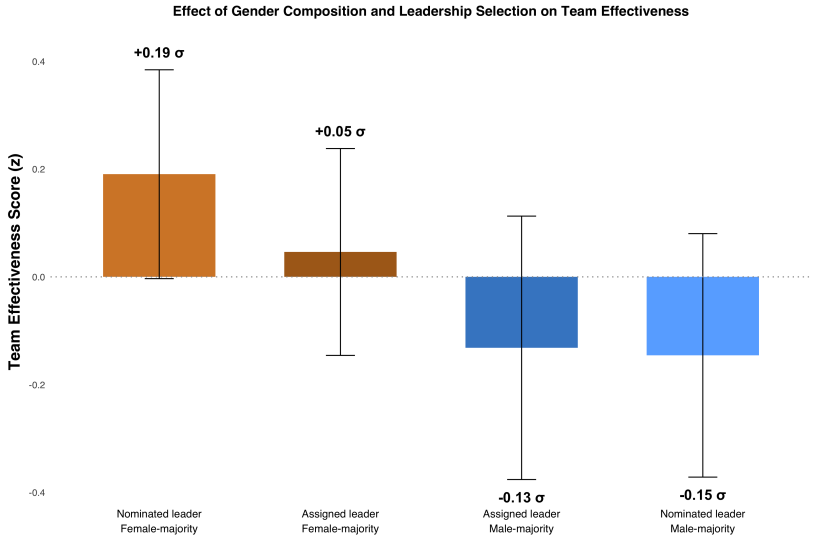


Figure 7: Effect on Gender Composition & Leadership Selection on Team Effectiveness

Notes: This figure shows team effectiveness scores for each treatment group. Team effectiveness index is constructed using exploratory factor analysis from responses to five survey items, and standardized to have a have mean of zero and a standard deviation of one. Each bar shows the mean of team effectiveness score for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

4.3 Effect on Leader Effectiveness, Leadership Quality, and Contributions

Although perceptions of teamwork could be a mechanism that explain the differences in performance between female and male-majority teams that nominate their own leaders, they do not explain differences in performance between female-majority teams. Female-majority teams, regardless of their leadership structure, show similar levels of teamwork. The differences in performance between teams with a female-majority remain unexplained. One hypothesis could be that since the primary difference between female-majority teams is their leadership structure, measures of leader effectiveness and quality could shed light on these differences and explain the variation in performance between female-majority teams. Thus, in Figures 11-13, I test whether there are differences in the effectiveness, quality, and contributions of team leaders–as perceived by their peers–across treatment arms, and especially between female-majority teams.

Each figure plots one of the following three measures of leadership: (i) a standardized team leader effectiveness score (1–10 score measured at the endline); (ii) a standardized leadership index composed of five key leadership dimensions— coordination, motivation,

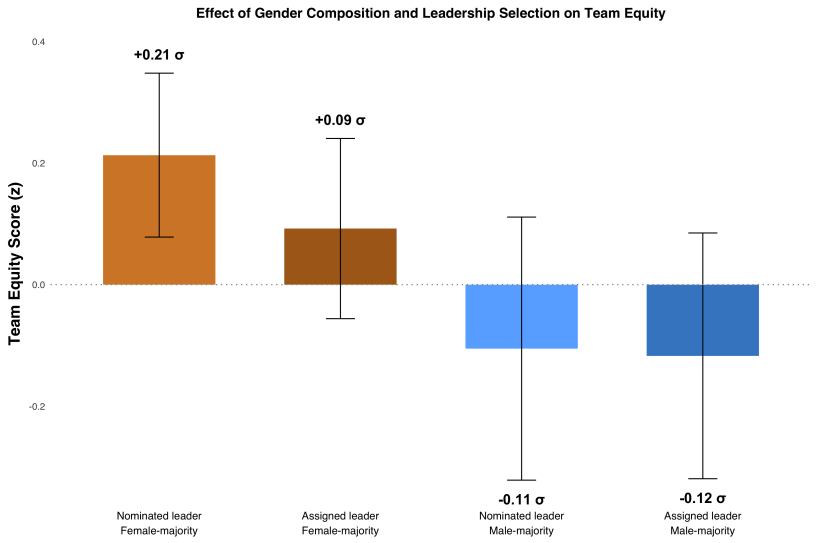


Figure 8: Effect on Gender Composition & Leadership Selection on Equitable Teamwork

Notes: This figure shows team equity scores for each treatment group. Team equity index is constructed using exploratory factor analysis from responses to six survey items, and standardized to have a have mean of zero and a standard deviation of one. Each bar shows the mean of team equity score for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

conflict resolution, task distribution, and openness to team members' ideas; [7] (iii) teammates' perception of how much (in percentage) of the work do team leaders contribute to. Results are shown in Figure 9 (Effectiveness), Figure 10 (Leadership Index), and Figure 11 (Leader Contribution) with corresponding point estimates and standard errors in tabular form in Table 4.

In Figure 9 and Column 1 of Table 4 female-majority teams with nominated leaders are perceived as having the most effective leadership, scoring 0.26σ (p-value < 0.1) higher than male-majority teams with nominated leaders and substantially higher (0.43σ, p-value < 0.01) than female-majority teams with assigned leaders. There are no statistically significant difference in leader effectiveness rating between male and female-majority teams with assigned leaders and between male-majority teams.

Figure 10 and Column 2 of Table 4 present differences in the leadership index–a holistic measure of perceived leadership quality. Here, too, nominated leaders of female-majority teams are perceived to perform the best. They score 0.42σ (p-value < 0.01) higher on the

[7] I construct a continuous index of perceived leadership effectiveness by extracting standardized factor scores from a one-factor exploratory factor analysis estimated using maximum likelihood. Internal consistency is high, with Cronbach's alpha = 0.93.

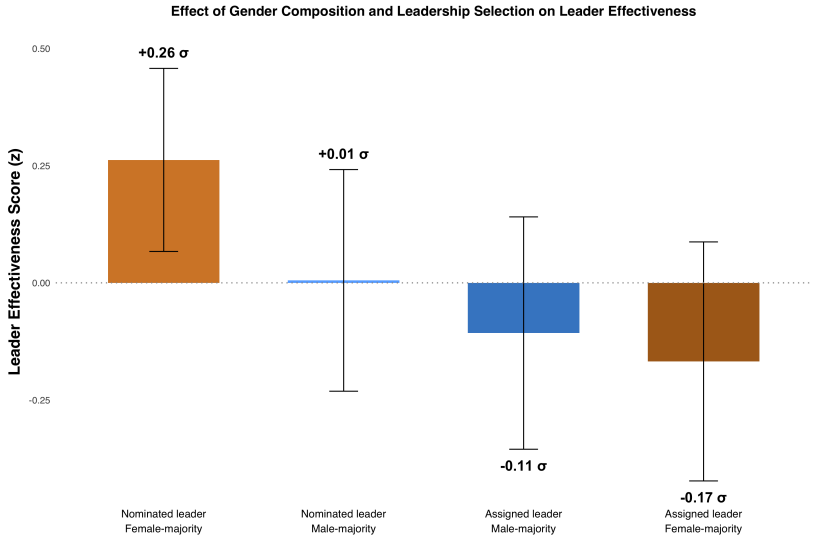


Figure 9: Effect on Gender Composition & Leadership Selection on Leader Effectiveness Rating

Notes: This figure shows leader effectiveness scores for each treatment group and pairwise comparisons. Leader effectiveness score is based on a single question, where respondents rate their team leader on a scale of 1-10. The score is calculated from the subset of individuals who are not team leaders, and standardized to have a mean of zero and a standard deviation of one. Each bar shows the mean of leader effectiveness z-score for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

leadership index than male-majority teams with nominated leaders and 0.33σ (p-value $<$ 0.05) higher than female majority teams with assigned leaders. As before, there are no statistically significant differences between the other groups.

Finally, peer-reported leader contributions are reported in Figure 11 and Column 3 of Table 4. As before, the leaders of female-majority teams with nominated leaders are perceived to be the most effective according to this measure. Teams with assigned leaders–male and female-majority–both report their leaders contribute approximately 33% to their collaborative work. Put differently, members of these teams perceive that their leaders contribute their "fair share" (for a 3 member team, equal or a fair share of work would be 33.33%). Teams with nominated leaders do more than their fair share–especially nominated leaders of female-majority teams are perceived to contribute the most on average (43.5%) and significantly more by 8.34pp (p-value $<$ 0.01) than assigned leaders of female-majority teams.

The results across leadership outcomes–leader effectiveness, leader quality, and leader contribution–highlight that leadership selection matters consistently for leaders of female-

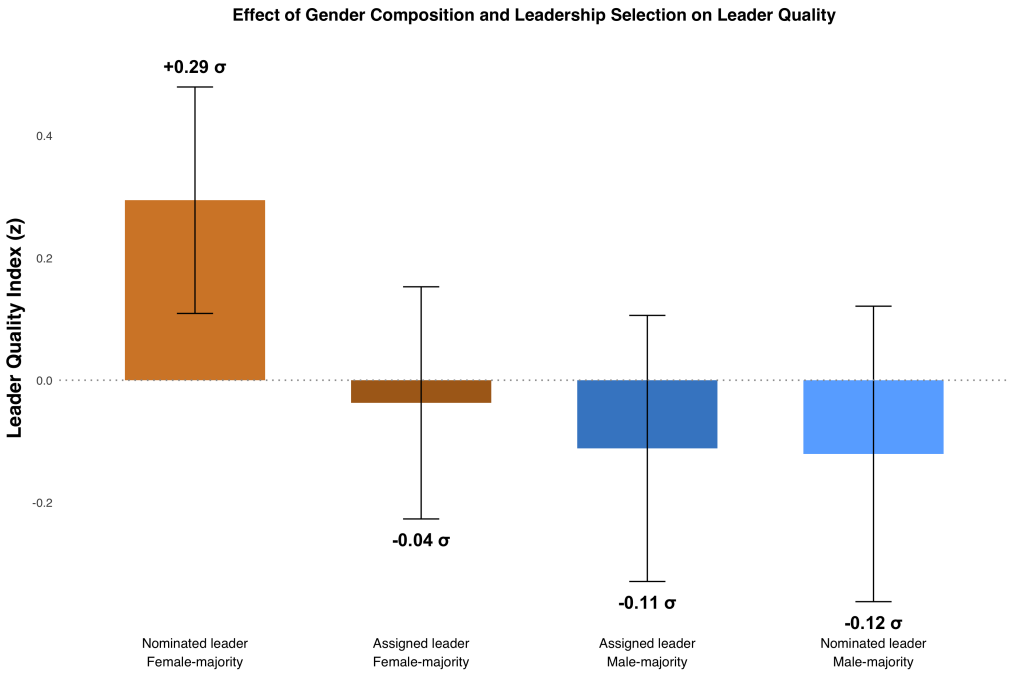


Figure 10: Effect on Gender Composition & Leadership Selection on Leader Quality

Notes: This figure shows scores from an index of leadership for each treatment group. Leadership index is constructed using exploratory factor analysis, using five survey items that correspond to different aspects of leadership. The score is calculated from the subset of individuals who are not team leaders and standardized to have a mean of zero and a standard deviation of one. Each bar shows the mean of leadership index z-score for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

majority teams. Assigning a leader (vs. nominating one) reduces leader effectiveness by 0.43σ ($p < 0.01$), reduces leader quality by 0.33σ ($p < 0.05$), and decreases peer-reported leader contribution by 8.34 pp ($p < 0.01$) among teams with female-majority. These large and significant differences in leadership between female-majority teams emphasize the importance of leadership selection for female-majority teams. By contrast, among male-majority teams, assignment yields small and statistically insignificant differences in effectiveness (-0.11σ) and quality ($+0.01\sigma$), and a marginally significant difference in leader contribution (-5.36 pp, p-value < 0.1). This seems to suggest that leadership selection is less important for male-majority teams.

Taken together, the leadership results explain the variation in performance between female-majority teams and reinforce the variation in performance between female and male-majority teams with nominated leaders. The result also complement the teamwork out-

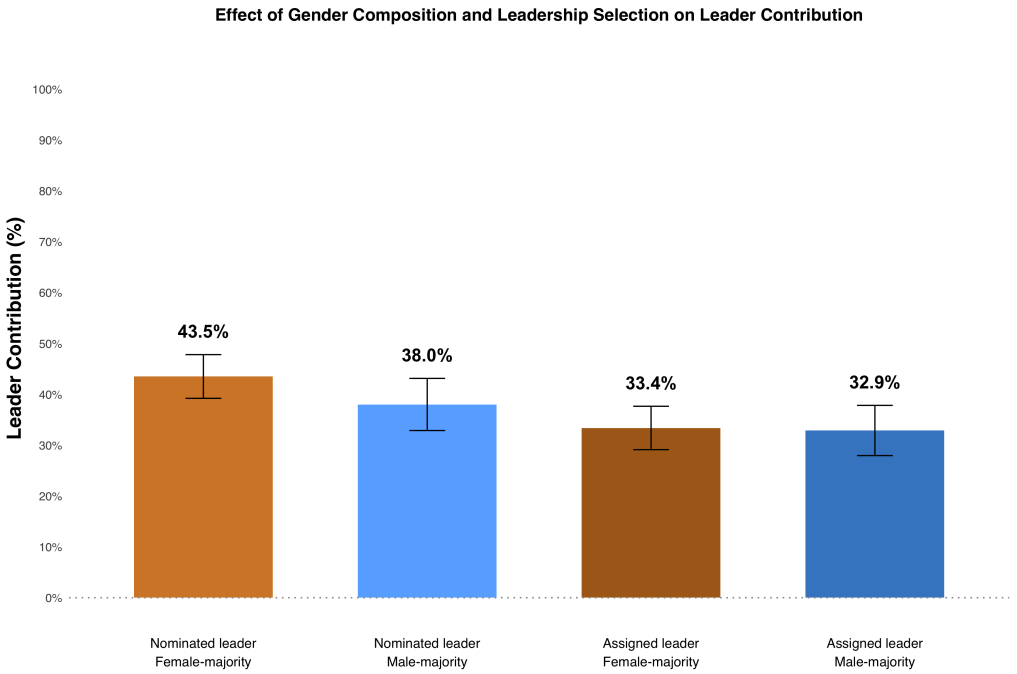


Figure 11: Effect on Gender Composition & Leadership Selection on Leader Contribution

Notes: This figure shows leader contribution (in %) for each treatment group. Leader contribution is calculated only from the subset of individuals who are not team leaders, with each team member reporting how much their leader contributed to the team's work. Each bar shows the mean of leader contribution for the indicated group. Error bars represent 95% confidence intervals based on robust standard errors clustered at the team level.

comes in the earlier subsection: gender composition shapes the general teamwork climate (effectiveness and equity), whereas leadership selection—especially peer nomination—shapes how leaders are evaluated and how much they contribute, with these effects concentrated in female-majority teams. This pattern helps to reconcile why female-majority teams with nominated leaders perform consistently better than other teams. Lastly, the results emphasize the importance of leaders for a team's performance. Specifically, leadership quality (Figure 10) – a holistic measure of a leader's perceived performance–closely maps onto the overall performance of teams (Figure 5), much more than teamwork-related outcomes. This suggests that while teamwork is important for how individuals perceive the experience of working in a team, leadership seems to correspond more strongly to a team's performance.

4.4 Variation in Characteristics of Team Leaders

An important question that arises from the differences in leadership' effectiveness, quality, and contributions is whether leaders themselves differ systematically between teams. Specifically, do certain types of leaders-such as those who are nominated versus assigned, or those leading female-majority versus male-majority teams-possess different baseline characteristics that might explain variation in their effectiveness and performance? To examine this question, I report differences in the' background of leaders (gender, caste, and income) and their ability (IQ, emotional intelligence, and preference to lead) across teams. I present these differences in Table 5.

Panel A of Table 5 compares female- and male-majority teams within each leadership structure. Leaders in female-majority teams, regardless of their leadership structure, are substantially more likely to be women (+46.9 pp & +42.4 pp, $p < 0.001$), consistent with the higher proportion of women in those teams. Beyond gender, there are few systematic differences: leaders in female-majority teams do not differ in socioeconomic background, cognitive ability, or emotional intelligence, though they exhibit a slightly higher preference for leadership ($+0.17\sigma$ & $+0.28\sigma$, for female-majority teams with nominated and assigned leaders, respectively, compared to their male-majority counterparts–though neither are statistically significant).

Panel B compares leadership structures within each gender composition. Among male-majority teams, assigning a leader—as rather than nominating one—reduces the share of low-income leaders (–18.2 pp, $p < 0.10$) and selects leaders with substantially higher emotional intelligence ($+0.63\sigma$, $p < 0.001$). Among female-majority teams, the only difference is in PAGE scores ($+0.32\sigma$ higher for assigned leaders). The higher PAGE scores among the assigned leaders are partly mechanical: in these arms, the leaders were deliberately chosen based on their emotional intelligence (proxied by PAGE) measured at baseline. Although assigned leaders exhibit lower preference for leadership (-0.18σ & -0.28σ, for female-majority and male-majority teams with assigned leaders, respectively) neither are statistically significant.

These results suggest that while team composition largely shapes who becomes a leader (e.g. increasing representation of women), the leader selection method shapes what kind of leader is chosen (e.g. one with a higher preference to lead or higher emotional intelligence). This distinction may help explain earlier results: nominated leaders, especially in female-majority teams, who are more motivated to lead, tend to contribute more, and are evaluated more favorably by their teammates. Assigned leaders, while more emotionally skilled, may be less inclined to take ownership of the leadership role, which could help explain their lower contributions and effectiveness (observed in earlier tables). This misalignment is

Table 5: Team Leader Characteristics by Leadership Selection & Gender Composition Group

	Female (=1)	Upper (=1)	Low income (=1)	Raven (z)	PAGE (z)	Leader pref. (z)
Panel A: Within Leadership Structure						
Nominated leader: Female-majority (=1)	0.469***	0.065	-0.073	0.072	0.218	0.174
	(0.088)	(0.095)	(0.093)	(0.200)	(0.203)	(0.172)
Assigned leader: Female-majority (=1)	0.424***	-0.020	0.153	-0.222	-0.083	0.275
	(0.091)	(0.091)	(0.097)	(0.207)	(0.151)	(0.199)
Panel B: Within Gender Composition						
Female-majority teams: Assigned (=1)	-0.032	-0.100	0.044	-0.268	0.323*	-0.183
	(0.086)	(0.091)	(0.092)	(0.196)	(0.176)	(0.172)
Male-majority teams: Assigned (=1)	0.012	-0.015	-0.182*	0.026	0.625***	-0.284
	(0.094)	(0.095)	(0.098)	(0.210)	(0.181)	(0.199)
Panel C: Full sample						
Female-majority (=1)	0.447***	0.022	0.039	-0.076	0.076	0.220*
	(0.063)	(0.066)	(0.068)	(0.144)	(0.131)	(0.132)
Assigned (=1)	-0.005	-0.059	-0.063	-0.130	0.466***	-0.227*
	(0.070)	(0.066)	(0.067)	(0.143)	(0.126)	(0.131)
Observations (leaders)	203	203	203	203	203	203

Notes: Each cell reports the coefficient from a separate OLS regression of the listed characteristic on the comparison indicator (Panel A: *female-majority* vs *male-majority* within nominated/assigned teams; Panel B: *assigned* vs *nominated* within female-/male-majority teams; Panel C: same two comparisons in the full team-leader sample). Standard errors clustered by team are in parentheses. Intercepts omitted. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

also reflected in the low correlation ($\rho = 0.08$) between willingness to lead and emotional intelligence as measured by PAGE.

Interestingly, there are no differences in the gender and caste identity of leaders regardless of whether they are assigned externally or nominated by peers. This is surprising in India's context, where gender and caste play a prominent role in shaping social dynamics. It would appear that other factors, such as the willingness to lead, take precedence in an environment where college teams are required to work collaboratively on projects with strong financial incentives.

4.5 Free Riding

The analysis so far indicates that female-majority teams benefit from agency in choosing their own leaders, but male-majority teams do not. To better understand this phenomenon, I further examine another source of variation. Specifically, I analyze whether the patterns of free-riding differ by team gender composition and leadership structure. I construct an indicator for *free rider* that is equal to 1 if both peers of a team member report that the individual contributed nothing (or 0%) to the project and 0 otherwise. This measure there-

fore captures cases of complete non-participation as perceived unanimously by teammates. The results of free-riding are presented in Table 6.

Table 6: Free-riding by Team Gender Composition and Leader Assignment

	Free rider (=1)
Panel A: Within Leadership Structure	
Nominated leader: Female-majority (=1)	-0.016
	(0.027)
Assigned leader: Female-majority (=1)	0.030
	(0.019)
Panel B: Within Gender Composition	
Female-majority teams: Assigned (=1)	-0.018
	(0.024)
Male-majority teams: Assigned (=1)	-0.064***
	(0.023)
Observations (individuals)	569

Notes: Each cell reports the coefficient from a separate OLS regression. Heteroskedasticity-robust standard errors clustered at the team level are in parentheses. Dependent variable is an indicator for free-riding (0/1). Intercepts omitted. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Panel A compares female- and male-majority teams within each leadership structure. The estimates show no statistically significant differences between female- and male-majority teams when leaders are nominated: female-majority teams report 1.6 pp fewer free-riders, but not significantly ($p > 0.10$). However, descriptive evidence helps to clarify the pattern. Among teams with nominated leaders, men and women in female-majority teams free-ride at roughly equal rates, whereas in male-majority teams all free-riders are men. This aligns with the idea that in male-majority settings, men who are not chosen as leaders may be less willing to cooperate or contribute. When leaders are externally assigned, the difference in free-riding between female- and male-majority teams remains small and insignificant.

Panel B compares team with the same gender composition. Among female-majority teams, assigning a leader has little effect on free-riding. In contrast, among male-majority teams, assigning a leader based on emotional intelligence significantly reduces free-riding by 6.4 pp ($p < 0.01$). This suggests that in male-majority teams, externally assigning a leader may mitigate within-team conflict or disengagement that arises when peers compete for leadership.

These findings support the mechanism proposed earlier: in male-majority teams that choose their own leaders, male teammates who are not selected appear more likely to withdraw effort, consistent with prior evidence that male-to-male competition can undermine cooperation (Sutter et al., 2009; Eckel & Grossman, 2001; Castillo et al., 2010). In contrast, teams with female-majority show more balanced participation and are less affected by the

leadership selection process.

4.6 Contribution and Evaluation of Women Leaders

Prior literature at the intersection of gender and leadership suggests that a team's gender composition and how leaders are selected shape the representation and role of women leaders in important ways. Certain settings in particular–such as male-dominated fields–are detrimental to women's leadership. In previous sections, I reported a somewhat surprising finding: irrespective of a team's gender composition and leadership selection process, women leaders are well-represented and are not marginalized from leadership roles. However, contingent on women assuming leadership roles, are there differences between how women and men leaders contribute and are evaluated? The answers to this question are presented in Table 7.

Table 7: Differences in Perceived Performance of Male & Female Leaders

	All teams	Female-majority teams	Male-majority teams
Contribution (peer-reported)	5.608**	-0.373	11.710***
	(2.228)	(3.407)	(3.275)
TL effectiveness (peer, z)	0.038	-0.214	0.223
	(0.113)	(0.198)	(0.160)
TL quality index (peer)	0.088	-0.259*	0.201
	(0.103)	(0.145)	(0.172)
Observations (individuals)	371	199	172

Notes: Entries are coefficients on *Female Leader* (vs male leader) from separate OLS regressions. Standard errors clustered by team are in parentheses. Columns split the sample by team gender composition. Intercepts omitted. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

I find that on average, women leaders are perceived to contribute more than male leaders (+5.6 pp, p-value < 0.05). This average masks strong heterogeneity by team gender composition: in male-majority teams, the contribution gap is large and precisely estimated (+11.7 pp, p-value < 0.01), whereas in female-majority teams it is near zero and imprecise. In female-majority teams, female leaders contribute about the same as male leaders but receive slightly lower peer ratings: the leadership quality index is -0.26σ and marginally significant (p-value < 0.10), and effectiveness is also negative but imprecise. In male-majority teams, the pattern changes in contribution—female leaders contribute more (+11.71 pp, p-value < 0.01), but their ratings do not rise accordingly. Although their leader effectiveness (+0.22σ) and leader quality (+0.20σ) ratings are positive, they are statistically indistinguishable from zero. In other words, female leaders appear to get no ratings premium when they contribute substantially more in male-majority teams, and they may face a modest ratings penalty in female-majority teams even when their contribution is comparable.

In Appendix Table A3, I also report within-team rating gaps by rater gender using team fixed effects (female rater minus male rater). Essentially, within the same team, do male and female teammates rate their leaders differently based on the gender of the team leader? I find that across both male-led and female-led teams, these gaps in ratings are small and statistically indistinguishable from zero. In other words, I find no evidence of same-sex favoritism or cross-sex penalties in peer scoring. Taken together with the main table, this suggests that the patterns I document–female leaders contributing more overall (especially in male-majority teams) but not receiving higher leadership ratings, and a modest ratings shortfall for female leaders in female-majority teams despite similar contribution–are not driven by rater-gender bias. Rather, these patterns might indicate that peers, regardless of their gender, might have higher expectations for how much female leaders should contribute to be considered as effective as male leaders. Alternatively, the divergence might also reflect differences in how teams translate observed contributions of leaders into evaluations of leaderships across contexts.

4.7 Robustness Checks

To assess the robustness of the main results, I examine whether the findings are sensitive to sample restrictions and alternative model specifications. The main specification, presented in Table 2, includes year-fixed effects to account for the randomization of teams within academic cohorts.

First, I re-estimate the main outcomes after excluding two teams whose gender composition does not clearly align with the predominant male- or female-majority categories used in the design (the two teams are gender-balanced but coded as female-majority in the main analysis). Table A4 shows that the results remain nearly identical to the baseline specification: female-majority teams with nominated leaders continue to have the highest submission rates and overall performance. The estimated effects are consistent in magnitude and significance in all four outcomes, indicating that the main findings are not driven by these atypical teams.

Next, I assess whether the results are robust to omitting year fixed effects. Although randomization occurred within each academic cohort, removing year fixed effects allows me to check that the results are not solely an artifact of controlling for between-year variation. Table A5 reports the estimates from this specification. The results remain substantively unchanged: female-majority teams with nominated leaders continue to outperform all other groups across all outcomes. They are about 20 percentage points more likely to submit their projects ($p < 0.01$) and achieve total project scores roughly 0.51σ higher than male-majority teams with nominated leaders ($p < 0.01$). Among female-majority teams, those with assigned leaders continue to perform significantly worse (-0.38σ, $p < 0.05$). Across all

four outcomes, including the share of teams that reached the finalist stage, the estimates are nearly identical in magnitude and direction to the main specification, confirming that the findings are not dependent on the inclusion of year fixed effects.

Finally, I re-estimate the main specification including additional individual-level controls for socioeconomic status and ability. Specifically, I control for indicators of caste group, household income, leadership preferences, cognitive ability, and emotional intelligence. These variables account for background characteristics that could potentially influence student and team performance. Table A6 presents the results. The inclusion of these controls has little effect on the estimates: female-majority teams with nominated leaders continue to show significantly higher completion rates and overall performance–approximately 0.48σ higher than male-majority teams with nominated leaders ($p < 0.01$)–and female-majority teams with assigned leaders continue to perform worse (–0.36σ, $p < 0.05$). The coefficients remain similar in size and significance across all outcomes, including the probability of reaching the finalist stage.

Overall, the results across these alternative samples and specifications demonstrate that the main findings are highly robust. The advantage of female-majority teams with nominated leaders persists across all models and is not driven by cohort-level differences, outliers, or differences in observable student characteristics.

4.8 Attrition

There is no attrition for any of the main performance outcomes, since these are collected at the team-level. As such, I have outcome data for all the teams that were part of the study. In addition to team-level data, I also collected individual-level data via the endline survey–these data were then aggregated at the team-level, specifically for outcomes related to teamwork and leadership. Although the overall response rate for the endline is high, (93.3%), I check for differential attrition between treatment groups with 'missingness' as the outcome. I report these results in Table A7. I find differential attrition between teams with nominated leaders. Specifically, male majority teams that nominated their own leader have 7.1pp higher attrition than female-majority teams with a nominated leader. I find no evidence of differential attrition between other treatment groups.

Using Lee (2009) bounds, I re-estimate the differences in teamwork and leadership outcomes between female- and male-majority teams with nominated leaders to assess whether the main results are sensitive to differential attrition. The bounded estimates, reported in Table A8, are very similar to the original coefficients, suggesting that the main findings for teamwork and leadership are robust to correcting for potential selection due to attrition.

5 Discussion and Conclusion

This field experiment in India examines the effect of a team's gender composition and method of leadership selection on performance, teamwork, and the representation and effectiveness of leaders. I show that when female-majority teams have the agency to choose their own leaders, they outperform male-majority teams (regardless of leadership structure) and also outperform female-majority teams with leaders assigned based on emotional intelligence. In general, female-majority teams demonstrate better engagement and teamwork, which might reflect more prosocial behavior when women collaborate. The study also shows that the process of leadership selection in mixed-gender teams matter, especially when women are in a majority. The choice of leadership selection impacts the perceived effectiveness, quality, and contribution of leaders, and closely mirrors improvements in a team's performance.

With regards to the role of women in teams, some results of this study complement findings in other settings–more cooperation among women, lower evaluation of women leaders, and the legitimizing effect of democratic selection for women leaders. However, other findings depart from earlier studies in subtle but important ways. For example, I find no evidence of the under-representation of women in leadership roles, regardless of whether they're nominated by peers or assigned based on emotional intelligence, or whether they're in male or female-majority teams. On the surface, this might appear somewhat puzzling. Even in more gender-equal Western contexts, women leaders are under-represented in male-dominated STEM fields. How is it that in an arguably less gender-equal context of India, women are able to assume leadership roles?

One reason could be women's higher preference for leadership compared to men. Expressing this preference during team deliberations might help them assume leadership roles. Additional evidence from the baseline suggests that women assume leadership roles despite gender bias. For example, men in the study report higher levels of hostile sexism compared to women[8]. In one of the questions related to women in the workplace, slightly more than 50% men agree or strongly agree with the statement *Many women ask for special treatment at work and call it "equality"* compared to 21% women. This aligns with the type of prejudices men might have about women in gender-segregated work environments. However, the election of women leaders, especially in male-majority teams, might suggest that women are able to successfully negotiate these prejudices and assume leadership roles.

There could also be differences in the broader context of developing countries that might shape women's preferences. For instance, women might not take their education–and,

[8]Adapted from the hostile sexism inventory in Glick and Fiske (1996). The inventory consists of a 3-item index, which asked participants to respond on a 5-point Likert scale to questions related to whether women get upset too easily; whether women ask for special treatment at work; and whether women make their work problems seem bigger than they are. Men report 0.46σ (p-value < 0.001) higher hostile sexism than women.

more generally, their freedom–for granted. Evidence from developing countries suggests that women's educational decisions involve inter-generational intrahousehold bargaining–specifically, in patriarchal cultures that value obedience, women are expected to negotiate with their parents to pursue higher levels of education (Ashraf et al., 2020). Early qualitative work from my field site indicate patterns of intrahousehold bargaining. Women at the college encounter opposition from their families to entering the labor market after graduation, especially since the labor market in technology and engineering would most likely involve moving away from home, living independently, and being exposed to modern mores. Most women in this context prepare themselves to negotiate with their parents to 'allow' them to enter the labor market. These dynamics for college-going women in India are different from their counterparts in Western countries, and could shape women's motivation and their willing to participate and lead during college in unexpected ways.

Finally, as with any experimental study, external validity is a natural concern. I highlight certain features of the study and the sample to help alleviate those concerns. First, the study's gender mix (50.8% women, 49.2% men) is close to the composition of Indian higher education: women constitute roughly 48% of total higher-education enrollment and about 43% of STEM enrollment in 2021–2022, with national trends showing a gradual rise in women's participation.[9] Second, the institutional context is typical: roughly 60% of colleges in India are located in rural areas, which my field site reflects. Third, the team task mirrors the environments many college graduates enter—mixed-gender, project-based teams—so the core study design (who teams are composed of and how leaders are selected) maps directly onto early-career work settings. While generalization beyond short-horizon student teams—especially to higher-stakes professional settings—should be made with caution, the study provides novel evidence from a developing country context of how team gender composition and leadership selection interact to shape collaboration and performance.

[9]Based on the 'All India Survey on Higher Education', AISHE 2021-2022. While more recent data are unavailable, trends point to a gradual increase in the enrollment of women.

Tables

Table 3: Teamwork by Gender Composition and Leadership Structure

	Team effectiveness	Team equity
Panel A: Within Leadership Structure		
Nominated leader: Female-majority (=1)	0.336**	0.318**
	(0.134)	(0.125)
Assigned leader: Female-majority (=1)	0.178	0.209*
	(0.138)	(0.127)
Panel B: Within Gender Composition		
Female-majority teams: Assigned (=1)	-0.144	-0.121
	(0.127)	(0.103)
Male-majority teams: Assigned (=1)	0.014	-0.012
	(0.137)	(0.141)
Observations (individuals)	569	569

Notes: Each cell reports the coefficient on the listed indicator from a separate OLS regression with year fixed effects. "Team Effectiveness" and "Team Equity" are both indices constructed from five questions each in the endline survey, standardized to have a mean of 0 and standard deviation of 1. Heteroskedasticity-robust standard errors, clustered at the team level, are in parentheses. Intercepts omitted. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Leadership by Gender Composition and Leadership Structure

	Leader effectiveness	Leader quality	Leader contribution (%)
Panel A: Within Leadership Structure			
Nominated leader: Female-majority (=1)	0.257*	0.415***	4.359
	(0.144)	(0.139)	(2.952)
Assigned leader: Female-majority (=1)	-0.060	0.075	1.379
	(0.173)	(0.150)	(3.347)
Panel B: Within Gender Composition			
Female-majority teams: Assigned (=1)	-0.430***	-0.332**	-8.349***
	(0.160)	(0.131)	(2.986)
Male-majority teams: Assigned (=1)	-0.112	0.009	-5.369*
	(0.150)	(0.153)	(3.057)
Observations	371	371	371

Notes: This only includes the subsample of individuals who are not team leaders. Peers–who are not team leaders–evaluate team leaders on these three outcomes. Each cell reports the coefficient on the listed indicator from a separate OLS regression with year fixed effects. Heteroskedasticity-robust standard errors, clustered at the team level, are in parentheses. Intercepts omitted. "Leader Effectiveness" is based on a survey question that asked respondents to rate their team leader on a scale of 1-10, which here is standardized to have a mean of 0 and a standard deviation of 1. "Leader Quality" is an index based on five survey questions that capture various components of leadership such as coordination, motivation, conflict resolution, task distribution, and openness to team members' ideas. It is standardized to have a mean of 0 and a standard deviation of 1. "Leader contribution" is expressed as a percentage of total team contributions. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

References

Aparicio Fenoll, A. and S. Zaccagni (2022, December). Gender mix and team performance: Differences between exogenously and endogenously formed teams. *Labour Economics 79*, 102269.

Apesteguia, J., G. Azmat, and N. Iriberri (2012a, January). The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field. *Management Science 58*(1), 78–93. Publisher: INFORMS.

Apesteguia, J., G. Azmat, and N. Iriberri (2012b, January). The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field. *Management Science 58*(1), 78–93. Publisher: INFORMS.

Ashraf, N., N. Bau, C. Low, and K. McGinn (2020, May). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics 135*(2), 1095–1151.

Berge, L. I. O., K. S. Juniwaty, and L. H. Sekei (2016, November). Gender composition and group dynamics: Evidence from a laboratory experiment with microfinance clients. *Journal of Economic Behavior & Organization 131*, 1–20.

Born, A., E. Ranehill, and A. Sandberg (2022, March). Gender and Willingness to Lead: Does the Gender Composition of Teams Matter? *The Review of Economics and Statistics 104*(2), 259–275.

Brandts, J., D. J. Cooper, and R. A. Weber (2015, November). Legitimacy, Communication, and Leadership in the Turnaround Game. *Management Science 61*(11), 2627–2645. Publisher: INFORMS.

Castillo, M., R. Petrie, M. Torero, and L. Vesterlund (2013, March). Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics 99*, 35–48.

Chemin, M. (2021, June). Does appointing team leaders and shaping leadership styles increase effort? Evidence from a field experiment. *Journal of Economic Behavior & Organization 186*, 12–32.

Chen, J. and D. Houser (2019, December). When are women willing to lead? The effect of team gender composition and gendered tasks. *The Leadership Quarterly 30*(6), 101340.

De Paola, M., F. Gioia, and V. Scoppa (2022, September). Female leadership: Effectiveness and perception. *Journal of Economic Behavior & Organization 201*, 134–162.

Deming, D. J. (2017, November). The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Deserranno, E., M. Stryjan, and M. Sulaiman (2019, October). Leader Selection and Service Delivery in Community Groups: Experimental Evidence from Uganda. *American Economic Journal: Applied Economics 11*(4), 240–267.

Eckel, C., L. Gangadharan, P. J. Grossman, and N. Xue (2021, July). Chapter 7: The gender leadership gap: insights from experiments. Section: A Research Agenda for Experimental Economics.

Edin, P.-A., P. Fredriksson, M. Nybom, and B. Öckert (2022, April). The Rising Return to Noncognitive Skill. *American Economic Journal: Applied Economics 14*(2), 78–100.

Englmaier, F., S. Grimm, D. Grothe, D. Schindler, and S. Schudy (2024, August). The Effect of Incentives in Nonroutine Analytical Team Tasks. *Journal of Political Economy 132*(8), 2695–2747. Publisher: The University of Chicago Press.

Gangadharan, L., T. Jain, P. Maitra, and J. Vecci (2016, November). Social identity and governance: The behavioral response to female leaders. *European Economic Review 90*, 302–325.

Glick, P. and S. T. Fiske (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology 70*(3), 491–512. Place: US Publisher: American Psychological Association.

Hardt, D., L. Mayer, and J. Rincke (2025, May). Who Does the Talking Here? The Impact of Gender Composition on Team Interactions. *Management Science 71*(5), 4131–4152. Publisher: INFORMS.

Hoogendoorn, S., H. Oosterbeek, and M. van Praag (2013, July). The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment. *Management Science 59*(7), 1514–1528. Publisher: INFORMS.

Karpowitz, C. F., S. D. O'Connell, J. Preece, and O. Stoddard (2024, September). Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams. *Journal of Political Economy 132*(9), 3077–3114. Publisher: The University of Chicago Press.

Lamiraud, K. and R. Vranceanu (2018, January). Group gender composition and economic decision-making: Evidence from the *Kallystée* business game. *Journal of Economic Behavior & Organization 145*, 294–305.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102. Publisher: [Oxford University Press, The Review of Economic Studies, Ltd.].

Levy, D. M., K. Padgitt, S. J. Peart, D. Houser, and E. Xiao (2011, January). Leadership, cheap talk and *really* cheap talk. *Journal of Economic Behavior & Organization 77*(1), 40–52.

Montes, F., R. C. Jimenez, and J.-P. Onnela (2018, October). Connected but segregated: social networks in rural villages. *Journal of Complex Networks 6*(5), 693–705.

Reuben, E. and K. Timko (2018, December). On the effectiveness of elected male and female leaders and team coordination. *Journal of the Economic Science Association 4*(2), 123–135.

Sutter, M., R. Bosman, M. G. Kocher, and F. v. Winden (2009, September). Gender pairing and bargaining—Beware the same sex! *Experimental Economics 12*(3), 318–331.

Truffa, F. and A. Wong (2025, July). Undergraduate Gender Diversity and the Direction of Scientific Research. *American Economic Review 115*(7), 2414–2448.

Vial, A. C., J. L. Napier, and V. L. Brescoll (2016, June). A bed of thorns: Female leaders and the self-reinforcing cycle of illegitimacy. *The Leadership Quarterly 27*(3), 400–414.

Weidmann, B. and D. J. Deming (2021). Team Players: How Social Skills Improve Team Performance. *Econometrica 89*(6), 2637–2657. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18461.

Weidmann, B., J. Vecci, F. Said, D. J. Deming, and S. R. Bhalotra (2024, July). How Do You Find a Good Manager?

Weidmann, B. and Y. Xu (2024, September). PAGE: A Modern Measure of Emotion Perception for Teamwork and Management Research. arXiv:2410.03704 [cs].

Yang, Y., T. Y. Tian, T. K. Woodruff, B. F. Jones, and B. Uzzi (2022, September). Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences 119*(36), e2200841119. Publisher: Proceedings of the National Academy of Sciences.

Appendix

A. Additional Tables and Figures

Table A1: **Covariate Balance by Gender and Team Composition**

	Female				**Male**			
Variable	Maj. Female	Maj. Male	Diff	p-value	Maj. Female	Maj. Male	Diff	p-value
Hindu	0.958	0.979	-0.021	0.317	0.917	0.932	-0.015	0.698
Low income	0.701	0.792	-0.091	0.203	0.574	0.620	-0.046	0.268
Mother completed school	0.243	0.188	0.055	0.249	0.343	0.292	0.051	0.299
PAGE score	14.893	14.667	0.226	0.320	14.500	14.708	-0.208	0.195
Raven score	5.332	5.146	0.186	0.541	5.500	5.740	-0.240	0.257
Upper caste	0.272	0.263	0.009	0.981	0.333	0.279	0.054	0.403
Assigned leader	0.500	0.490	0.010	0.766	0.509	0.490	0.020	0.704
Preference for leading	7.173	6.938	0.235	0.399	6.222	5.911	0.311	0.543
Observations	214	96			108	192		

Notes: The table reports mean values of baseline covariates by student gender and team gender composition. Differences are from t-tests with year fixed effects. Stars to indicate statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: **Covariate Balance by Team Leader Selection Groups**

Variable	**Mean (Nominated)**	**Mean (Assigned)**	**Difference**	***p*-value**
Female	0.507	0.508	0.001	0.969
Female-majority team	0.520	0.535	0.015	0.831
Hindu	0.955	0.936	−0.019	0.304
Low income	0.678	0.657	−0.021	0.587
Mother completed school	0.274	0.261	−0.013	0.716
PAGE score	14.804	14.673	−0.131	0.697
Raven score (std.)	0.109	−0.110	−0.219	0.119
Team lead pref. (std.)	0.022	−0.022	−0.045	0.751
Upper caste	0.287	0.281	−0.006	0.866
Year 1	0.618	0.634	0.016	0.815
Year 2	0.275	0.267	−0.007	0.909
Year 3	0.108	0.099	−0.009	0.837
Observations	*102*	*101*		

Notes: The table reports group means and t-test p-values comparing teams with assigned versus nominated leaders. Stars denote significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Joint F-test p-value for all covariates: 0.541.

Table A3: Within-team rating gap (Female rater – Male rater) by leader sex

	Male leader	Female leader	Diff-in-gaps (F–M)
TL effectiveness (peer, z)	0.018	0.227	0.210
	(0.278)	(0.197)	(0.341)
TL quality index (peer)	0.293	0.061	-0.231
	(0.280)	(0.221)	(0.357)

Notes: Entries are within-team rating gaps (Female rater – Male rater) from OLS with team fixed effects; standard errors clustered by team in parentheses. The "Female leader" column equals the male-leader gap plus the FemaleLeader×FemaleRater interaction. "Diff-in-gaps (F–M)" is the interaction coefficient. Intercepts omitted.

Table A4: Performance by Gender Composition and Leadership Structure (Drop tie teams)

	Submitted	Project score (z)	Project score (z), 0 for non-submit	Finalists (%)
Panel A: Within Leadership Structure				
Nominated leader: Female-majority (=1)	0.211***	0.170	0.510***	0.123**
	(0.080)	(0.221)	(0.190)	(0.056)
Assigned leader: Female-majority (=1)	-0.014	0.279	0.098	0.039
	(0.090)	(0.262)	(0.213)	(0.066)
Panel B: Within Gender Composition				
Female-majority teams: Assigned (=1)	-0.174**	-0.041	-0.374**	-0.020
	(0.074)	(0.182)	(0.176)	(0.063)
Male-majority teams: Assigned (=1)	0.050	-0.150	0.038	0.064
	(0.092)	(0.278)	(0.216)	(0.053)
Observations (teams)	201	155	201	201

Notes: Each cell reports the coefficient on the listed indicator from a separate OLS regression with year fixed effects. Heteroskedasticity-robust standard errors clustered at the team level are in parentheses. Intercepts omitted. "Finalists" is an indicator for Top-20 teams. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5: Performance by Gender Composition and Leadership Structure (Without YFE)

	Submitted	Project score (z)	Project score (z), 0 for non-submit	Finalists (%)
Panel A: Within Leadership Structure				
Nominated leader: Female-majority (=1)	0.212***	0.159	0.509***	0.090*
	(0.077)	(0.219)	(0.185)	(0.054)
Assigned leader: Female-majority (=1)	-0.016	0.268	0.087	0.005
	(0.088)	(0.252)	(0.208)	(0.062)
Panel B: Within Gender Composition				
Female-majority teams: Assigned (=1)	-0.178**	-0.028	-0.383**	-0.020
	(0.073)	(0.181)	(0.174)	(0.063)
Male-majority teams: Assigned (=1)	0.051	-0.137	0.039	0.066
	(0.092)	(0.281)	(0.217)	(0.053)
Observations (teams)	203	156	203	203

Notes: Each cell reports the coefficient on the listed indicator from a separate OLS regression without year fixed effects. Heteroskedasticity-robust standard errors clustered at the team level are in parentheses. Intercepts omitted. "Finalists" is an indicator for Top-20 teams. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A6: Performance by Gender Composition and Leadership Structure (With Controls)

	Submitted	Project score (z)	Project score (z), 0 for non-submit	Finalists (%)
Panel A: Within Leadership Structure				
Nominated leader: Female-majority (=1)	0.206***	0.112	0.475**	0.126**
	(0.079)	(0.204)	(0.186)	(0.054)
Assigned leader: Female-majority (=1)	-0.007	0.261	0.100	0.036
	(0.088)	(0.262)	(0.208)	(0.064)
Panel B: Within Gender Composition				
Female-majority teams: Assigned (=1)	-0.165**	-0.039	-0.356**	-0.016
	(0.071)	(0.179)	(0.170)	(0.060)
Male-majority teams: Assigned (=1)	0.048	-0.189	0.019	0.074
	(0.092)	(0.272)	(0.215)	(0.053)
Observations (teams)	203	156	203	203

Notes: Each cell reports the coefficient on the listed indicator from a separate OLS regression that includes controls. Heteroskedasticity-robust standard errors clustered at the team level are in parentheses. Intercepts omitted. "Finalists" is an indicator for Top-20 teams. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A7: Attrition by Team Gender Composition and Leadership Structure

	Missing (=1)
Panel A: Within Leadership Structure	
Nominated leader: Female-majority (=1)	-0.071**
	(0.034)
Assigned leader: Female-majority (=1)	0.011
	(0.027)
Panel B: Within Gender Composition	
Female-majority teams: Assigned (=1)	0.030
	(0.025)
Male-majority teams: Assigned (=1)	-0.052
	(0.036)
Observations	569

Notes: Each cell reports the coefficient from a separate OLS regression. Heteroskedasticity-robust standard errors clustered at the team level are in parentheses. The dependent variable is an indicator for missingness (1 = missing). Intercepts omitted. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A8: Lee (2009) bounds for nominated-leader sample (female − male)

Outcome	Lower bound [95% CI]	Upper bound [95% CI]	Sel. rate (F)	Sel. rate (M)	Trimmed group	Trim fraction α
Team effectiveness (index)	0.345 [0.147, 0.552]	0.544 [0.254, 0.772]	0.963	0.891	Female-majority	0.074
Team equity (index)	0.266 [0.065, 0.485]	0.483 [0.292, 0.661]	0.963	0.891	Female-majority	0.074
Leader effectiveness (std)	0.241 [0.057, 0.433]	0.379 [0.157, 0.562]	0.963	0.891	Female-majority	0.074
Leadership quality (index)	0.340 [0.151, 0.525]	0.538 [0.258, 0.743]	0.963	0.891	Female-majority	0.074

Notes: Bounds are for the nominated-leader subsample; contrast is female-majority minus male-majority. 95% CIs are from a cluster bootstrap by team. The selection indicator equals 1 when the outcome is observed (non-missing). The higher-retention arm is trimmed to match the other arm.

Table A9: Balance (Male-majority teams): Assigned vs Nominated leader

	Group means		Test	
Variable	Mean: Nominated	Mean: Assigned	Diff (A–N)	p-value
Student & household				
Female (=1)	0.333	0.333	0.000	0.330
Hindu (=1)	0.952	0.940	-0.013	0.627
Low income (=1)	0.728	0.624	-0.104*	0.069
Mother completed school (=1)	0.252	0.262	0.011	0.831
Aptitude & preferences				
Page short total score	14.653	14.745	0.092	0.855
Raven total score (z)	0.163	−0.077	-0.240	0.233
Team leader preference (z)	−0.164	−0.176	-0.012	0.952
Upper caste (=1)	0.269	0.280	0.011	0.830
Cohort (dummies)				
Year 1 (=1)	0.776	0.787	0.012	0.891
Year 2 (=1)	0.204	0.191	-0.013	0.879
Year 3 (=1)	0.020	0.021	0.001	0.977
Observations (teams)	49	47	–	–

Notes: Entries are means in each group. "Diff (A–N)" reports Assigned minus Nominated with significance stars. Two-sided p-values from OLS of each covariate on an Assigned indicator. Joint F-test (all covariates): $p = 0.656$. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A10: Balance (Female-majority teams): Assigned vs Nominated leader

	Group means		Test	
Variable	Mean: Nominated	Mean: Assigned	Diff (A–N)	p-value
Student & household				
Female (=1)	0.668	0.667	-0.002	0.324
Hindu (=1)	0.958	0.936	-0.022	0.397
Low income (=1)	0.632	0.673	0.041	0.452
Mother completed school (=1)	0.294	0.269	-0.025	0.628
Aptitude & preferences				
Page short total score	14.943	14.635	-0.309	0.500
Raven total score (z)	0.059	−0.093	-0.152	0.445
Team leader preference (z)	0.194	0.161	-0.033	0.865
Upper caste (=1)	0.303	0.282	-0.021	0.681
Cohort (dummies)				
Year 1 (=1)	0.472	0.500	0.028	0.774
Year 2 (=1)	0.340	0.346	0.007	0.944
Year 3 (=1)	0.189	0.154	-0.035	0.640
Observations (teams)	53	54	–	–

Notes: Entries are means in each group. "Diff (A–N)" reports Assigned minus Nominated with significance stars. Two-sided p-values from OLS of each covariate on an Assigned indicator. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B. Photographs of Competition and App Examples

Figure B1: In-person Competition Featuring the Top-20 Teams


राज एक्सप्रेस
नवीन बार्चे
जवाहरलाल इंस्टीट्यूट ऑफ टेक्नालॉजी बोरावां में, हावर्ड यूनिवर्सिटी की रिसर्च द्वारा आयोजित...
रूरल इण्डिया एप आईडिया चैलेंज में 203 टीमों के 609 इंजीनियर्स ने लिया हिस्सा
प्रथम स्थान के विजेता के 15 हजार, द्वितीय स्थान 12 हजार एवं तृतीय स्थान प्राप्त करने वाली टीम ने जीता 10 हजार का इनाम
20 टीमों के 60 इंजीनियर्स ने अंतिम चरण की प्रतियोगिता में जीते पुरस्कार
पूर्व केन्द्रीय कृषि राज्यमंत्री अरूण यादव ने विजेताओं को दी शुभकामनाएं
जवाहरलाल इंस्टीट्यूट ऑफ टेक्नालॉजी बोरावां में शनिवार 12 जुलाई को हावर्ड यूनिवर्सिटी की रिसर्च टीम द्वारा इंजीनियरिंग पाठ्यक्रमों के विद्यार्थियों के लिए रूरल इण्डिया एप आईडिया चैलेंज प्रतियोगिता का आयोजन किया गया।
इस प्रतियोगिता की शुरूआत मई 2025 में एक बेसलाईन सर्वे से हुई थी । जिसमें बी.टेक. और एम.सी.ए. पाठ्यक्रमों कुल 609 विद्यार्थियों ने भाग लिया था । तत्पश्चात उन्हें 3-3 के समूह में कुल 203 टीमों में विभाजित किया गया । हावर्ड यूनिवर्सिटी के रिसर्चर आशुतोष बुराड़िया ने जानकारी देते हुए बताया कि हमारी रिसर्च टीम वर्क और लीडरशिप पर केन्द्रित है । इसके साथ ही विद्यार्थियों को तकनीकी क्षमताओं, टीम बिल्डिंग, लीडरशिप, एप्लीकेशन डेवलपमेंट एवं तकनीकी ज्ञान का आंकलन कर सही दिशा में उपयोग पर केन्द्रित करना है । श्री बुराड़िया ने आगे कहा कि रूरल इण्डिया पर आधारित प्रोजेक्ट को प्रेजेन्टेशन के माध्यम से सभी टीमों द्वारा तैयार कर जमा करना था । हमारी टीम द्वारा इनमें से 20 यूनिक प्रोजेक्ट का चयन किया गया जिसका आज शनिवार को जेआयटी बोरावां में आयोजित प्रतियोगिता में ऑनलाईन प्रेजेन्टेशन के माध्यम से चयन किया गया । इस प्रक्रिया में रिसर्च टीम का हिस्सा रही रिसर्च असिस्टेंट सिद्रा अली एवं 2 निर्णायक तेजस एवं मृदुल ने बेस्ट 3 आईडियाज का ऑनलाईन चयन किया गया ।
प्लेसमेंट अधिकारी एवं कार्यक्रम संयोजक प्रो. नितिन नामदेव ने जानकारी देते हुए बताया कि प्रथम पुरस्कार संस्था की बी.टेक.प्रथम वर्ष की टीम उम्मेहानी , कार्तिकेय और अमन को रू 15 हजार नगद प्रदान किये गये । द्वितीय स्थान पर प्रथम वर्ष
शुरू होने से यात्रियों की परेशानी समाप्त होगी।
जवाहरलाल इंस्टीट्यूट ऑफ टेक्नालॉजी बोरावां में हावर्ड यूनिवर्सिटी
रूरल इण्डिया एप आईडिया चैलेंज में 203 टीमों के 609 इंजीनियर्स ने लिया हिस्सा
नवभारत न्यूज
खरगोन, निप्र। जवाहरलाल इंस्टीट्यूट ऑफ टेक्नालॉजी बोरावां में शनिवार 12 जुलाई को हावर्ड यूनिवर्सिटी की रिसर्च टीम द्वारा इंजीनियरिंग पाठ्यक्रमों के विद्यार्थियों के लिए रूरल इण्डिया एप आईडिया चैलेंज प्रतियोगिता का आयोजन किया गया।
पूर्व केन्द्रीय कृषि राज्यमंत्री ने विजेताओं को दी बधाई
पूर्व केन्द्रीय कृषि राज्यमंत्री अरूण यादव ने अपने संदेश में संस्था में शनिवार को आयोजित रूरल इण्डिया एप आईडिया चैलेंज प्रतियोगिता के विजेताओं को बधाई एवं शुभकामनाएँ देते हुए कहा कि आज का दौर आधुनिक सोच को क्रियान्वित करने का युग है । हमारे विद्यार्थियों ने एप्लीकेशन टूल्स का उपयोग करते हुए शिक्षा, कृषि, प्रबंधन, स्मार्ट वर्क, डाटाबेस के प्रोजेक्ट के रूप में एप डिजाईन किये गये । सभी विजेताओं को मैं बधाई एवं शुभमानाएँ देता हूँ । आने वाला समय आपका है , आप सभी मेहनत व लगन से पढ़ाई कर संस्थान और अपने परिवार को गौरवान्वित करें । मैं हावर्ड यूनिवर्सिटी के रिसर्चर आशुतोष बुराड़िया एवं उनकी टीम को भी धन्यवाद करना चाहता हूँ कि आपने हमारे विद्यार्थियों को आधुनिक तकनीकी के क्षेत्र में एक नया मार्ग दिखाया है ।
की टीम प्रियंका नकुल और अंकित को राशि रू 12 हजार नगद प्रदान किये । तृतीय स्थान पर अक्षा, सुफियान और नैना की टीम को राशि रू 10 हजार नगद प्रदान किये । इसके साथ ही 17 टीमों को सांत्वना पुरस्कार के रूप में नगद 3 हजार प्रत्येक टीम को प्रदान किये । साथ ही हावर्ड यूनिवर्सिटी की रिसर्च टीम द्वारा प्रतिभागिता प्रमाण पत्र भेजे जावेंगें ।
संस्थान के प्राचार्य डॉ. अतुल उपाध्याय , वरिष्ठ प्राध्यापकों एवं सभी विभागाध्यक्षों व शिक्षकों ने विजेता टीमों को पुरस्कार सौंपकर उनके उज्जवल भविष्य की कामनाएँ की । इस आयोजन में डॉ. संजय चौहान, डॉ. संतोष पँवार, डॉ. मनीष देशमुख , प्रो. कमलेश पाटीदार, प्रो. विनय देउलकर , प्रो. लवकेश पाटीदार का सहयोग सराहनीय रहा।

Figure B2: Local Media Coverage of the Competition

C. Survey Instruments

C.1 PAGE Assessment

I constructed a measure of emotional perceptiveness called PAGE (Perceiving AI Generated Emotions) with 28 items for India's context. The assessment is based on Weidmann and Xu (2024), who designed the assessment for a sample in the United States. PAGE was constructed using the prompts from Weidmann and Xu (2024) but using Indian faces and expression. For example, I used the following prompt to generate the face in Figure C3: "Create a hyper-realistic image of a young Indian women showing expression anxiety. Eyes looking sideways, frowned eyebrows, biting lips. Detailed skin texture and natural lighting. Wearing a white tshirt. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background." I adapted prompts for the original paper to generate a range expression for Indian faces. I provide examples of the PAGE assessment in this section.

In terms of its psychometric properties, the PAGE has a moderate Cronbach's Alpha of 0.70. The tests shows good psychometric properties in the other respects as well. For instance, the distribution of PAGE total scores approximately normal (M = 14.7, SD = 3.37, skew = -0.01, kurtosis = -0.18). Fewer than 1% of participants scored at either the minimum or maximum value, indicating no floor or ceiling effects. The distribution of PAGE scores are shown in Figure C1. I also conduct a principal component analysis of PAGE. The principal component analysis (PCA) indicated a strong first component (eigenvalue = 2.1), with a clear 'elbow' in the scree plot after the first component, as depicted in Figure C2. Subsequent components each accounted for substantially less variance. This pattern supports a unidimensional structure of the PAGE, consistent with the intended measure of a single underlying construct of emotional intelligence.

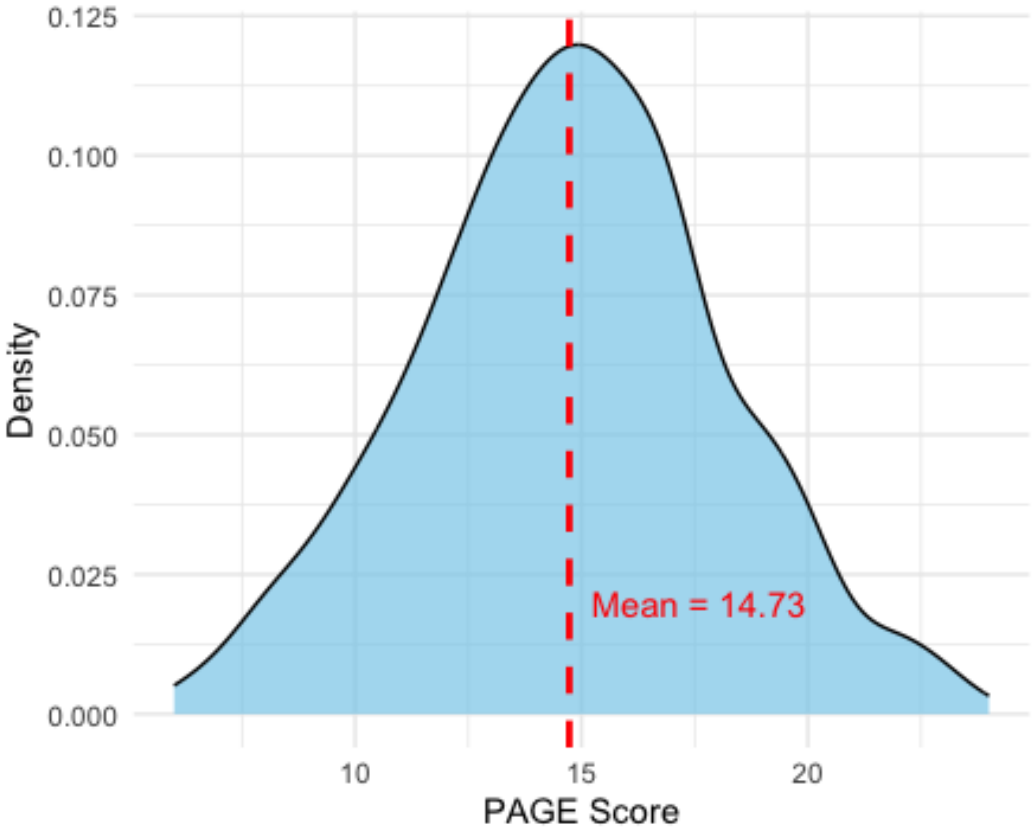


Figure C1: Distribution of PAGE Score

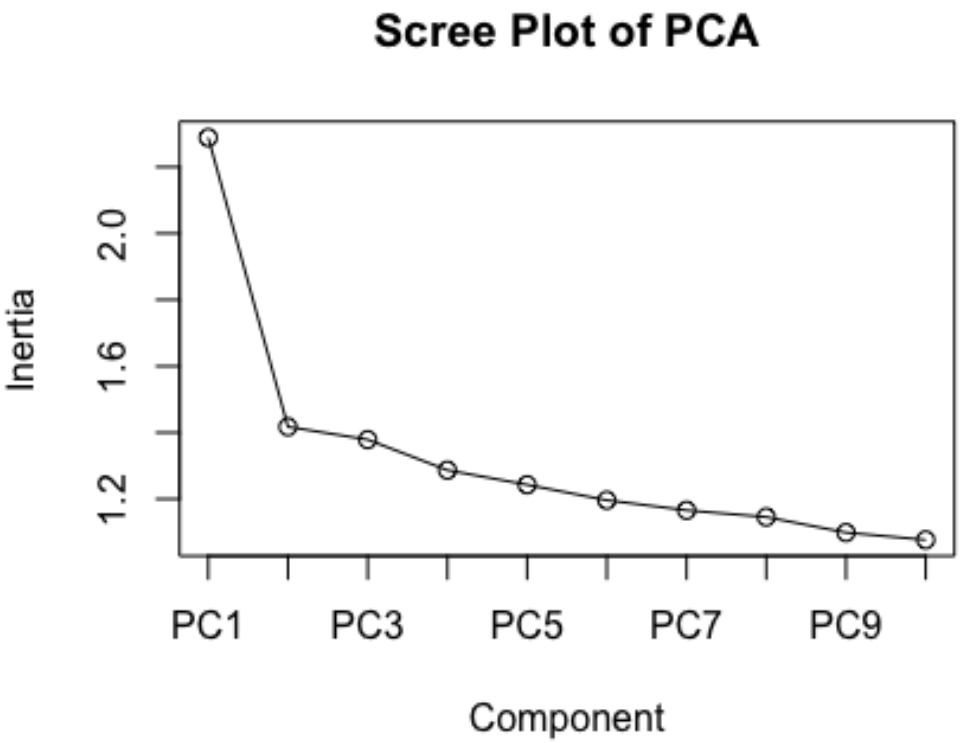
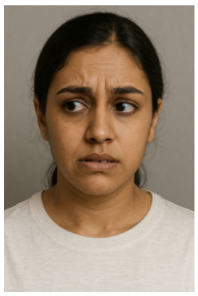


Figure C2: Scree Plot of PCA for PAGE

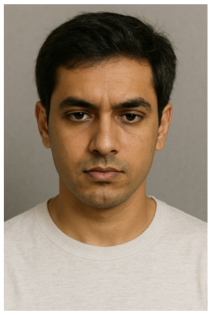
What is the person in the picture thinking or feeling? तस्वीर में दिख रहा व्यक्ति क्या सोच रहा है या क्या महसूस कर रहा है?



- ◯ Embarrassment - शर्मिंदगी
- ◯ Fear - डर
- ◯ Anxiety - चिंता
- ◯ Contemplation - चिंतन - deep thinking
- ◯ Contentment - संतुष्टि - satisfied
- ◯ Confusion - भ्रम - not understanding

Figure C3: PAGE Item Example 1

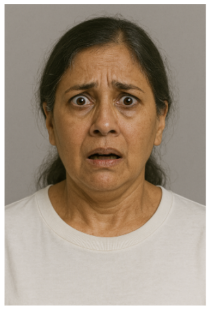
What is the person in the picture thinking or feeling? तस्वीर में दिख रहा व्यक्ति क्या सोच रहा है या क्या महसूस कर रहा है?



- ◯ Disappointment - निराशा
- ◯ Concentration - एकाग्रता - focused
- ◯ Confusion - भ्रम - not understanding
- ◯ Interest - रुचि - wanting to know more
- ◯ Doubt - संदेह - unsure
- ◯ Contentment - संतुष्टि - satisfied

Figure C4: PAGE Item Example 2

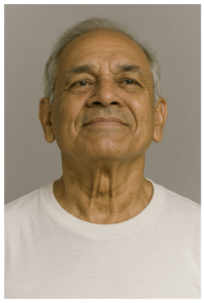
What is the person in the picture thinking or feeling? तस्वीर में दिख रहा व्यक्ति क्या सोच रहा है या क्या महसूस कर रहा है?



- ◯ Shame - लज्जा - guilty feeling
- ◯ Confusion - भ्रम - not understanding
- ◯ Fear - डर
- ◯ Anxiety - चिंता
- ◯ Awe - विस्मय - amazed
- ◯ Surprise - आश्चर्य

Figure C5: PAGE Item Example 3

What is the person in the picture thinking or feeling? तस्वीर में दिख रहा व्यक्ति क्या सोच रहा है या क्या महसूस कर रहा है?



- ◯ Awe - विस्मय - amazed
- ◯ Joy - हर्ष - very happy
- ◯ Amusement - मनोरंजन
- ◯ Contempt - तिरस्कार
- ◯ Pride - गर्व
- ◯ Interest - रुचि

Figure C6: PAGE Item Example 4

C.2 Survey Instruments: Teamwork and Leadership

Team Equity Index. To capture students' perceptions of fairness and inclusion within teams, I constructed a *Team Equity Index* using six Likert-scale items: (i) my ideas were valued by the team; (ii) I was able to contribute meaningfully to the project; (iii) workload was distributed equitably; (iv) all team members had equal opportunity to contribute; (v) members from all backgrounds were equally respected; and (vi) I felt comfortable and included within the team. Responses were coded from 1 (Strongly disagree) to 5 (Strongly agree). The index was derived from a one-factor exploratory factor analysis with regression-based scoring and standardized to mean zero and unit variance. Internal consistency is high ($\alpha = 0.80$).

Team Effectiveness Index. The *Team Effectiveness Index* measures perceived team functioning and performance. It is based on five items: (i) our team worked together effectively; (ii) our team made good use of time and resources; (iii) team members communicated well; (iv) our team resolved conflicts successfully; and (v) I am satisfied with our team's performance. Each item was rated from 1 (Strongly disagree) to 5 (Strongly agree). Scores were extracted using one-factor exploratory factor analysis and standardized (mean = 0, SD = 1). Reliability is high ($\alpha = 0.92$).

Leadership Index. To assess leadership quality, I constructed a *Leadership Index* from five items capturing coordination, motivation, conflict resolution, task distribution, and openness to team members' ideas. Each team member rated the team leader on these dimensions from 1 (Strongly disagree) to 5 (Strongly agree). A single-factor model was estimated using exploratory factor analysis to produce standardized factor scores. The index has strong internal consistency ($\alpha = 0.93$).

Leader Effectiveness Score. Participants were also asked "On a scale of 1-10, how would you rate your team leader's overall effectiveness?" 1 = not effective at all and 10 = extremely effective.

C.3 Survey Instruments: Project Evaluation Rubric

App Project Evaluation Rubric

Total Points: 20 (4 points per criterion)

A Problem Definition Clarity (4 points)

4 - Excellent

- Clearly identifies a specific rural problem with context or data/evidence.
- Shows understanding of target users and their basic challenges.
- Problem statement is rural-focused and well-defined.
- Demonstrates awareness that this is an issue worth addressing.

3 - Proficient

- Identifies a legitimate rural problem, though articulation may lack depth.
- Shows basic understanding of target users.
- Problem statement is clear enough to understand what they aim to solve and why it matters.
- Makes connection to the rural context, even if general.

2 - Developing

- Identifies a problem that could be relevant to rural areas, though the statement is general.
- Limited understanding of specific user needs or context.
- Problem statement lacks specificity but the core issue is understandable.
- Weak attempt to connect to the rural setting.

1 - Beginning

- Problem statement is unclear, overly broad, or not rural-specific.
- Little to no understanding of target users or context.
- Fails to articulate why this problem matters.
- No clear connection to rural challenges.

B Solution Feasibility, Especially in Rural Context (4 points)

4 - Excellent

- Shows strong understanding of rural constraints (connectivity, literacy, devices).
- Most features are realistic and appropriate for rural implementation.
- Includes adaptations for rural context (simple interface, offline features, etc.).
- Demonstrates awareness of the limitations of an app-based solution.
- Feasible with only minor adjustments needed.

3 - Proficient

- Shows basic awareness of rural constraints.
- Most features can be implemented with basic technology.
- Considers rural constraints in design.
- Demonstrates moderate awareness of limitations.
- Feasible with moderate adaptation.

2 - Developing

- Limited consideration of rural constraints, but not entirely unrealistic.
- Some features would be challenging to implement in rural settings.
- Would need significant adaptation, but core idea has potential.
- Shows superficial awareness of limitations.

1 - Beginning

- Little to no consideration of rural constraints.
- Features are unrealistic for rural deployment.
- Core idea lacks feasibility or potential.
- Shows no awareness of limitations.

C UI Simplicity & Accessibility (4 points)

4 - Excellent

- Interface is clean, simple, and user-friendly.
- Effective use of icons, visual cues, voice prompts, and local language elements.
- Clear and logical navigation with minimal cognitive load.
- Considers different literacy levels and accessibility.
- Interface would be usable by rural users.

3 - Proficient

- Interface is functional and intuitive.
- Includes visual elements supporting usability.
- Navigation is logical, though some tasks may require extra steps.
- Interface adequate for target users.

2 - Developing

- Interface functional but somewhat complex or unclear.
- Limited use of visual aids or accessibility features.
- Navigation possible but not intuitive.
- May challenge some users.

1 - Beginning

- Minimal or unclear information on user interaction.
- Heavy reliance on text without visual support.
- Navigation is confusing or illogical.
- No consideration of literacy or accessibility.

D Technical Innovation (within constraints) (4 points)

4 - Excellent

- Includes more than two creative or unique features.
- Thoughtful use of technologies (QR code, photo input, voice, SMS, offline mode, etc.).
- Demonstrates original thinking within no-code constraints.
- Shows clear understanding of technical possibilities.

3 - Proficient

- Has one or two creative or unique features.
- Uses available technologies effectively.
- Shows creativity, though somewhat conventional.
- Demonstrates good understanding of technical constraints.

2 - Developing

- Mostly conventional but functional.
- Limited creativity or feature variety.
- Some understanding of technical possibilities.

1 - Beginning

- Basic or conventional with no unique features.
- Minimal use of technology.
- Little evidence of innovation or understanding of constraints.

E Potential Social Impact (4 points)

4 - Excellent

- Demonstrates meaningful, measurable potential impact (possibly with numbers).
- Shows understanding of how the solution addresses root causes.

- Considers sustainability, scalability, and long-term effects.
- Includes realistic mechanisms for adoption and behavior change.

3 - Proficient

- Explains potential benefits to users clearly.
- Shows the solution addresses a real need.
- Sets reasonable expectations for impact.
- Includes a basic plan for user adoption.

2 - Developing

- Describes potential benefits vaguely.
- Shows limited understanding of how impact would occur.
- Weak connection between problem and benefits.
- Minimal consideration of sustainability or adoption.

1 - Beginning

- Benefits are unclear or unrealistic.
- Fails to explain how positive change would occur.
- Poor connection between problem and proposed solution.
- No mention of adoption or implementation.