Unpacking Team Diversity: Quantifying Relationships between Skills Diversity, Demographic Diversity, and Academic Outcomes

Zhonghan Xie, University of Michigan

Zhonghan (Jonas) Xie is a Master of Public Policy and Master of Science in Information student at University of Michigan, with a keen interest in education policy evaluation and big data in higher education. He is passionate for finding policy solutions to improve educational and career outcomes of college students from diverse backgrounds. Before studying in United States, Zhonghan accomplished his bachelor's degree in public administration from Sun Yat-sen University in China. During his undergraduate years, Jonas interned at a local educational NGO in Guangzhou for four years and deeply engaged in development and evaluation of several educational programs for rural-to-urban migrant students in urban China.

Rebecca L Matz, University of Michigan

Becky Matz is a Research Scientist on the Research and Analytics team at the Center for Academic Innovation at the University of Michigan. She directs and supports research projects across Academic Innovation's portfolio of educational technologies. Her research expertise is in assessing the efficacy of software tools that support student learning and success, analyzing quantitative equity disparities in STEM courses across institutions, and developing interdisciplinary activities for introductory chemistry and biology courses. Matz holds a B.S. in Chemistry from the University of Illinois and an M.S. in Educational Studies and a Ph.D. in Chemistry from the University of Michigan.

Dr. Robin Fowler, University of Michigan

Robin Fowler is a Technical Communication lecturer and a Engineering Education researcher at the University of Michigan. Her teaching is primarily in team-based engineering courses, and her research focuses on equity in communication and collaboration as well as in group design decision making (judgment) under uncertainty. She is especially interested in how power relationships and rhetorical strategies affect group judgment in engineering design; one goal of this work is to to understand factors that inhibit full participation of students who identify with historically marginalized groups and investigate evidence-based strategies for mitigating these inequities. In addition, she is interested in technology and how specific affordances can change the ways we collaborate, learn, read, and write. Teaching engineering communication allows her to apply this work as she coaches students through collaboration, design thinking, and design communication. She is part of a team of faculty innovators who originated Tandem (tandem.ai.umich.edu), a tool designed to help facilitate equitable and inclusive teamwork environments.

Unpacking Diversity: Quantifying Relationships between Demographic Diversity, Skills Diversity, and Learning Outcomes

Zhonghan Xie, Rebecca L. Matz, Robin Fowler
University of Michigan

Research Brief for an Empirical Research Paper

Abstract

Team diversity has been central to debates on institutional inclusion practices and discussions of team-based pedagogy. Using administrative records and survey data from a large undergraduate engineering course, we constructed multiple indices capturing team-level demographic and skill diversity. We found that demographic dissimilarity is not significantly correlated with final course grades but helps mitigate the negative impact on underrepresented minority students. We also found that skill diversity in terms of bringing different skill expertise into teams is associated with better academic performance. We generally found no evidence that team diversity harms learning outcomes. These findings underscore the heterogeneity of diversity effects in teams and suggest the need for nuanced design and pedagogical strategies to foster well-functioning team environments.

I. Introduction

Team-based and collaborative instructional approaches have been widely adopted across higher education based on evidence that structured team collaboration can improve student engagement, deepen learning, and build interpersonal skills such as communication and leadership. As these approaches proliferate, team formation has become a critical pedagogical question: how instructors compose teams shapes not only group dynamics but also individual student experiences and outcomes. One central dimension of team formation is diversity, whether and how teams should be composed across lines of gender, race, background, or experience. Prior research has shown that campus-level diversity and students' diversity experiences are often positively associated with learning and cognitive development (e.g., Bowman, 2010). However, the relationship between team-level diversity and student outcomes is more complex and less well understood.

The conceptualization of team composition diversity typically considers at least two dimensions: demographic diversity, which includes gender, race/ethnicity, and nationality diversity, and skill diversity, which considers students' prior knowledge, skills, and abilities relevant to completing course tasks (e.g., Bell et al., 2011; Horwitz & Horwitz, 2007). These dimensions may shape student outcomes through different mechanisms. Demographic diversity may primarily influence belonging, inclusion, and access to peer networks, while skill diversity may primarily affect how teams allocate work, share expertise, and solve complex problems. At the same time, team-level diversity may entail trade-offs: potentially improving learning and work products through exposure to new perspectives or complementary expertise, while also increasing coordination costs.

Existing research shows mixed findings about the relationships between team diversity and performance. Vasilescu et al. (2015) found that team diversity in gender and prior experience is associated with higher productivity. Other research shows demographic diversity may not have a direct effect (Horwitz & Horwitz, 2007) or has a neutral or marginally negative effect on team performance when holding skill diversity constant (Hamilton et al., 2012). Studies in higher education settings also show mixed results. Research at a leading U.K. university found that greater gender diversity on teams fosters a more inclusive and psychologically safe environment, leading to improved academic outcomes for all students (Contreras et al., 2022). Darova and Duchene (2023) found a U-shaped relationship with teamwork quality, that is, very homogeneous and very heterogeneous groups collaborated better.

Still, critical gaps remain regarding how different types of diversity affect team dynamics and how diversity affects individual learning**.** Prior literature shows that collaborative learning experiences can be inequitable and are affected by gender- and race-based stereotypes (e.g., Sekaquaptewa & Thompson, 2003; Henderson, 2023). The contribution and learning of individuals, especially students of color and female students, are fully realized only within inclusive team climates that actively recognize and value their perspectives. Research on the relationships between team diversity and individual learning is therefore essential to better determine whether diversity initiatives enhance outcomes for those most vulnerable to inequitable team processes and to identify whether targeted interventions are needed to foster inclusion in team-based learning environments.

Therefore, this research brief addresses the following research questions:
RQ1: What are the relationships between team skill diversity and academic outcomes?
RQ2: What are the relationships between team demographic diversity and academic outcomes?
RQ3: How do these relationships differ for historically marginalized students (e.g., students of color and female students)?

II. Data and Methodology

a) Data Source

With exemption from review from our Institutional Review Board, we focused on 1,221 students who enrolled in a required course for undergraduate engineering students (deidentified as Engineering or ENGR 150 herein) at the University of Michigan (U-M), a large research-intensive university, during the fall terms from 2021 to 2024. Students choose from a variety of ENGR 150 sections by topic ranging from robotics mechanisms to vehicle design. In each section, students work in a consistent team, usually of four to six students, for the duration of the term on a variety of assignments and projects. The total enrollment in each section is typically between 40 and 60 students each semester.

A subset of ENGR 150 instructional teams used an online collaboration support platform called Tandem to support teamwork. Tandem provides periodic surveys to track and assess team health as well as tailored online lessons throughout the semester (Fowler et al., 2021). At the beginning of the semester, Tandem also collects students' teamwork preferences and self reflections on their prior exposure to course-related skills, which helps instructors facilitate team formation. The overall response rate was approximately 96.5%. We obtained demographic characteristics (sex, underrepresented minority status, and student domestic vs. international status as defined by U-M) and ENGR 150 course grades and prior academic performance information from a U-M student records database. The prior academic performance metric was grade point average in other courses (GPAO; Young et al., 2023), which is the same as the student's cumulative GPA, except that it removes the effect of the course enrollment of interest (ENGR 150, in this case).

The descriptive statistics of our sample are shown in Table 1. Among these students, 33% were female and 20% were underrepresented racial/ethnic minorities (URM). Most were domestic students; 8% were international students. The average course grade (GPA) is 3.8

which corresponds to between an A- and A letter grade at U-M.

b) Operationalization

Our key explanatory variable is team diversity. There is extensive literature about quantifying team demographic diversity at the team and institutional levels (Chang & Yamamura, 2005; Darova & Duchene, 2024). Following similar settings as in Darova & Duchene (2024), we defined the team demographic dissimilarity index within a group as the average dissimilarity across all possible member pairs along multiple identity dimensions (sex, ethnicity, and student origin). Formally, for a group with *n* members and *K* identity dimensions, the team demographic diversity index (*DD*) is defined as

$$DD = \frac{1}{\binom{n}{2}} \sum_{i>j}^{n} \frac{1}{K} \sum_{k=1}^{K} 1(x_{ik} \neq x_{jk})$$

where $1(x_{ik} \neq x_{jk})$ is an indicator that equals 1 if members *i* and *j* in the same team differ on identity dimension *k*, and 0 otherwise. Thus, *DD* reflects the proportion of mismatched identity attributes averaged across all pairs and all identity dimensions within the group. Conceptually, this measure captures the extent to which group members differ demographically, normalized by the number of possible member pairs, thereby providing a standardized indicator of within-team demographic diversity.

Prior research in team diversity (e.g. Harrison & Klein, 2007) showed that within-team diversity is multifaceted and that different “diversity types”, i.e. variety, separation and disparity, may reflect fundamentally different compositions of difference with distinct theoretical mechanisms. Given the data availability, we operationalized skill diversity using the three complementary measures, skill coverage, skill composition, and skill disparity, to capture the distinct dimensions of how skills are distributed across team members.

We obtained students’ skill information from the beginning-of-semester survey, where students self-reported their experience in course-related skills using a four-item Likert scale, ranging from no prior experience in a specific skill to having applied the skill many times. We impute missing entries (3.5% of the students) using within-term skill means.

We computed skill coverage to capture the variety of skill resources available to each student team. We defined the skill coverage as the share of skill dimensions for which the team has at least one member meeting a proficiency threshold τ. In this paper, we set the threshold to $\tau = 4$, which stands for at least one member in the team indicating they have used the skill many times prior to the class.As different students on each team may have their own strengths in terms of course-related skills and may view the survey scale in different ways, we used the average pairwise cosine distance to capture the composition and relative emphasis of skills.

Finally, we included two dispersion-based measures to capture skill disparity, the difference in skill proficiency between team members within the same team. The within-team standard deviation of GPAO measures the dispersion in overall academic performance between members. The second, average skill range, captures within-team skill proficiency gaps at the skill level: for each skill, we computed the difference between the most and least proficient team members, then averaged the gap across all skills. Formally, average skill range is defined as follows, where p denotes the number of team members and X_{ij} denotes student i's self rating on skill j.

c) Model Specification

We used a linear model to estimate the relationship between team diversity and individual academic performance. Formally, the model is specified as follows.

$$Y_{it} = \beta_0 + \beta_1 DD_{it} + \gamma' X_{it} + \lambda_t + \epsilon_{it}$$

where Y_{it} represents the ENGR 150 course grade outcomes for individual i in term t; X_{it} represents the set of individual control variables, including sex, race/ethnicity, student origin, and GPAO. We also included the section-term fixed effects λ_t in the model to capture instructional differences between course sections along dimensions like assessment, grading, and teaching style Our key variable of interest (β_1) estimates the relationship between team diversity and grade outcomes.

III. Results

a) Demographic diversity and academic performance outcome

Table 2 shows the regression estimation results of team demographic dissimilarity. Generally, we did not find a statistically significant relationship between team demographic diversity and students' academic performance. Across all specifications, the estimated coefficient on demographic dissimilarity is small and statistically insignificant after controlling for GPAO, student demographics, and section and term fixed effects.

Several student-level characteristics are strongly associated with academic outcomes. Prior academic performance (GPAO) is the strongest predictor of course grades across all models ($\beta = 0.424$, $p<0.01$). Female students earn higher course grades across all models ($\beta = 0.097$, $p<0.01$ in Columns 1 and 3; $\beta = 0.138$, $p<0.1$ in Column 2). URM status is negatively associated with course grades in all specifications ($\beta = -0.056$, $p<0.05$; $\beta = -0.057$, $p<0.05$; $\beta = -0.248$, $p<0.01$). International student status is negatively associated with course grades, though none of these coefficients are statistically significant ($\beta \approx -0.04$).

We also included interaction terms between demographic diversity and sex and URM status to explore heterogeneous effects by student subgroups. The interaction between URM status and team demographic dissimilarity is positive and statistically significant ($\beta = 0.448$, $p<0.05$; Column 3). That is, higher team demographic diversity is associated with higher course grades for URM students relative to non-URM students. This indicates that demographic diversity is beneficial for URM students and mitigates the negative association of URM status and academic performance. The interaction between female status and demographic dissimilarity is negative but statistically insignificant ($\beta = -0.089$, SE = 0.153; Column 2).

b) Skill diversity and academic performance

Table 3 compares multiple measures of team skill diversity and their statistical relationships with individual academic performance. Measures capturing dispersion or distance in skills show no statistically significant relationship with course grades. The coefficient of within-team standard deviation of GPAO is negative but small and statistically insignificant ($\beta = -0.046$, SE = 0.042; Column 4). Similarly, average pairwise cosine distance, which captures dissimilarity in students' skill profiles, is negatively associated with grades, though the estimate is not statistically significant ($\beta = -0.160$, SE = 0.328; Column 5). However, skill coverage, which reflects the breadth of distinct skills represented within a team, is positively and statistically significantly associated with better academic outcomes. Students in teams with higher skill coverage earn higher course grades ($\beta = 0.113$, $p<0.05$; Column 6), holding individual demographic characteristics and team demographic dissimilarity constant. Skill range, which captures the difference between the highest- and lowest-skilled team members, is positively but not significantly associated with course grades ($\beta = 0.027$, SE = 0.021; Column 7), suggesting that the magnitude of skill gaps within teams is not independently predictive of performance.

Taken together, our results show that team skill diversity in the sense of skill coverage relates to better academic performance. It suggests that having access to a broader set of complementary skills within a team may be more beneficial than differences in skill levels or profiles alone. We did not find evidence suggesting that greater variability in team members' skills directly translates into better individual academic performance. Consistent with prior models, GPAO remains the strongest predictor of course performance ($\beta \approx 0.42$, $p<0.01$). Coefficients of individual demographic controls are similar across the different model specifications.

IV. Discussion

Using longitudinal administrative records from a required undergraduate engineering course, this study documents nuanced and heterogeneous relationships between team composition and student academic outcomes. Contrary to existing research in the context of higher education, we did not find that overall team demographic dissimilarity is directly associated with course grades. This finding parallels evidence from a comparable undergraduate business course setting, where aggregated team demographic dissimilarity showed only marginal and mixed associations with academic performance, and effects emerged primarily in non-linear specifications (Xie et al., 2026). However, we found that higher demographic dissimilarity may provide compensatory benefits for underrepresented minority students in terms of higher course grades. It suggests that the effects of demographic diversity are not uniform across students but depend on how team composition intersects with individual social identities.

Our results of multiple team skill diversity measures further highlight the importance of construct choice in diversity research. Measures capturing skill dispersion or distance, including within-team GPAO standard deviation, average pairwise cosine distance, and skill range, are not significantly associated with academic performance. In contrast, skill coverage is positively and significantly associated with course grades. This pattern suggests that academic benefits arise not from dispersion in skill levels per se, but from access to complementary areas of expertise. Skill diversity brings complementary skills that expand a team's collective capacity, which further supports teams in problem-solving and leads to better academic performance.

These findings should be interpreted considering two key limitations. First, the analysis draws on a required course at one institution, which may limit the extent to which the patterns generalize across disciplinary or institutional contexts. Second, team skill measures rely on students' self-reported assessments, which may be subject to systematic bias in self-evaluation. Rather than offering definitive conclusions, this study aims to encourage further research across varied settings and to prompt instructors to consider both demographic and skill diversity when formulating student teams and designing team-based courses.

This study contributes to the literature by showing that the influence of team diversity depends critically on both the dimensions of diversity and its operationalization. Our findings also have important implications for pedagogical design in engineering education. Instructors should move beyond generic calls for "more diversity" in teams and instead attend to which dimensions of diversity are being introduced and for what purpose. Team formation strategies that increase diversity may help support underrepresented minority students, while ensuring high levels of skill coverage can improve overall academic performance. By aligning team design with pedagogical goals and providing targeted support, instructors can more effectively leverage both demographic and skill diversity to promote equitable learning outcomes.

V. Acknowledgements

The authors are grateful for funding provided by the University of Michigan's Enhancing Engineering Education (E3) program, and by the National Science Foundation (DUE 2120252). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting Specific about Demographic Diversity Variable and Team Performance Relationships: A Meta-Analysis. *Journal of Management*, *37*(3), 709–743. https://doi.org/10.1177/0149206310365001

Chang, M. J., & Yamamura, E. (2005). Quantitative approaches to measuring student body diversity: Some examples and thoughts. In *Higher education in a global society: Achieving diversity, equity and excellence* (pp. 369–386). Emerald Group Publishing Limited. https://www.emerald.com/insight/content/doi/10.1016/s1479-358x(05)05018-7/full/html

Contreras, V., Orsini, C., & Ozcan, B. (2022). *Effects of team diversity on performance, perceptions, and predictions: Experimental evidence of gender composition and language* (Social Policy Working Paper). LSE Department of Social Policy.

Darova, O., & Duchene, A. (2024). Diversity in Teams: Collaboration and Performance in Experiments with Different Tasks. *MPRA Paper, MPRA Paper*, Article 120519. https://ideas.repec.org//p/pra/mprapa/120519.html

Fowler, R., Alford, L. K., Sheffield, S., Hayward, C., Henderson, T. S., & Matz, R. L. (2021, July). Supporting Equitable Team Experiences Using Tandem, an Online Assessment and Learning Tool. In 2021 ASEE Virtual Annual Conference Content Access. https://www.doi.org/10.18260/1-2--37787

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2012). Diversity and Productivity in Production Teams. In A. Bryson (Ed.), *Advances in the Economic Analysis of Participatory and Labor-Managed Firms* (Vol. 13, p. 0). Emerald Group Publishing Limited. https://doi.org/10.1108/S0885-3339(2012)0000013009

Henderson, T. S. (2023). Understanding Access to Learning Opportunities in Collaborative Projects: Gendered Social Hierarchies in Student Teams. *Studies in Engineering Education*, *4*(1). https://doi.org/10.21061/see.101

Horwitz, S. K., & Horwitz, I. B. (2007). The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. *Journal of Management*, *33*(6), 987–1015. https://doi.org/10.1177/0149206307308587

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*(1), 68–74. https://doi.org/10.1016/S0022-1031(02)00508-5

Vasilescu, B., Posnett, D., Ray, B., van den Brand, M. G. J., Serebrenik, A., Devanbu, P., & Filkov, V. (2015). Gender and Tenure Diversity in GitHub Teams. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, 3789–3798. https://doi.org/10.1145/2702123.2702549

Xie, Z., Matz, R. L., & Fowler, R. (2026, April). Unpacking Diversity: Heterogeneous and U-Shaped Relationships between Team Diversity and Learning Outcomes. Presented at the 2026 AERA Annual Meeting.

Young, N. T., Matz, R. L., Bell, E. F., & Hayward, C. (2023). How Researchers Calculate Students' Grade Point Average in Other Courses Has Minimal Impact. PLoS ONE, 18(8), e0290109. https://doi.org/10.1371/journal.pone.0290109

Appendix

Table 1: Descriptive Statistics

Variable	**N**	**Mean**	**Standard Deviation**
Course GPA	1221	3.83	0.40
GPAO	1221	3.51	0.50
Team Diversity Dissimilarity	1221	0.42	0.14
Skill Cosine Distance	1221	0.06	0.03
Team Skill Coverage	1221	0.66	0.28
Average Skill Range	1221	1.72	0.51
Within-team GPAO Standard Deviation	1221	0.41	0.24
Female	1221	0.33	0.47
URM	1221	0.20	0.40
International Student	1221	0.08	0.27

Notes: The table presents the descriptive statistics of students in ENGR 150 from Fall 2021 to Fall 2024. The number in each cell represents the mean or proportion of the variable.

Table 2: Relationships between Team Demographic Diversity and Academic Performance

	(1)	(2)	(3)
Diversity Dissimilarity	0.058	0.081	-0.017
	(0.071)	(0.081)	(0.077)
Female	0.097**	0.138*	0.097**
	(0.020)	(0.073)	(0.020)
URM	-0.056**	-0.057**	-0.248**
	(0.024)	(0.024)	(0.080)
International	-0.048	-0.047	-0.038
	(0.035)	(0.035)	(0.035)
Female X Dissimilarity		-0.089	
		(0.153)	
URM X Dissimilarity			0.448**
			(0.180)
GPAO	0.424**	0.424**	0.424**
	(0.019)	(0.019)	(0.019)
Intercept	2.250**	2.241**	2.285**
	(0.088)	(0.090)	(0.089)
Section Fixed Effects	Yes	Yes	Yes
Term Fixed Effects	Yes	Yes	Yes
N	1221	1221	1221
R-squared	0.39	0.39	0.40

Note: The table presents the relationships between team demographic diversity and individual academic performance outcomes (course grades) of students in ENGR 150 from Fall 2021 to Fall 2024 terms. Team demographic diversity is measured by the multi-level demographic dissimilarity index. All models include section and term fixed effects. Standard errors are presented in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 3: Relationships between Team Skill Diversity and Academic Performance

	(4)	(5)	(6)	(7)
Group GPAO SD	-0.046			
	(0.042)			
Cosine Distance		-0.160		
		(0.328)		
Skill Coverage			0.113**	
			(0.036)	
Skill Range				0.027
				(0.021)
Female	0.101**	0.102**	0.102**	0.101**
	(0.020)	(0.020)	(0.020)	(0.020)
URM	-0.055**	-0.055**	-0.053**	-0.055**
	(0.024)	(0.024)	(0.024)	(0.024)
International	-0.040	-0.040	-0.038	-0.042
	(0.034)	(0.034)	(0.034)	(0.034)
Diversity Dissimilarity	0.061	0.065	0.058	0.047
	(0.071)	(0.072)	(0.071)	(0.072)
GPAO	0.418**	0.425**	0.423**	0.425**
	(0.020)	(0.019)	(0.019)	(0.019)
Section Fixed Effects	Yes	Yes	Yes	Yes
Term Fixed Effects	Yes	Yes	Yes	Yes
Intercept	2.319**	2.282**	2.225**	2.221**
	(0.094)	(0.086)	(0.085)	(0.093)
N	1221	1221	1221	1221
R-squared	0.39	0.39	0.40	0.39

Note: The table presents the relationships between team skill diversity and individual academic performance outcomes (course grades) of students in ENGR 150 from Fall 2021 to Fall 2024 terms. Team skill diversity is measured by the group-level standard deviation of GPAO, average pairwise cosine distance, skill coverage, and average range of skill. All models include section and term fixed effects. Standard errors are presented in parentheses. * p<0.1, ** p<0.05, *** p<0.01