

# Try Before You Buy: Are There Benefits to a Random Trial Period before Students Choose Their Collaborative Teams?

Sukhada Samudra,<sup>a</sup> Cynney Walters,<sup>b</sup> Destiny Williams-Dobosz,<sup>c</sup> Aarati Shah,<sup>d</sup> and Peggy Brickman<sup>e\*</sup>

<sup>a</sup>Department of Genetics, and <sup>e</sup>Department of Plant Biology, University of Georgia, Athens, GA 30602; <sup>b</sup>Oak Ridge Institute for Science and Education Fellowship, Division of Foodborne, Waterborne, and Environmental Diseases (DFWED), Centers for Disease Control and Prevention (CDC), Atlanta, GA 30333; <sup>c</sup>Department of Educational Psychology, University of Illinois Urbana-Champaign, Champaign, IL 32611; <sup>d</sup>College of Dentistry, University of Florida, Gainesville, FL 32611

## ABSTRACT

The cognitive and performance benefits of group work in undergraduate courses depend on understanding how to structure groups to promote communication and comfort while also promoting diversity and reducing conflict. The current study utilized social network analysis combined with self-reported survey data from 555 students in 155 groups to understand how students identified group members whom they wished to work with. Students' willingness to work with their peers was positively associated with behavioral traits pertaining to attention, participation, and preparedness in class. We tested whether preventing students from choosing their group members until completing a multiweek period of random assignment to different groups each week influenced group selection criteria, and we found little effect. Students continued to depend on demographic similarities such as gender and ethnicity when selecting groupmates and enforcing random interactions before the group formation did not influence group satisfaction and/or grades. Random interactions before group formation did influence the willingness of students to continue working with peers who were persistently poorly rated based on behavioral attributes and contribution to the group work. Thus, the effort of random assignment could be beneficial to identify struggling students and improve collaboration.

## INTRODUCTION

National science education policies promote collaboration as a key scientific practice. Students who engage in collaboration demonstrate greater learning, achievement, attitudes toward learning, and persistence in STEM courses and programs (Lou *et al.*, 1996; Springer *et al.*, 1999; Metoyer *et al.*, 2014). In addition to providing important opportunities for scientific investigation, laboratory courses also provide excellent opportunities to foster development of the social skills necessary for effective collaboration (Seifert *et al.*, 2009; Corwin *et al.*, 2015). Instructors need to be proactive in forming groups in labs to ensure effective collaboration (Johnson and Johnson, 1999; Kreijns *et al.*, 2003). Currently, there is a lack of consistent advice for these instructors who wish to structure teams to ensure diversity in gender, ethnicity, or prior academic performance (Donovan *et al.*, 2018; Wilson *et al.*, 2018). This study characterizes factors students use when forming groups in laboratory classes, including the role of random assignment, and then analyzes subsequent outcomes in terms of conflict and performance to provide evidence-based suggestions for instructors.

Kristy Jean Wilson, *Monitoring Editor*

Submitted Feb 10, 2023; Revised Sep 29, 2023; Accepted Oct 11, 2023

CBE Life Sci Educ March 1, 2024 23:ar2

DOI:10.1187/cbe.23-01-0011

\*Address correspondence to: Peggy Brickman (brickman@uga.edu).

© 2024 S. Samudra *et al.* CBE—Life Sciences Education © 2024 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 4.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

### The Case for Student Choice in Forming Collaborative Groups

Students who are allowed to select their own group members report positive outcomes that include: reported satisfaction (Bacon *et al.*, 1999; Connerley and Mael, 2001; Chapman *et al.*, 2006; Myers, 2012); higher initial group cohesiveness, communication, enthusiasm, and confidence in each other (Strong and Anderson, 1990; Chapman *et al.*, 2006; Ciani *et al.*, 2008); higher grades (Mahenthiran and Rouse, 2000); and greater ownership of group tasks (Mello, 1993). Students who are comfortable with their groups also demonstrate greater content mastery, with the greatest predictor of comfort being friendship status (Theobald *et al.*, 2017).

However, researchers also document negative outcomes when students are allowed to select their own groups (Feichtner and Davis, 1984) including: negative opinions of the course, instructors, projects, and classmates (Brickell *et al.*, 1994) and poor test results in physical sciences laboratory classes (Lawrenz and Munch, 1984). The negative perceptions of self-selected groups might derive from the fact that on their own, students often select group members that they already know or who are from similar cultural, ethnic, racial, and academic backgrounds (Jalajas and Sutton, 1984; Chapman *et al.*, 2006; Rienties *et al.*, 2014; Freeman *et al.*, 2017). This creates homogeneous groups, a situation referred to as homophily in the field of social network analysis (SNA; McPherson *et al.*, 2001).

Both negative and positive features are attributed to homophily, making it challenging for instructors to know whether they should intervene to ensure greater heterogeneity in groups. On the negative side, homogeneous groups that lack gender and national diversity exhibit lower cognitive complexity in their collaborative group work, lower performance, and lower quality ideas during collaboration (Watson *et al.*, 1993; McLeod *et al.*, 1996; Curşeu and Pluut, 2013). In addition, failure to integrate students into diverse groups may perpetuate a lack of the equity and inclusivity of classrooms (Malcom-Piqueux and Bensimon, 2017; Ruedi *et al.*, 2020). On the flip side, there are positive reports for maintaining homophily. Students from historically underserved groups may achieve a greater sense of comfort when they can choose their own group members. For example, LGBTQIA students report problems with assigned groups, preferring to be able to work with someone who would be accepting of their identities (Cooper and Brownell, 2016). Similarly, international students seek comfort in students with similar cultural backgrounds (Hendrickson *et al.*, 2011).

Instructors who choose to assign student groups to enhance diversity must consider a host of characteristics including ethnicity, nationality, gender, age, sexual orientation, prior academic performance, and teamwork experience for which there may be a paucity of experimental evidence of the benefits or balancing these factors in group settings. Most studies categorize gender as a dichotomous variable that misses nuanced aspects of the personal and cultural construction of gender (Knaak, 2004). Even with that limitation, in some research settings gender composition is found to have no impact on final outcomes of group work (Takeda and Homberg, 2014). While in other cases, for example, problem solving in Physics classrooms, homogeneous-gender groups and mixed gender groups perform better than men-dominated groups with a single woman (Heller and Hollabaugh,

1992). These researchers note that in a gender unbalanced group with only one woman, men are more likely to dominate discussions (Heller and Hollabaugh, 1992). This tendency for men to receive more attention from their group members and exert stronger influence compared with women is seen in multiple sociological settings (Carli, 2001). Perhaps this tendency for men to dominate in group work could be alleviated through greater gender balance.

Socially-constructed groups with enhanced diversity do demonstrate increased interactions and resulting transfer of knowledge. In a quasiexperimental study, postgraduate students in groups randomized to increase diversity with respect to gender and nationality show similar team cohesion and performance compared with groups where students self-selected their members (Rienties *et al.*, 2014). However, more “knowledge spillovers” (unintentional, informal, and uncompensated transfer of knowledge among the interacting individuals) occur in the randomized groups. The researchers conclude that randomizing groups can have positive effects on informal and formal interactions between the students from diverse backgrounds beyond the group (Rienties *et al.*, 2014).

### Theoretical Framework

We employ a social cohesion perspective to guide our understanding of how demographic factors and comfort with team members influences group interactions that result in greater achievement (Speer *et al.*, 2001). Social cohesion emphasizes that the effects of cooperative learning on students and their achievement derive primarily from the quality of group interactions (Battistich *et al.*, 1993). Slavin’s (2014) excellent review of theories related to cooperative learning, explains that the level of social cohesiveness of a group determines the quality of the interactions between its members. Better social cohesion results in greater elaborated explanations, peer modeling, practice, and peer correction which subsequently results in enhanced learning. There are several indicators of poor social cohesion including low levels of communication and self-reported conflict. In addition, other researchers characterize social cohesion through the constructs of friendship and comfort. We know from prior research that when selecting their group members, students often choose peers from similar gender, ethnic, racial, and social backgrounds with whom they are more comfortable (Chapman *et al.*, 2006; Freeman *et al.*, 2017). In addition, while studying group dynamics and student learning in a large biology undergraduate classroom, Theobald and colleagues (2017) confirm that friendship is the major predictor for students’ perception of comfort while working in a group as well as a major contributor to learning. Social cohesion theory thus helps focus our selection of quantitative measures of group social cohesion used in this study (communication, conflict, and satisfaction) as well as the SNA we employ to examine how student collaborative groups form and are sustained.

### Selection of Measures of Social Cohesion

Several measures can help monitor and identify social cohesion and performance in groups (Jehn and Mannix, 2001; Lejk and Wyvill, 2001; Curşeu *et al.*, 2012; Brickman *et al.*, 2021). Jehn and Mannix’s (2001) conflict identification items measure: 1) task conflict which occurs when students debate conflicting ideas or points of view; 2) process conflict which occurs

over disagreements about group members' responsibilities to schedule and complete tasks; and 3) relationship conflict which is emotional in nature and involves disagreements about personal issues. Task and process conflict are seen as productive group interactions because they foster intellectual debate. Teams at high performance levels demonstrate low but increasing levels of process conflict over time, moderate task conflict, and low levels of relationship conflict. Curşue and colleagues (2012) find that an item that measures communication frequency between team members has a positive effect on discussion of ideas (task conflict) in their structural equation model. They also find that higher levels of relationship conflict have a negative effect on the cognitive complexity of a summative group assignment. Finally, one study reveals that a single categorical item can be more effective at identifying conflict than Jehn and Mannix's (2001) scale, suggesting that more qualitative measures may be needed to help students qualify group cohesive interactions (Brickman *et al.*, 2021).

### Social network analysis (SNA)

Social networks can characterize and define relationships between individuals within a group (Wasserman and Faust, 1994). SNA is a useful tool that can be used to visualize student interactions to better characterize group selection criteria and social cohesion. There are four basic assumptions of SNA: 1) individuals and their actions are interdependent; 2) information is transmitted via relational ties between individuals; 3) relationship patterns (social structures) can influence individual actions by providing opportunities for and/or constraints on individual behavior; and 4) social network models conceptualize relationship patterns among actors in a network (Carolan, 2014; Wasserman and Faust, 1994). SNA is grounded in systemic empirical data and relies on the use of graph theory, mathematical, statistical, and computational models to represent complex social interactions (Freeman, 2004). SNA can be used to draw statistical inferences from the relational data to understand flow of information /attitudes within actors, and to explore learning outcomes of a classroom (Carolan, 2014). For example, network analysis tied to the statistical approach of exponential graph random models (ERGMs) reveals sociodemographic clustering in friendship networks (Goodreau *et al.*, 2009). Clustering in SNA is defined by connections formed between two actors (nodes) in a network because of shared attributes or sociodemographic, behavioral, and/or intrapersonal characteristics (McPherson *et al.*, 2001). Clustering observed as a structural property of a network can provide insight about relationship patterns including homophily. Actors who get locked into a fixed position in a network develop less social ties which inherently restrict information flow (Carolan, 2014).

Prior SNA studies reveal that a student's position within communication and interaction networks correlate with performance (Brunn and Brewe, 2013; Grunspan *et al.*, 2014) and that students' performance can be influenced by GPA, gender, attendance, and number of ties within a network (Buchenroth-Martin *et al.*, 2017). Additional data-driven SNA approaches are useful in evidence-based revision of active learning curricula, comparing strategies for promoting peer interaction and group learning, and understanding group formation in a classroom (Buchenroth-Martin *et al.*, 2017). We

utilize SNA in this study as a visual (qualitative) tool to understand group formation in an undergraduate laboratory classroom setting.

### Study Goal

In this study, we investigate the impact of student choice of groups on facets of social cohesion. We compare student groups who completely self-assemble without instructor intervention (referred to as unstructured) to student groups who are only allowed to select their members after an initial period of random assignment (referred to as structured.) We hypothesize that students who self-select their group members after random assignment exhibit lower levels of clustering and homophily seen through SNA. We also examine other facets of social cohesion including quantitative measures of conflict, satisfaction, and communication frequency, a qualitative analysis of group selection criteria, and detailed study of case studies of problematic group interactions, and how they compare between structured and unstructured groups.

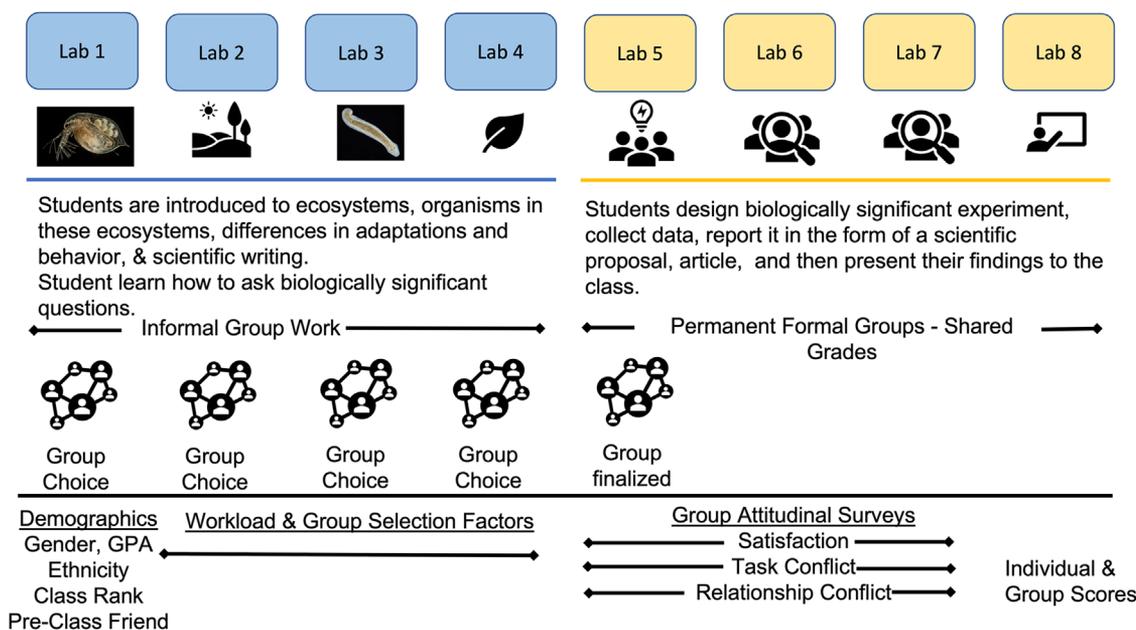
Specifically, we investigate these research questions using a mixed-method approach:

1. Are there quantitative differences in positive or negative outcomes for groups that have structured opportunities to select their members?
2. What group selection criteria do students use when forming groups and do they differ between structured and unstructured groups?
3. Do groups that have structured opportunities to select their members make different decisions regarding working with poorly rated group members compared with unstructured groups?

## METHODS

### Instructional Context

Our study population was enrolled in an inquiry-based introductory biology lab (Principles of Biology II laboratory) at University of Georgia in the spring semester of 2019. All participants provided informed consent to our study which was deemed exempt according to the University of Georgia Institutional Review Board (STUDY00005732). Laboratory curriculum and assignments are outlined in Figure 1. During the first 4 wk of lab, informal group work was used to help students explore aquatic and terrestrial biodiversity field activities and introduce scientific processes and the writing-intensive curriculum. Formal student groups were formed during wk 5. For the following 3 wk (5–7), groups agreed on a biologically significant question, designed a scientific experiment to answer that question, performed an introductory literature review, drafted a research proposal, collected and presented data, and communicated their findings in a research article. The formal group grade for these weeks comprised submission of a research proposal and bibliography. Students presented their data as a group and were graded as a group on this presentation. Assignments and grades were distributed so that 53% of points were derived from individual assignments and 46% from group assignments. After wk 8, students were also expected to perform data analysis and submit an individual research article. For the rest of the semester, students performed field work and associated experiments as a class activity, during which they



**FIGURE 1:** Data collection scheme. Lab numbers indicate the weeks when data was collected. Surveys for wk 1–4 included questions that asked students to select the members that they worked with that week indicate whether they were preclass friends, and rate their participation. Surveys in wk 5–7 asked students to indicate satisfaction and conflict with their permanent group members. Individual grades and total group grades were then collected after wk 8 at the end of the semester.

had at least two more weeks of informal group work where they could work with anyone they preferred.

Each laboratory section was taught by a single instructor and consisted of six tables each with a maximum capacity of four students who worked together for each group activity. Each section had a maximum of six groups with three or four students per group for a possible enrollment cap of 24 students. (See *Schematic Representation* in Supplemental Figure S1).

### Assigning structured and unstructured lab sections, survey distribution and data collection

Each instructor/graduate laboratory assistant (GLA) in our study was assigned to teach two lab sections. For every individual instructor, we assigned one of their lab sections as an unstructured lab and other as structured lab section to control for any specific instructor variables that might have influenced student interaction while selecting their group members. In the structured labs, students were randomly assigned to a group number using Microsoft excel and seated at the same numbered table. Name tags were generated that include table numbers so that students were ensured to sit where assigned. In the unstructured labs, we also provided name tags for students for the first 4 wk. These name tags did not provide a table or group number and were used only to ensure that students get to know the names of their group members. Students in unstructured sections were made aware that they could self-select their group members and change groups each week, if they desired. In both sections, students were allowed to choose their own group in wk 5 for formal group work that continued for several weeks.

Student enrollment data was used to create unique surveys (Qualtrics, 2020) for each lab section. Surveys were emailed at the start of the lab so that students could complete the survey

at the end of lab activities for that week. Reminder emails were sent every 2 d until the beginning of the next lab class to those students had not responded. Data was collected from a total of 30 lab sections (15 structured and 15 unstructured.) As most of the group grade was associated with formal group work, we anticipated observing maximum group related conflicts in the wk 5–7. Hence, we collected data from wk 1–7 along with the final grades. (Schematic representation of the study provided in Figure 1.)

### Student demographics

Student demographics were collected during the first week of lab to identify the occurrence of homophily in group formation (Table 1; Supplemental Table S1). Students who preferred not to provide their biological sex were given “other” designation. We incorrectly provided students with categories referring to biological sex when we asked for them to indicate their gender in our survey (Supplemental Material, Appendix 1). We have corrected this in Table 1, and we refer readers to this excellent resource for researchers on the subject published after we collected our data (Sullivan, 2020). We defined racial and ethnic URM status using NSF’s definition of historically underrepresented minorities in Science and Engineering (NCSES, 2021): this included students who identified themselves as Black, Hispanic, Native American, Alaskan natives, and/or Pacific Islanders, and biracial. As expected for a predominantly White institution, we observed a low percentage of URM students (<20%). Subsequently, we combined the five classifications into a single URM status for statistical analysis (Table 1). We also collected demographic data on students who reported their ethnicity as Asian (Indian, Korean, Chinese, Filipino, and Japanese; Supplemental Table S1). Students who did not wish to disclose their ethnicity and/or who chose every single option from the

TABLE 1: Demographics of the students – (A) Gender and Ethnicity and (B) class rank

A							
Labs	Number	Biological sex			URM status		
		Female	Male	Other	URM	Non URM	Other
Structured	323	207	113	3	49	256	18
Unstructured	330	224	105	1	64	254	12
B							
Labs	Class Rank						
	Freshman	Sophomore	Junior *	Senior	Prefer not to answer		
Structured	33	201	67	16	6		
Unstructured	27	182	94	27	0		

drop-down menu in the survey were compiled together and given “Other- ethnicity” status.

No statistically significant differences were observed between student demographics (biological sex and URM status) between structured and unstructured labs (gender:  $p = 0.4048$ ; URM:  $p = 0.226$ ; Pearson’s Chi-squared test). The average self-reported GPA of structured labs was 3.48 out of 4 and of unstructured labs was 3.50 out of 4. When we compared class rank, we found that structured labs had more students in their junior year than unstructured labs. ( $p = 0.03469$ ; Pearson’s Chi-squared test).

#### Instruments/Survey questions and survey design

To understand student’s perception of group members in both the early phase of group formation as well as subsequent formal group activities, survey questions were designed to gather both qualitative and quantitative characteristics that students used to select and rate group members (Table 2). Social networks were constructed using students’ reports of the names of the students that they worked with that week and asking students to identify any students that they worked with who were preclass friends. We defined preclass friend as, “someone the student considered as a friend before the term” (Theobald *et al.*, 2017).

For every group member they worked with for that week, they were asked to select positive or negative traits from a pre-populated list of group selection factor categories that we developed (Table 2). These qualitative group selection factor items were generated by our preliminary exploratory study during previous semester. For this preliminary study, we asked a total of 74 students (three unstructured and two structured labs) over the period of the first 4 wk to provide pairs of traits they used to define a peer as a good or poor group member. Two members of the research team (S.S. and D.W.-D.) began the analysis by open coding (also called “initial coding”) 530 examples that students provided for good group members traits and 527 examples that students provided for poor group-member traits (Strauss and Corbin, 1998; Saldaña, 2021). Although we had not identified a priori codes or themes, we knew we were looking to explain the phenomena of social comfort, trust, and conflict. After discussing and characterizing the responses, the two coders worked to create a codebook using the words of the participants, a process called in vivo coding (Charmaz, 2006; Saldaña, 2021). Two additional researchers (P.B. and J.L.) consulted on the code book and suggested revisions. As new codes emerged, the first two indepen-

dent coders (S.S. and D.W.-D.) analyzed previous answers to again to look for the new codes (Schwandt, 2014). We continued this iterative coding process until no new codes were revealed in the data (Charmaz, 2006; Saldaña, 2021). All examples that students provided were coded to consensus between the two raters. Three main themes emerged during coding: cognitive traits, attributes, and behavioral traits (Supplemental Figure S2). The most common trait that students provided to define poor and good group members was behavioral (50%). The three most common behavioral traits mentioned were used to create a group selection factor item (Table 3; Supplemental Figure S2)

Quantitative measures of group dynamics were utilized to measure aspects of conflict, satisfaction, shared workload, and frequency of communication. We assessed communication frequency within the group every week (from survey 1–7) using scale provided by Curşeu *et al.* (2012). To identify free riders in early phase of group formation (wk 1–4), we asked students to score individual group members using a numerical Likert scale from 5 (very well) to 1 (very poor) to indicate how well they shared the groups’ workload each week. (Figure 1, Supplemental Materials, Appendix 1). Additional quantitative measures of group dynamics were administered during wk 5–7 and included prior validated items on group satisfaction and conflict (Jehn and Mannix, 2001; Van der Vegt *et al.*, 2001). Both shared workload and identification of poor and/good group member traits were used to confirm whether group selection factors were significantly different from our preliminary data analysis. See Supplemental Materials, Appendix 1 for all survey items in details and Figure 1 for data collection scheme.

#### SNA to visualize group formation

We used SNA to visualize the difference in how student groups formed in structured and unstructured labs during the early phase of group formation (wk 1–4) and to compare how students selected their final group members in wk 5. R packages: *statnet*, *ergm*, *ergm.count*, *ergm.rank*, and *latentnet* (R Core Team, 2022) with the R3.6.1 (Hunter *et al.*, 2008; Krivitsky and Handcock, 2008) were used for plotting networks. Every laboratory section was confined by the physical space and time of the day so that students from one lab section could not work with the students from other lab sections: thus, each laboratory section was considered to be an individual network with its boundary. Each lab had a maximum of 24 students. Each student was asked to self-report which three other students

**TABLE 2: Research questions and measures**

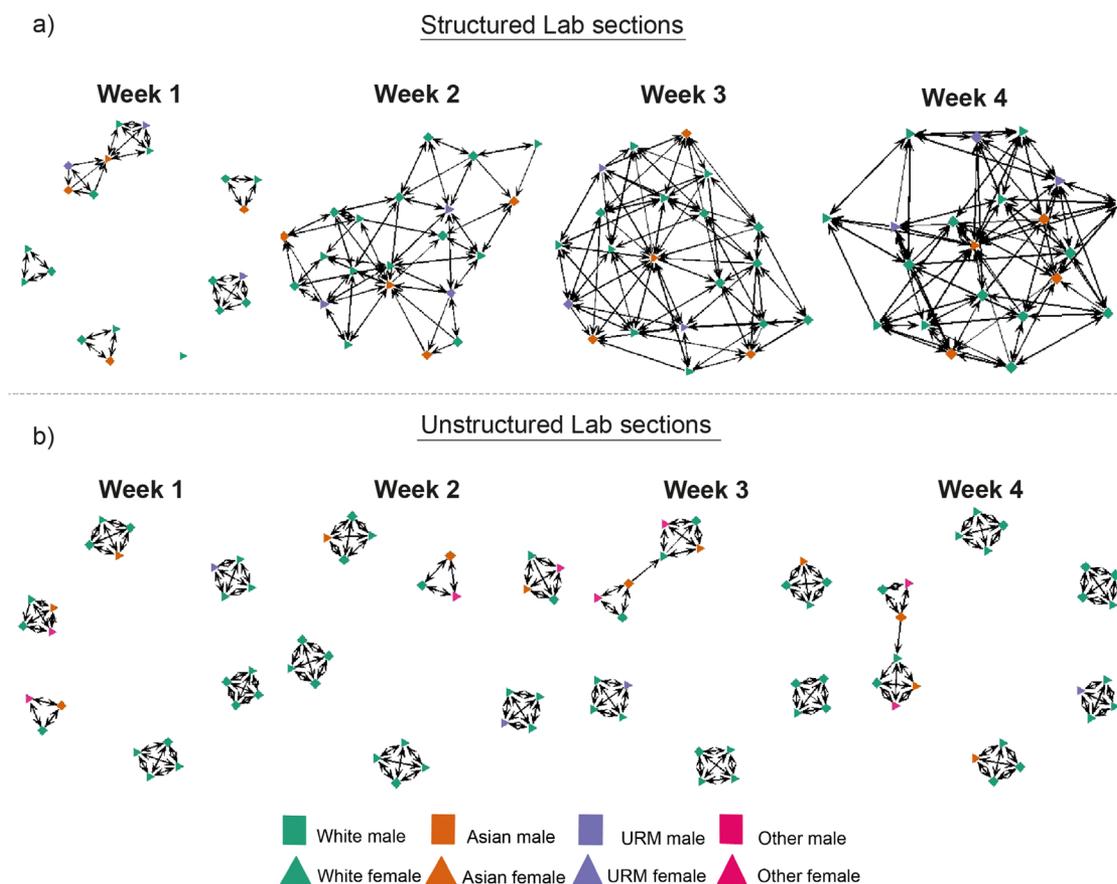
Research questions	Measures	Example items
1. Are there quantitative differences in positive or negative outcomes for groups that have structured opportunities to select their members?	Conflict (Jehn and Mannix, 2001) Satisfaction (Van der Vegt <i>et al.</i> , 2001) Communication frequency (Curşeu <i>et al.</i> , 2012)	1. I am pleased with the way my teammates and I work together. 2. How much conflict of ideas is there in your work group? 3. How much relationship tension is there in your work group? 4. Indicate the frequency of communication between members of your group in the lab today.
2. What group selection criteria do students use when forming groups and do they differ between structured and unstructured groups?	Group selection factor item SNA (biological sex, preclass friend, race/ethnicity)	1. For each of the students that you worked with today, select only those that you would like to work with again. 2. What factored into your decision for the last question? a) Well prepared for class and knows the material well in advance. b) Pays attention. c) Participates in discussion and offers meaningful suggestions. d) Does not come to class prepared or does not know the material in advance. e) Does not seem interested in the class. f) Does not participate in discussions/listen to others.
3. Do groups that have structured opportunities to select their members make different decisions regarding working with poorly rated group members compared with unstructured groups?	Shared workload (Likert scale 1 very poor –5 very good) Group selection factor item (negative comments)	1. For each student that you worked with in lab today, rank how well you felt they shared the workload as a group member: shared workload includes discussing ideas, using equipment, recording data, presenting your group's ideas, asking relevant questions, etc. a) very poor b) poor c) moderate d) good e) very good 2. For each of the students that you worked with today, select only those that you would like to work with again. 3. What factored into your decision for the last question? Negative options: a) Does not come to class prepared or does not know the material in advance. b) Does not seem interested in the class. c) Does not participate in discussions/listen to others.

they worked with in each lab session. Ideally each student had a chance to work with a maximum of 23 students over the period of 4 wk. Using cumulative data from wk 1 through 4, directed binary social networks were plotted where 0 represented no interaction between the two students and 1 represented interaction between two students. These binary networks indicated whether two students worked together or not. The directionality indicated if one (single arrow) or both (double-headed arrow) of the student actors reported the interaction. Because randomization was used to ensure that students

had a maximum number of chances to work with other students in the structured labs, we confirmed that we had one large network in these sections at the end of 4 wk (representative section plotted in Figure 2). Although, because of random assignment, it was possible for students in the structured labs to work with the same students on multiple weeks, but it was rare. In comparison, smaller network patterns were observed in unstructured sections (representative section plotted in Figure 2) indicating that student actors only worked with a subset of students for all 4 wk.

**TABLE 3: Top three behavioral traits students reported in construction of group selection item**

Behavioral trait category	Description for good group member	Description for poor group member
Attention/Interest	Pays attention	Does not seem interested in the class
Participation	Participates in the discussion and offer meaningful suggestions	Does not participate in the discussions/listen to others
Preparedness	Well prepared for class and knows material well in advance	Does not come to class prepared or does not know the material in advance



**FIGURE 2:** Representative examples of directed binary social networks for wk 1–4 in (a) structured and (b) unstructured lab sections. Each line represents a tie between two students (nodes of the network). The directionality of the tie is represented by the direction of the arrow. If the arrows are not bidirectional, it means only one student reported the interaction. The color of nodes indicates ethnicity. Green = White, Orange = Asians, Violet = URM, and Pink = Other. Triangles represent female students and squares represent male students.

In 4 sections we saw no difference in the networks that formed in structured and unstructured labs. Discussions with the GLAs for those lab sections revealed that they had inappropriately randomized their unstructured lab sections and allowed us to additionally randomize the structured sections. So, for these two GLAs both their lab sections were randomized. Therefore, we removed a total four lab sections (two unstructured and two structured sections associated with these GLAs) from further data analysis.

### SNA to identify homophily

In our study, we wished to determine whether the level of the homophily observed between two nodes of the network (students) could be attributed to shared characteristics like ethnicity, biological sex, and preclass friend status. We identified homophily by qualitative review of social networks that formed. As our classrooms were predominantly white students who reported their biological sex as female, to understand the assortative mixing tendencies, we used differential homophily while plotting and analyzing networks. Differential homophily accounts for the fact that the assortative mixing tendencies for formation of groups are different, depending on the friendship status of the individual students involved (Goodreau *et al.*, 2009) (<https://github.com/eehh-stanford/SNA-workshop>).

We analyzed directed valued networks after formal groups were formed (wk 5). In a valued network, 0 ties indicated no interaction; 1 tie indicated interaction between two students; and 2 ties represented interaction between two students who were preclass friends. We analyzed group formation qualitatively by visually analyzing social networks to find groups with at least two URM students working together in that group who were not preclass friends and at least two students with same biological sex who were preclass friends. For plotting the valued networks, we used the tutorial and code published by The Statenet Development Team (Krivitsky *et al.*, 2021).

### Case Studies: Identification of persistently poorly rated students (PPR students)

We used two criteria for identifying students who were persistently poorly rated by their peers: 1) shared workload, and 2) perceived poor-group behavior. Shared workload scores were calculated for each individual student by averaging all scores they received from all their group members for the first 4 wk and used to identify reported free-riding or social loafing behavior. In addition, data from the group selection factor item was compiled for every student for both negative and positive responses. To normalize data, we calculated the total number of comments an individual received that week and then

calculated the percentage of those comments which were negative and positive. With an average group size of four students, each student could receive a maximum of nine positive or nine negative comments. We determined the number of students receiving negative comments every week in both structured and unstructured labs as well as the type of negative comment received from the prepopulated list. We defined a PPR student as any student who received greater than 50% negative comments for two or more weeks.

### Statistical analysis

Quantitative survey items that were used to compare satisfaction, conflict, shared workload, and communication frequency were converted into numerical value on Likert scale. For the satisfaction score the numerical values assigned ranged from strongly agree = 5, agree = 4, neutral = 3, disagree = 2, strongly disagree = 1. For task and relationship conflict scores, the numerical values assigned were none = 1, Rarely/little = 2, Some = 3, Often/much = 4, very much/very often = 5. For the shared workload score the numerical values assigned were very poor = 1, poor = 2, neutral = 3, good = 4, very good = 5. For the communication frequencies, a -5 to +5 scale was converted to 0 to 11. A score of six indicates an average level of communication (Curşeu *et al.*, 2012).

We used the Shapiro-Wilk test for checking the normality assumption for the satisfaction, task conflict, relationship conflict, process conflict, communication frequencies, shared workload score, and final grades. As the data is not normally distributed, nonparametric tests were used for data analysis. To compare the difference between structured and unstructured labs scores, the Wilcoxon rank sum test with confidence interval of 0.95 was used where required. We used multiple regressions to examine whether randomizing students for the first 4 wk before they select their formal group has any effect on group satisfaction, conflict, and communication frequencies. We were interested in investigating the impact of this treatment, biological sex, race/ethnicity, GPA, year, and week of semester on the outcomes mentioned above. Using regression, we could investigate the relative impact of these factors simultaneously (Field *et al.*, 2012). For regression analysis we used multilevel modeling. Our data was collected over multiple weeks from multiple sections. As described in Theobald (2018), the student interactions in our data collection were not independent of each other. This nonindependence of sampling was accounted for by random effects in multilevel regression models using the guideline and R code described in Theobald (2018).

### Regression Model Selection

We first generated a highly inclusive initial model with treatment (structured/unstructured labs), biological sex, ethnicity, GPA, year, and week as fixed effects. We then calculated intra-class correlation score (ICC) for two random effects, student random effects (repetitive measures) and section random effects (clustering). ICC score for repetitive measures was 0.42 and for clustering was 0.05. We then included both the effects in the initial model. To find the best supported model we used Akaike information criterion (AIC) validation with a relative fit index (Premo *et al.*, 2018). The model with lower AIC value was considered a more optimal model (Burnham and Anderson, 2004). As described in Theobald (2018), once the initial

model was established, backward stepwise regression was used to remove fixed factors from the model. AIC values between the models were then compared (Theobald, 2018). We used the model with the lowest AIC with fewer predictors as recommended in Theobald (2018). The details are in Supplemental Table S4. This process was repeated to examine factors predicting every outcome. The best models were used to investigate any difference between structured and unstructured labs. The details are presented in Supplemental Table S5.

To compare the frequency of reported poor group members between structured and unstructured labs, we performed a Chi-squared test to identify any statistically significant difference. For this purpose, we used the total number of students and the total number of students who received at least one negative comment for a particular week. Descriptive statistical analysis of satisfaction and relationship conflict scores were calculated using Microsoft Excel for students who we identified as PPR and their respective group members. We used one assignment – the individual research article – to determine performance levels of PPR compared with their respective group members. For group performance, we combined both the group research proposal and group presentation scores. These major writing assessments for the experiment the groups conducted during wk 5–7 were the best qualitative measure to contrast individual with group performance.

All the analyses were performed in R v4.1.2 using packages: lme4, psych, clinfun, pastecs, pgrimess, remotes, and car (Revelle, 2022; R Core Team, 2022)

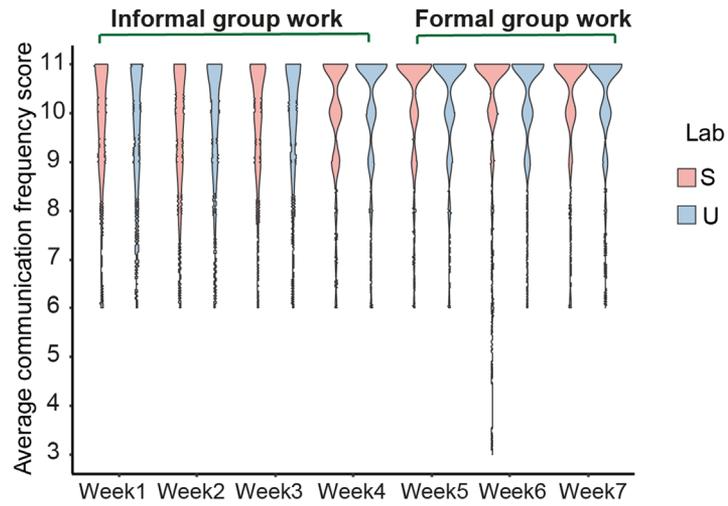
## RESULTS

### Question 1: Are there quantitative differences in positive or negative outcomes for groups that have structured opportunities to select their members?

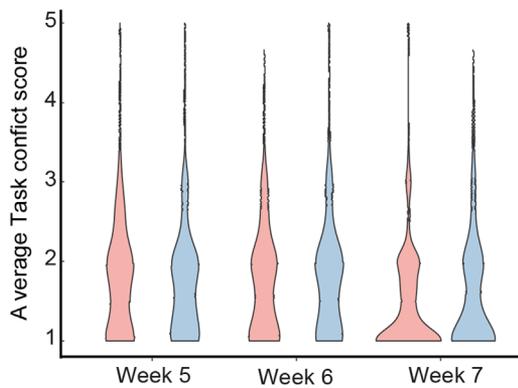
We did not find any statistically significant difference in positive or negative quantitative outcomes between the groups in structured and unstructured lab sections. Using multilevel regression models, we investigated whether randomizing students for the first 4 wk before they select their formal group had any effect on group satisfaction, conflict, and communication frequencies. According to our theoretical model, we anticipated increased communication frequencies and hence lower relationship conflicts in structured lab sections as compared with unstructured labs during wk 5 to 7. However, we did not find any difference between the sections (Figure 3a; Supplemental Table S5). Relationship conflict items indicated more emotional exchange and tension within the group with a score of one indicating an absence of any relationship conflict. The average relationship conflict score in both structured and unstructured lab settings was lower than 1.5, indicating rare incidents of reported relationship conflict or there were no relationship conflicts within the groups at that time point (Figure 3c; Supplemental Tables S5 and S6).

Overall students reported that they were satisfied with their team members. The average process conflict score was lower than 1.5 in both the lab settings indicating a relative lack of conflict over responsibilities pertaining to group tasks during wk 5–7 with no significant differences between the structured and unstructured labs. Our initial predication was that with higher communication frequencies and lower relationship conflict scores, we would see higher task conflict and satisfaction

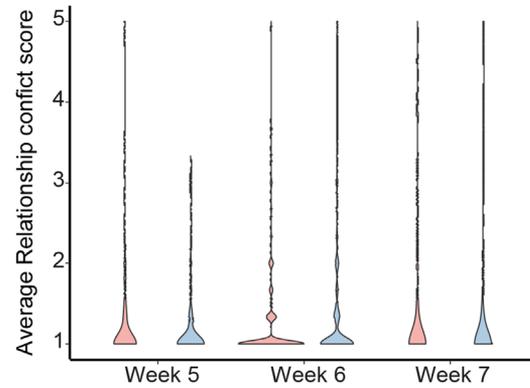
a) Communication frequency



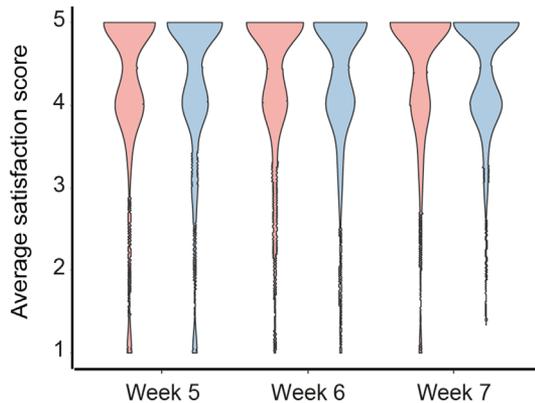
b) Task Conflict



c) Relationship Conflict



d) Satisfaction



e) Final grades

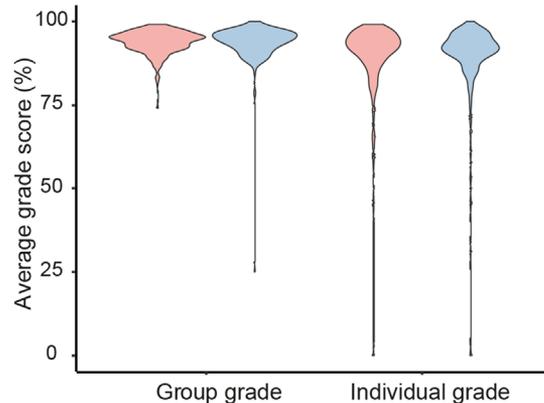


FIGURE 3: Differences in average communication frequency, task conflict, relationship conflict, satisfaction, and final grades between structured and unstructured lab sections. Data was collected from 13 unstructured and 13 structured lab sections. Pink represents data from structured labs, and blue represents data from unstructured labs for (A) communication frequency reported from wk 1–7, (B) Task conflict, (C) Relationship conflict, (D) Satisfaction scores reported from wk 5–7 when formal groups were formed, and (E) Final average group and individual grades.

**TABLE 4: Qualitative analysis of differential homophily – a) Preclass friend status and biological sex and b) Ethnicity/Race.**

A				
Lab sections	Groups with students who were preclass friends*	Number of all female groups who were preclass friends*	Number of all male groups who were preclass friends*	Number of all groups with students of different biological sex who were preclass friends*
Structured	34	17	1	16
Unstructured	34	13	3	18
B				
Lab sections	Total number of groups	Number of groups with students who were preclass friends*	Number of groups with URM students with no preclass friend status*	Number of groups with Asian students with no preclass friend status*
Structured	77	34	5	3
Unstructured	78	34	8	5

\*At least two students in a group.

scores in structured labs. Task conflict items measured student perceptions of discussion and exchange of ideas, with a score of five indicating these discussions occur very often and score of one indicating a discussion did not occur at all. In both the structured and unstructured labs, the average task conflict score was lower than two (Figure 3b; Supplemental Tables S5 and S6), indicating that students either did not report the differences in opinions and discussions among the group members or there was a pattern of less discussion. A satisfaction score of five indicated that students strongly agree that they felt satisfied working with that group each week. A satisfaction score of one indicated that students strongly disagree that they felt satisfied working with that group in a given week. Both the structured and unstructured labs had average satisfaction scores greater than four with no significant difference observed (Figure 3d; Supplemental Tables S5 and S6).

The average group grade and individual grade were 94% (0.94) and 90% (0.90), respectively in both the structured and unstructured labs with no significant difference (Figure 3e; Supplemental Table S6). There were a total three students who failed the class and hence their grade is zero. (Two students were from structured labs, and one was from an unstructured lab.) We also did not find any difference between communication frequencies and shared workload score in structured and unstructured labs in wk 1–4 (Supplemental Tables S8–S10)

### Question 2: What group selection criteria do students use when forming groups and do they differ between structured and unstructured groups?

The qualitative SNA indicated that there was no significant difference between structured and unstructured lab sections when it came to student's choice of group members. This was in accordance with the theoretical framework where we hypothesized that choice of selection was driven by preclass friend status, ethnicity, and/or gender. Students in unstructured labs were given sole choice in which they worked with from the start of the course, while students in structured labs had random exposure to different students in wk 1–4 before deciding on their formal groups in wk 5. We followed interaction patterns using SNAs for all students throughout the first 4 wk, and in wk 5 to understand the degree to which being a preclass friend or similar biological sex drove the student's choice of group. Out

of 26 labs we analyzed, only one lab had an equal proportion of male and female students. The remaining lab sections had female to male ratios of at least 1.5, so we considered differential homophily to determine whether male students were clustering together because of their biological sex, preclass friend status, or both. We found that there were equal numbers of groups with preclass friend status in both structured and unstructured labs (Table 4). The number of groups with same sex students who were preclass friends and groups with different biological sex who were preclass friends were also comparable (Table 4). We identified four groups out of 77 in structured labs where male students who were not preclass friends worked together. However, we could not conclude whether biological sex might have a role in relationship to friendship. The number of groups in both structured and unstructured labs, where clustering was observed because of preclass friend status, URM status, or Asian status, was comparable. All the groups with students of mixed biological sex or mixed ethnicity had preclass friend status (Table 4). Details about individual lab sections were provided in Supplemental Tables 2 and 3. In cases of isolated URM students where there was absence of preclass friends or other URM students in a group, same biological sex groups were formed. A similar trend was found with Asian students. The number of groups with this trend was comparable between both structured and unstructured lab sections. We found a comparable number of groups (22 in unstructured and 15 in structured lab sections) without any distinguishing pattern of preclass friend status, ethnicity, or gender. These groups had a mix of White, Asian, and URM ethnicity and a higher proportion of female students. The higher proportion of female students was expected, as in each individual lab section, there were more female students than male students (Supplemental Table 3).

### Question 3: Do groups that have structured opportunities to select their members make different decisions regarding working with poorly rated group members compared with unstructured groups?

We observed different patterns in how groups formed and in the reporting of negative group behavior patterns in structured and unstructured labs. In unstructured labs, students formed a group in wk 1, and then they continued to work with the same

students throughout the semester even though they had an opportunity to change the groups with some exceptions. 15 students (out of 330) changed the group from wk 1 to 2. These students continued to work with the same group after wk 2; four students (out of 330) changed the group from wk 2 to 3. These students continued to work with the same group after wk 3; three students (out of 330) who worked with the same group for the first 4 wk changed their group in wk 5. We did not have any specific information about why they changed their group. These students were not identified as PPR students as per our analysis.

Students in unstructured labs were less likely to report negative group behaviors but more likely to stay with a PPR group member even when they did report negative group behaviors. We compared the frequency of students reporting of poor group member behavior during both the group formation phase (wk 1–4) and the common group selection factors reported. In structured labs, 282 students received comments in wk 1, 271 in wk 2, 282 in wk 3, and 276 in wk 4. In unstructured labs, 291 students received comments in wk 1, 290 in wk 2, 284 in wk 3, and 283 in wk 4. A greater number of students received negative comments in structured labs (10–20%) compared with unstructured labs (<10%) with statistically significant differences observed in wk 3 and 4. ( $\chi^2$  [1,  $N = 282$ ] with Yates correction = 5.45,  $p = 0.019$  for wk 3) and ( $\chi^2$  [1,  $N = 276$ ] with Yates correction = 13.72,  $p = 0.0002$  for the wk 4; Figure 4a).

We used the “group selection factors” commonly reported by students (Table 2) to calculate the total number of students who received at least one negative comment in the category of behavioral traits in the first 4 wk. This distribution of students was represented as a Venn diagram in Figure 4b and was similar between both structured and unstructured labs. In structured lab sections, lack of interest was identified as the primary negative trait (100 students), followed by participation (81 students), then preparedness (63 students). In unstructured lab sections, the primary negative trait identified was also lack of interest (54 students), but it was almost equal to those identified with poor participation (52 students), with preparedness as the third ranking trait identified (24 students). 37.34% of students in structured labs compared with 33.33% of students in unstructured labs received at least two negative behavioral trait comments from their peers, while 9.6% of students in structured compared with 6.45% in unstructured labs received all three negative traits. A weekly distribution of the students who received negative comments is represented in Venn diagrams (Figure 4, c and d).

We attempted to identify whether negative comments were a result of random behavior or whether some students were PPR students. Using the definition described in the *Methods*, we identified three PPR students out of a total of 323 in structured lab sections and five PPR students out of 320 students in unstructured labs sections (Figure 5). There was only one PPR student identified per group. Case studies were taken of these eight PPR students to characterize the type of PPR behavior, group members’ response to this behavior, and influence of PPR students on group satisfaction, conflict, and performance.

### Case studies

Figure 5 depicts the percentage of negative comments received by PPR students (pseudonyms) in a) structured and c) unstruc-

tured labs as well as the shared workload scores in b) structured and d) unstructured labs for wk 1–4. PPR received at least two comments each week, except for wk 1 when Thomas received only one comment and it was negative. However, the rest of his group members did not complete the survey for that week. We identified three types of behavior patterns and grouped the students accordingly.

PPR group I students received *exclusively* negative comments along with shared workload scores for consecutive weeks that were below the class average. In all 4 wk, Odin and Cesar exemplify this category with shared workload scores of two or below for the labs where they received 100% negative comments. (Figure 5c; Class shared workload score average in Supplemental Table S10.) For wk 1, Cesar received 50% negative comments and Odin received 28.57% negative comments, and their shared workload score was below 3.5.

PPR group II students *consistently* received negative comments for all the labs they attended along with lower than class average of shared workload scores for consecutive weeks. We identified four students with this behavior pattern: Athena, Olivia, and Caliban consistently received negative comments for the 4 wk and received 50% or more negative comments for 2 wk. All three of them also received lower than class average of shared workload scores for three consecutive weeks. Thomas did not attend a lab in wk 4 and hence data from that week is missing for him. However, he received negative comments consistently in all the remaining 3 wk and had lower than class average of shared workload score (Supplemental Table S10) for consecutive 2 wk.

PPR group III students did not receive any negative comments for at least a week. However, for the weeks that they received negative comments, their shared workload scores were below class average. Two students fit into this category. Zeus received 0% negative comments for two consecutive weeks and Ophelia had 0% negative comments in 1 wk. Both Zeus and Ophelia had shared workload scores that were greater than the class average for those corresponding weeks. For both Zeus and Ophelia, the shared workload score for the weeks when they received more than 50% negative comments three and 3.5, which was below the class average (Supplemental Table S10). Ophelia received 16% negative comments for wk 2 with the shared workload score below the class average of 4.6.

Both structured and unstructured lab sections had all three types of PPR students. When comparing structured and unstructured lab sections, however, every group member who gave negative comments to the PPR students reported the desire of not working with them again. However, in unstructured lab sections, these students continued to work with PPR students in their formal groups in wk 5–7, while group members chose not to work with PPR students in structured sections (Table 5). Figure 6 a) and b) represents network data from groups of students who worked with Cesar (structured) and Odin (unstructured) from wk 1–4 along with their comments. In the unstructured lab sections, all the group members who provided negative comments for Odin specifically mentioned in the surveys that they did not want to work with him again. Contrary to their expressed intentions, these same group members continued to work with Odin in wk 5 when formal groups were formed (Figure 6d). As a result of randomization that occurred in structured sections, Cesar worked with seven different

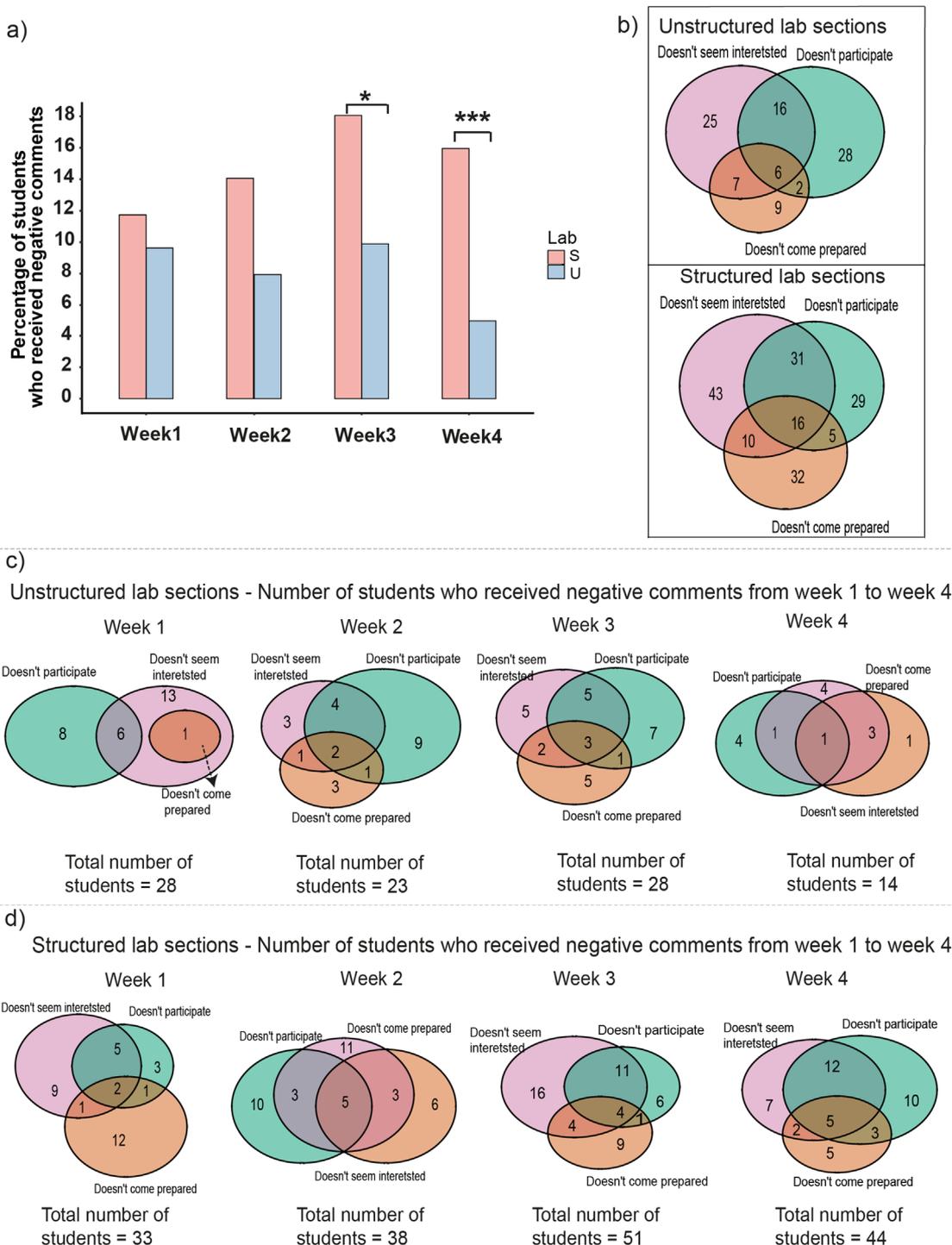
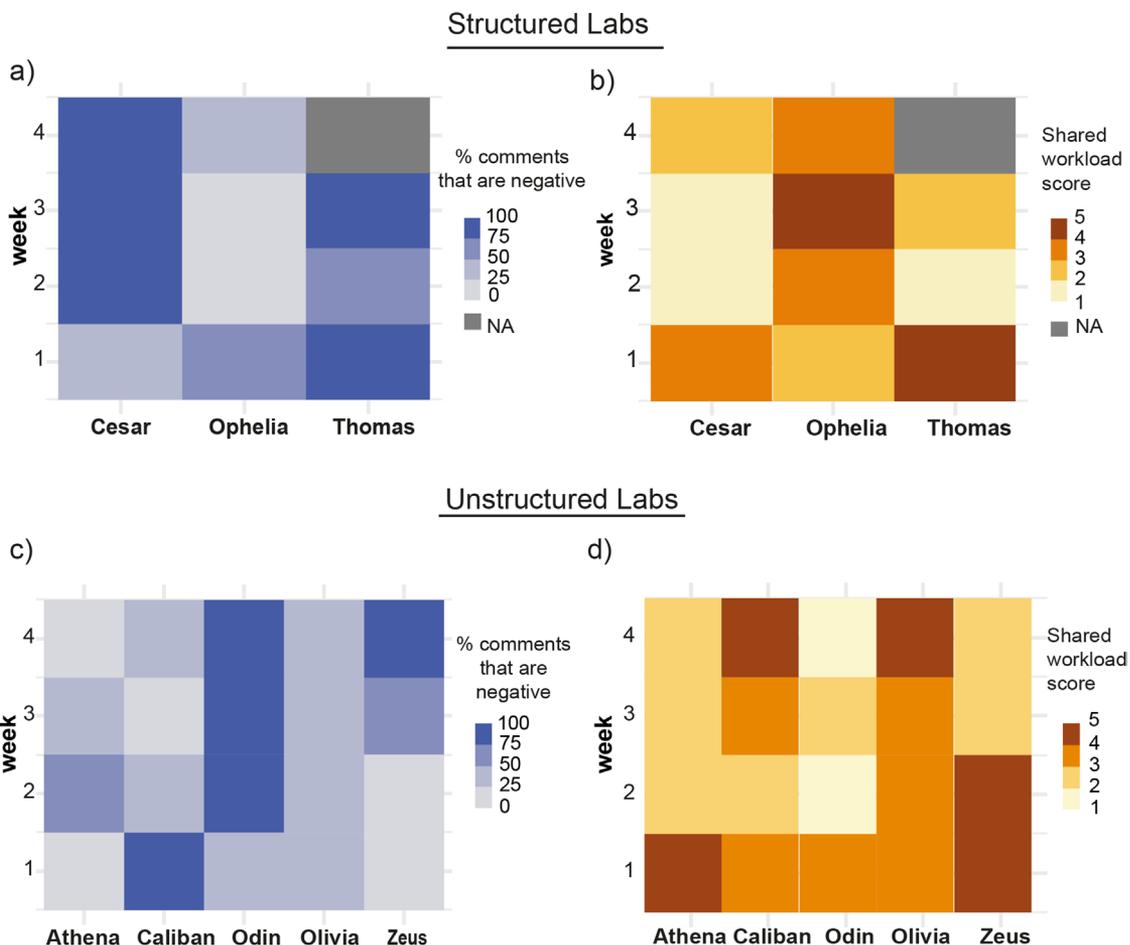


FIGURE 4: Reports of negative comments given by peers (group selection factors). (A) Percentage of students receiving negative comments from their group members across wk 1–4 in both the structured (pink) and unstructured (blue) lab sections. \*Pearson’s Chi-squared test with Yates’ continuity correction statistics ( $\chi^2 [1, N = 282]$  with Yates correction = 5.45,  $p = 0.019$ ) \*\*Pearson’s Chi-squared test with Yates’ continuity correction statistics ( $\chi^2 [1, N = 276]$  with Yates correction = 13.72,  $p = 0.0002$ ). (B) Venn diagrams comparing the number of students in the three negative behavioral group selection categories in structured and unstructured lab sections over wk 1–4. Each circle represents one category of group selection factor. The size and number in each circle indicate the number of students who received that category of comment from their group members. (C) Venn diagrams of weekly data of students with the negative comments in (C) unstructured and (D) structured lab sections. Green indicates “student doesn’t participate”, pink indicates a “student doesn’t seem interested,” and orange indicates a “student doesn’t come prepared for the class.”



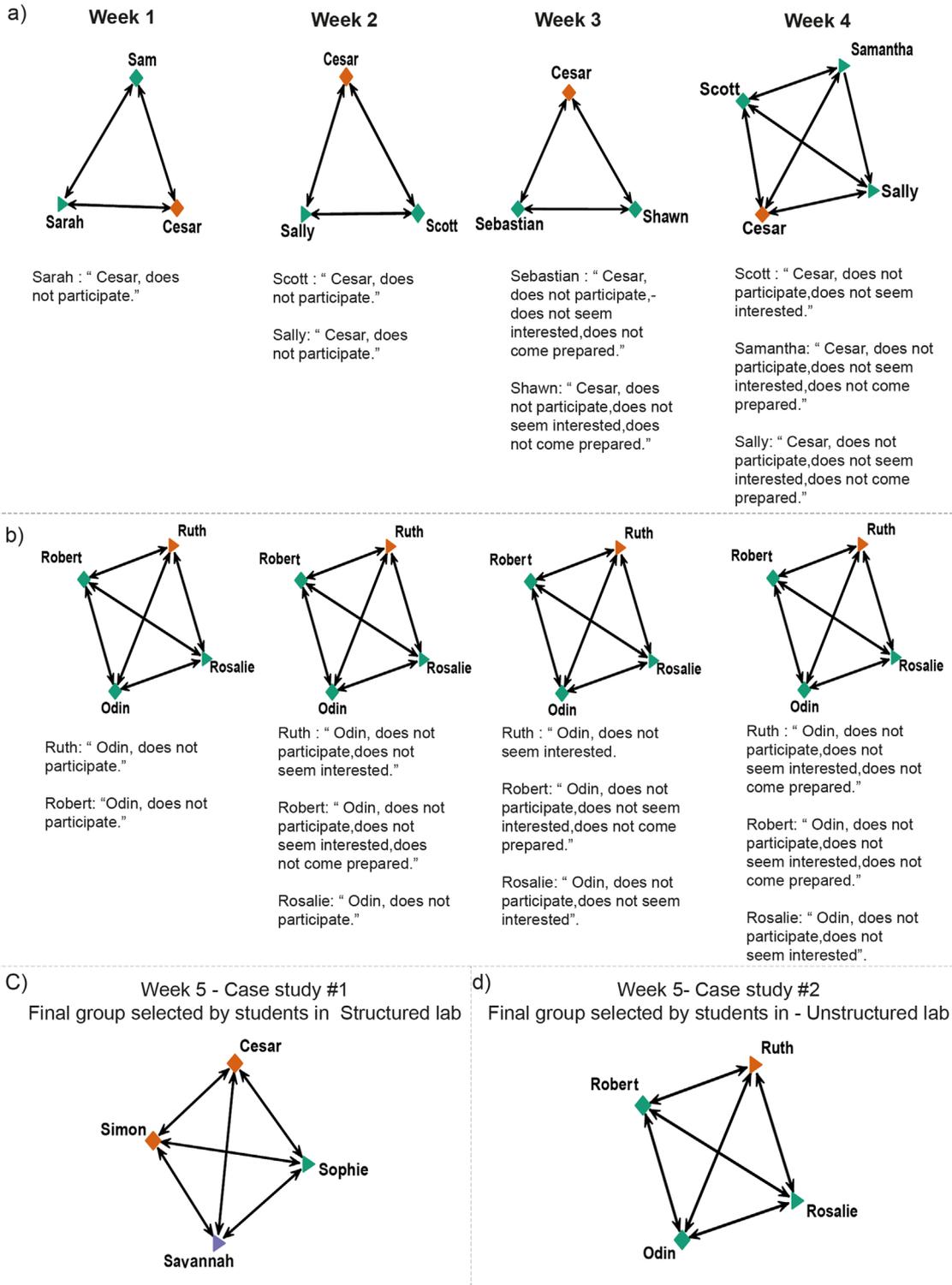
**FIGURE 5: Identification and Characterization of PPR students.** Heatmaps of the percentage of negative comments a student received from their group members that week in structured (A) or unstructured (C) lab sections. Color intensity is associated with the percentage and is indicated by a scale next to the heatmap. A gray square indicates that the student was either absent that week or did not receive any comments from his/her group member that week. Darker blue color indicates a higher percentage of negative comments. Heatmaps of the shared workload score of a student received from their group members in structured (B) and unstructured (D) lab sections. The maximum score is five and the minimum is one. The color intensity associates with the score received and is depicted by a scale next to the heatmap. A gray square indicates that the student was either absent that week or did not receive any score from his/her group member that week.

students for the first 4 wk before finalizing the group in lab 5. We note that two students, Scott and Sally, worked with Cesar more than once during these 4 wk as a result of randomization. All the students who gave negative comments to Cesar also reported in the surveys that they did not want to work with him again. In accordance with the expressed intentions, none of these students chose to work with Cesar when formal groups were formed in wk 5. (Both Cesar and Odin did not complete any surveys except the one in wk 1.)

Research articles and research proposal grades received also provided measures of group performance in the groups with PPR students (See *Methods*.) Data is summarized in Table 6. The average class grade for individual performance is 87.95%. Cesar, Ophelia, and Zeus have individual grades lower than the class average. The average class grade for group performance is 92.17%. Groups with Cesar, Thomas, and Ophelia have lower than the class average group grades. Apart from the group with

**TABLE 5: Group’s response to PPR students**

Pseudonym	PPR behavior group	Did the other group members continue working with the PPR students?
<b>Unstructured Lab sections</b>		
Odin	I	Yes
Athena	II	Yes
Caliban	II	Yes
Olivia	II	Yes
Zeus	III	Yes
<b>Structured Lab sections</b>		
Cesar	I	No
Thomas	II	No
Ophelia	III	No



**FIGURE 6: Social Networks with Qualitative Comments from PPR group (I) students.** Each network image represents the number of students in a group. Each line represents a tie between two students (nodes of the network). Directionality of the tie is represented by the direction of the arrow. If the arrows are not bidirectional, it means only one student reported the interaction. The color of nodes indicates ethnicity. Green = White, Orange = Asians, Violet = URM, and Pink = Other. Triangles represent female students and squares represent male students. (A) Case study #1 – Network data from a group of students who worked with Cesar from a structured lab section. (B) Case study #2 – Network data from a group of students who worked with Odin in an unstructured lab section. (C) Case study #1 – A network representing the final group selected by students in wk 5 in a structured lab. (D) Case study #2 – A network representing the final group selected by students in wk 5 in an unstructured lab.

TABLE 6: Group satisfaction, group performance, and individual performance of PPR students

Pseudonym	Group selection method	Average group satisfaction score (out of 5)	Group grade (%)	Individual grade (%)	Average individual grades of other group members (%)
PPR group (I)					
Cesar	Structured	3.85	90.00	76.73	82.60
Odin	Unstructured	3.81	97.64	91.86	92.23
PPR group (II)					
Thomas	Structured	4.86	90	88.43	86.08
Athena	Unstructured	3.97	87.64	89.03	93.08
Caliban	Unstructured	3.88	87.35	90.71	93.21
Olivia	Unstructured	3.51	94.11	98.23	75.29*
PPR group (III)					
Ophelia	Structured	3.51	97.64	55.29	91.06
Zeus	Unstructured	3.25	95.29	84.80	93.01

\*One student did not submit an assignment, hence, the group individual grade average is low.

Thomas, other groups with PPR students have satisfaction scores below the class average.

## DISCUSSION

### Students select group members who are friends or have similar demographics even after randomizing student interactions with multiple peers in the early phases of group formation

Instructors have received inconsistent recommendations on how to structure groups to balance students' comfort, group cohesiveness, and confidence in each other with diversity and work styles and preparation. We found no apparent benefits to structuring groups to ensure that students work with other students at random before selecting their final groups. Our results indicate that, while forming groups, students tended to first select their friends, then the people who share the same biological sex and/or similar racial/ethnic background; students who worked randomly with different students in wk 1–4 before deciding on their formal groups in wk 5 exhibited the same level of homophily as groups that were allowed to select their groupmates from the first week. One major limitation of our data was that it was collected in a predominantly White Institution with individual lab sections dominated by female students with low numbers of URM students.

We found that even though students had a chance to work with everyone in the class, they still preferred to form a formal group with their preclass friends. Nearly all the other groups in the lab sections with preclass friends were always mixed biological sex groups. We had higher number of women in any given laboratory section and small number of total students in each network, so we turned to qualitative visual analysis of SNAs to characterize interactions rather than conducting statistical tests.

Self-selected groups have been shown to be beneficial in prior research studies, and random assignment of groups does not appear to result in more long-lasting scholastic relationships among students. Chapman *et al.* (2006) compared random and self-selected groups and found that students reported better communication in self-selected groups, more enthusiasm, confidence in each other's abilities, and willingness to ask for help. Theobald *et al.* (2017) and Premo *et al.* (2022) also

observed higher levels of comfort and willingness to work with other students in groups with their friends. Lacey and colleagues (2020) conducted a similar intervention to our study. They constructed random pairings of students in lab courses and followed the students in all their courses over a year. They found that random pairings did not result in long-lasting working relationships. This was unfortunate because students who formed laboratory peer groups also formed education peer groups to help each other study, and these peer groups obtained similar overall grades. When Lacey and colleagues (2020) compared how students self-selected their groupmates, they found that students who attained high grades were more likely to report selecting their peers based on ability and work ethic. Students with lower attainment were more likely to indicate that they selected peers with similar backgrounds to work with. Lacey cautioned that instructors who attempt to engineer student groups by random assignment are unlikely to succeed in the formation of long-term peer support. Instead, they suggest embedding transient interactions (mini-breakout sessions) in which students change lab partners and specifically share skills and information that relate to challenging concepts which have been shown to increase course specific knowledge in other studies (De Hei *et al.*, 2018).

### Students choose to work with peers based on behavioral characteristics

Davies (2009) recommended the use of instructor-designed rubrics for identifying characteristics of an ideal team member that were enlisted from students themselves. Influenced by this recommendation, we performed preliminary data exploration to understand students' perception and how they defined good and bad group members (Supplemental Figure S2; Supplemental Table 3). The top three responses were behavioral traits associated with interest and engagement (participation and preparedness), and these terms were used to generate "group selection factors." Along with the behavioral characteristics, concerns over free riding group members and unequal distribution of work were reported as poor group member behaviors. This finding was consistent with Lacey and colleagues' (2020) results that the driving factor that students used when choosing to work with a peer was perceived work ethic. We found that

when asked a categorical yes/no question about willingness to work with a particular group member in the future, students' responses were more forthcoming compared with their responses to survey items using a Likert scale to rate their peers. Unwillingness to work with a group member in future was associated with participation and perception of unequal distribution of work members in the early phase of group formation. Premo and colleagues (2022) reported personal connections and contributions in the groups work were major predictors of student's willingness to work with a particular group member in future. Our data suggested a similar pattern of social predictors dictating student's choice of group members.

### **Students are more willing to report issues with group members in randomly assigned groups**

Unproductive teams often involve unequal distribution of labor and higher levels of interpersonal conflict (Livingstone and Lynch, 2000; Aggarwal and O'Brien, 2008; Pauli *et al.*, 2008; Shimazoe and Aldrich, 2010; Hall and Buzwell, 2012). However, students are reluctant to report group problems or to directly confront free-riders and /or social loafers (Strong and Anderson, 1990; Brickman *et al.*, 2021). We observed that students reported poor group member behavior using our single group selection factor item about a particular group member that directly contradicted the high scores they provided for group satisfaction, communication frequency and a low score for relationship conflict (Supplemental Table 4). In addition, although students reported unequal distribution of work as the major concern about group work, they simultaneously provided high ratings for their peers for shared workload in both structured and unstructured lab sections. These survey items were clearly less effective at identifying conflict or unequal distribution of workload. Our findings concur with what Brickman and colleagues (2021) found that students reported the presence of a problematic group member with a single "yes/no answer" survey question more effectively than task and relationship conflict items from Jehn and Mannix (2001).

Our data also indicated that fewer group work problems were reported in unstructured lab sections. In these sections, where students were allowed to choose their own groups from d 1, students continued to work with a poor group member they identified rather than confront the issue and change groups. In structured sections, when students identified a poor group member, they were more likely to avoid working with that group member in their final groups. Perhaps, the process of randomization helped students to avoid the social pressure that is associated with refusing to work with another student during the early phases of group formation.

SNA literature describes how relationships in social contexts generate a reputation cost for bad behavior and thus aid in collaboration (Burt *et al.*, 2013). In an undergraduate laboratory or classroom setting, it is possible to observe the relational embedding and structural embedding of the network. Relational embedding indicates two students with a relational history, such as friend or preclass friend status. Structural embedding is associated with students who have many mutual informal or formal connections, such as the connections enforced in our structured labs. The probability of bad behavior being discovered is higher in a more connected student net-

work (Burt *et al.*, 2013). This may explain the higher reported negative behavior in structured labs and the increased tendency to identify and avoid PPR students.

### **Categories of PPR students could be useful to instructors for interventions**

Prior work has revealed that students don't distinguish between strugglers, social loafers, and free-riding students in their definition of a good or poor group member (Freeman and Greenacre, 2011). Strugglers were defined as students who failed to contribute to the group or class because they were behind in their understanding of the class material (Freeman and Greenacre, 2011). Social loafers were defined as students who put in less effort because of a lack of identity or feeling of belongingness in a group (Davies, 2009). Free riders were defined as students who receive grades or rewards without putting in any effort, and often social loafing can lead to free riding (Watkins, 2004; Davies, 2009). None of these behaviors led to productive collaboration, but each of them might need different interventions to resolve.

Many interventions that have been suggested to resolve conflicts due to free riders might be inappropriate and harmful to struggling students. For example, constructive penalties for free-riding behavior might include allowing peers to anonymously provide peer ratings that could be used to adjust shared grades (Lejk and Wyvill, 2001; Johnston and Miles, 2004) or allow the instructor to intervene (Falchikov and Goldfinch, 2000). These actions might spur free riders into action and resolve group conflict. However, these same actions toward a struggling student might exacerbate their ability to perform successfully in the course. Researchers have demonstrated that groups that are performing poorly enact higher penalties from struggling students on peer ratings than groups that are performing well (Chang *et al.*, 2018).

Without instructor intervention, students might enact their own penalties to free riders that were even more destructive to collaboration. These could include: excluding that member from communication, completely removing peer support, or assigning tasks to students which were unsuitable and too difficult for them (Freeman and Greenacre, 2011). These socially destructive behaviors might increase relationship conflicts within already challenging group situations (Freeman and Greenacre, 2011), affect the learning environment for all students, and be particularly destructive to struggling students.

To support learning for all students, instructors need to identify the reason why certain students were not contributing and mediate. There could be a variety of reasons why a student was not attentive or prepared including conflicting deadlines, personal issues, sleepless night, etc. Our data suggest that students in structured labs were more likely to label someone as bad group member in the early phase of group formation after only one class session. Instructors should take this into account and wait until they see several reports to identify persistently poor behavior patterns to determine whether intervention is needed.

Instructors could also preemptively provide training in team dynamics and provide frequent opportunities to discuss and reflect on teamwork practices. Interventions such as those developed for business students such as SUIIT (Share, Understand, Integrate, Team Decision) that involve helping all students practice expressing their viewpoints (share), confronting opposing

views (understand), integrating other team members viewpoints (integrate), and finally agreeing on common solutions (team decision) have demonstrated a reduction in detrimental team conflict (O'Neill *et al.*, 2017, 2020). These methods could help students better identify the Students trained in this manner grow more confident over time in providing honest feedback and their team performance has been shown to improve, as well (Donia *et al.*, 2018). Using a tool such as the Enhancing Learning by Improving Process Skills in STEM (ELIPSS) designed specifically for undergraduate STEM students can help students develop a more accurate perception of how to engage in the effective communication, and teamwork (Czajka *et al.*, 2021). Students became much more accurate in their abilities to assess their own teamwork skills when they used the ELIPSS tool multiple times. This involved a primary round in which they reflect on their performance, a second round in which they receive feedback from their instructor, and finally assess themselves again.

### Implications: From an instructor's perspective, is it worth enforcing initial student interactions through randomizing?

We found no benefit to initially randomly assigning students to groups in terms of students' reported conflict scores (task, process, or relationship), satisfaction, or final group and individual grades. Randomly assigning students to groups also failed to significantly influence how students selected their teammates. However, in the small number of cases of PPR students, random assignment increased the willingness for students to report poor behavior and avoid forming groups with those students. In unstructured groups, students would rather stay with a problem group member that they were familiar with rather than confront the issue and change groups. Instructor-mediated intervention is needed to resolve group conflict and help struggling group members in these cases. For example, recommendations from Davies (2009) that ask students to define good and poor group members might help create clearer expectations about group work before these problems accelerate. We found that Likert survey items were not as effective as simple categorical "Yes/No" and "my fellow group member was not performing well and why" questions. Administering our group selection survey could help serve as a better measure to identify struggling students in early group formation and help the instructor to observe and mediate conflicts in real time. Because the number of PPR students is so low, more than 1 wk of reported problems would be recommended to identify significant issues.

### ACKNOWLEDGMENTS

We are indebted to all the students who consented to participate in the study as well as our University of Georgia Biology Education Research Group colleagues for their critical feedback and support. We also wish to thank Daniel Z. Grunspan who generously shared raw data to test the SNA code he developed and published as a primer, Paola Barriga who provided excellent support in R programming for data visualization, Austin Lannen who helped extracting and organizing survey data for SNA, and Kim Brown who helped us with the logistics of conducting our research experiment in the BIOL 1108L laboratory classroom. This material is based upon work supported by the National Science Foundation under Grant - NSF#1659423 – Research

Experiences for Undergraduates - UBERV2 program. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s), and do not necessarily reflect the views of the National Science Foundation.

### REFERENCES

- Aggarwal, P., & O'Brien, C. L. (2008). Social loafing on group projects: structural antecedents and effect on student satisfaction [reports - research]. *Journal of Marketing Education*, 30(3), 255–264. <http://dx.doi.org/10.1177/0273475308322283>
- Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst student team experiences: how a teacher can make the difference [reports - research]. *Journal of Management Education*, 23(5), 467–488.
- Battistich, V., Solomon, D., & Delucchi, K. (1993). Interaction processes and student outcomes in cooperative learning groups. *The Elementary School Journal*, 94(1), 19–32.
- Brickell, J. L., Porter, D. B., Reynolds, M. F., & Cosgrove, R. D. (1994). Assigning students to groups for engineering design projects: a comparison of five methods. *Journal of Engineering Education*, 83(3), 259–262.
- Brickman, P., Lannen, A., & Beyette, J. (2021). What to expect with group work. *Journal of College Science Teaching*, 50(3), 61–67.
- Brunn, J., & Brewes, E. (2013). Talking and learning physics: Predicting future grades from network measures and Force Concept Inventory pretest score. *Physical Review Physics Education Research*, 9, 020109.
- Buchenroth-Martin, C., DiMartino, T., & Martin, A. P. (2017). Measuring student interactions using networks: Insights into the learning community of a large active learning course. *Journal of College Science Teaching*, 46(3), 90.
- Burnham, K. P., & Anderson, D. R. (2004). Model selection and multimodel inference. *A Practical Information-Theoretic Approach*, 63(2002), 10.
- Burt, R. S., Kilduff, M., & Tasselli, S. (2013). Social network analysis: foundations and frontiers on advantage. *Annual review of psychology*, 64, 527–547.
- Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, 57(4), 725–741.
- Carolan, B. V. (2014). Measures for egocentric network analysis. In Carolan, B. V. (Ed.), *Social network analysis and education: theory, methods & applications* (pp. 139–168). New York, NY: SAGE Publications.
- Chang, Y., Brickman, P., & Tanner, K. (2018). When group work doesn't work: insights from students. *CBE—Life Sciences Education*, 17(3), ar42. <https://doi.org/10.1187/cbe.17-09-0199>
- Chapman, K. J., Meuter, M., Toy, D., & Wright, L. (2006). Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. *Journal of Management Education*, 30(4), 557–569.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis* (pp. 1–233). London, UK: Sage Publications.
- Ciani, K. D., Summers, J. J., Easter, M. A., & Sheldon, K. M. (2008). Collaborative learning and positive experiences: Does letting students choose their own groups matter? *Educational Psychology*, 28(6), 627–641.
- Connerley, M. L., & Mael, F. A. (2001). The importance and invasiveness of student team selection criteria. *Journal of Management Education*, 25(5), 471–494.
- Cooper, K. M., & Brownell, S. E. (2016). Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA students. *CBE—Life Sciences Education*, 15(3), ar37.
- Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, 14(1), es1.
- Curşeu, P. L., Janssen, S. E., & Raab, J. (2012). Connecting the dots: Social network structure, conflict, and group cognitive complexity. *Higher Education*, 63(5), 621–629.
- Curşeu, P. L., & Pluut, H. (2013). Student groups as learning entities: the effect of group diversity and teamwork quality on groups' cognitive complexity [reports - research]. *Studies in Higher Education*, 38(1), 87–103. <http://dx.doi.org/10.1080/03075079.2011.565122>
- Czajka, D., Reynders, G., Stanford, C., Cole, R., Lantz, J., & Ruder, S. (2021). A novel rubric format for providing feedback on process skills to STEM undergraduate students. *Journal of College Science Teaching*, 50(6), 48–56.

- Davies, W. M. (2009). Groupwork as a form of assessment: common problems and recommended solutions [information analyses reports - evaluative]. *Higher Education: The International Journal of Higher Education and Educational Planning*, 58(4), 563–584. <http://dx.doi.org/10.1007/s10734-009-9216-y>
- De Hei, M., Admiraal, W., Sjoer, E., & Srijbos, J.-W. (2018). Group learning activities and perceived learning outcomes. *Studies in Higher Education*, 43(12), 2354–2370.
- Donia, M. B., O'Neill, T. A., & Brutus, S. (2018). The longitudinal effects of peer feedback in the development and transfer of student teamwork skills. *Learning and Individual Differences*, 61, 87–98.
- Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE—Life Sciences Education*, 17(4), ar60.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287–322.
- Feichtner, S. B., & Davis, E. A. (1984). Why some groups fail: A survey of students' experiences with learning groups. *Journal of Management Education*, 9(4), 58–73.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (Vol. 3, pp. 245–311). London, UK: Sage Publishing.
- Freeman, L., & Greenacre, L. (2011). An examination of socially destructive behaviors in group work. *Journal of Marketing Education*, 33(1), 5–17.
- Freeman, S., Theobald, R., Crowe, A. J., & Wenderoth, M. P. (2017). Likes attract: Students self-sort in a classroom by gender, demography, and academic characteristics. *Active Learning in Higher Education*, 18(2), 115–126.
- Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1), 103–125.
- Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE—Life Sciences Education*, 13(2), 167–178.
- Hall, D., & Buzwell, S. (2012). The problem of free-riding in group projects: Looking beyond social loafing as reason for non-contribution. *Active Learning in Higher Education*, 14, 37–49.
- Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7), 637–644.
- Hendrickson, B., Rosen, D., & Aune, R. K. (2011). An analysis of friendship networks, social connectedness, homesickness, and satisfaction levels of international students. *International Journal of Intercultural Relations*, 35(3), 281–295.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), nihpa54860.
- Jalajas, D. S., & Sutton, R. I. (1984). Feuds in student groups: Coping with whiners, martyrs, saboteurs, bullies, and deadbeats. *Organizational Behavior Teaching Review*, 9(4), 94–102.
- Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, 44(2), 238–251.
- Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory Into Practice*, 38, 67–73.
- Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education*, 29(6), 751–768.
- Knaak, S. (2004). On the reconceptualizing of gender: Implications for research design. *Sociological Inquiry*, 74(3), 302–317.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353. [https://doi.org/Pii%20S0747-5632\(02\)00057-2](https://doi.org/Pii%20S0747-5632(02)00057-2)
- Krivitsky, P. N., & Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24(5). <https://doi.org/10.18637/jss.v024.i05>
- Krivitsky, P. N. M., Martina Handcock, M. S., Butts, C. T., Hunter, D. R., Goodreau, S. M., Klumb, C., & Bender de-Moll, S. (2021). Exponential Random Graph Models (ERGMs) using statnet. *1st European Conference on Social Networks, Barcelona*. Retrieved from [https://statnet.org/Workshops/ergm\\_tutorial.html](https://statnet.org/Workshops/ergm_tutorial.html)
- Lacey, M. M., Campbell, S. G., Shaw, H., & Smith, D. P. (2020). Self-selecting peer groups formed within the laboratory environment have a lasting effect on individual student attainment and working practices. *FEBS Open Bio*, 10(7), 1194–1209.
- Lawrenz, F., & Munch, T. W. (1984). The effect of grouping of laboratory students on selected educational outcomes. *Journal of Research in Science Teaching*, 21(7), 699–708. <https://doi.org/doi10.1002/tea.3660210704>
- Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 26(1), 61–72.
- Livingstone, D., & Lynch, K. (2000). Group project work and student-centred active learning: Two different experiences. *Studies in Higher Education*, 25(3), 325–345.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: a meta-analysis. *Review of Educational Research*, 66(4), 423–458. <https://doi.org/10.2307/1170650>
- Mahenthiran, S., & Rouse, P. J. (2000). The impact of group selection on student performance and satisfaction. *International Journal of Educational Management*, 14(6), 255–265.
- Malcom-Piqueux, L., & Bensimon, E. M. (2017). Taking equity-minded action to close equity gaps. *Peer Review*, 19(2), 5–9.
- McLeod, P. L., Lobel, S. A., & Cox, Jr, T. H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, 27(2), 248–264.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Mello, J. A. (1993). Improving individual member accountability in small work group settings [reports - descriptive]. *Journal of Management Education*, 17(2), 253–259.
- Metoyer, S. K., Miller, S. T., Mount, J., & Westmoreland, S. L. (2014). Examples from the trenches: Improving student learning in the sciences using team-based learning. *Journal of College Science Teaching*, 43(5), 40–47.
- Myers, S. A. (2012). Students' perceptions of classroom group work as a function of group member selection. *Communication Teacher*, 26(1), 50–64.
- National Center for Science and Engineering Statistics (NCSES). (2021). *Women, minorities, and persons with disabilities in science and engineering*. Special Report NSF 21-321. Alexandria, VA: National Science Foundation. Available at <https://nces.nsf.gov/wmpd>
- O'Neill, T. A., Hancock, S., McLarnon, M. J., & Holland, T. (2020). When the SUIT fits: Constructive controversy training in face-to-face and virtual teams. *Negotiation and Conflict Management Research*, 13(1), 44–59.
- O'Neill, T. A., Hoffart, G. C., McLarnon, M. M., Woodley, H. J., Eggermont, M., Rosehart, W., & Brennan, R. (2017). Constructive controversy and reflexivity training promotes effective conflict profiles and team functioning in student learning teams. *Academy of Management Learning & Education*, 16(2), 257–276.
- Pauli, R., Mohiyeddini, C., Bray, D., Michie, F., & Street, B. (2008). Individual differences in negative group work experiences in collaborative student learning. *Educational Psychology*, 28(1), 47–58. <https://doi.org/10.1080/01443410701413746>
- Premo, J., Cavagnetto, A., & Davis, W. B. (2018). Promoting collaborative classrooms: The impacts of interdependent cooperative learning on undergraduate interactions and achievement. *CBE—Life Sciences Education*, 17(2), ar32.
- Premo, J., Wyatt, B. N., Horn, M., & Wilson-Ashworth, H. (2022). Which group dynamics matter: social predictors of student achievement in team-based undergraduate science classrooms. *CBE—Life Sciences Education*, 21(3), ar51.
- Qualtrics. (2020). *Qualtrics*. [www.qualtrics.com](http://www.qualtrics.com)
- R Core Team. (2022). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research*. In: *R package version 2.2.9*, Evanston, IL.
- Rienties, B., Alcott, P., & Jindal-Snape, D. (2014). To let students self-select or not: That is the question for teachers of culturally diverse groups. *Journal of Studies in International Education*, 18(1), 64–83.

- Ruedi, B., Feder, M., Thompson, D., & Conn, E. (2020). AAAS Pavilion: SEA Change-Making diversity, equity, and inclusion in STEM the norm for higher education. *2020 Annual Meeting*.
- Saldaña, J. (2021). *The coding manual for qualitative researchers*. Thousand Oaks, CA: SAGE Publications Limited.
- Schwandt, T. A. (2014). *The Sage dictionary of qualitative inquiry* (3rd ed.). Los Angeles, CA: Sage Publications.
- Seifert, K., Hurney, C. A., Wigtill, C. J., & Sundre, D. L. (2009). Using the academic skills inventory to assess the biology major [reports - descriptive]. *Assessment Update*, *21*(3), 1–2.
- Shimazoe, J., & Aldrich, H. (2010). Group work can be gratifying: Understanding & overcoming resistance to cooperative learning. *College Teaching*, *58*(2), 52–57.
- Slavin, R. E. (2014). Cooperative learning and academic achievement: why does groupwork work? *Anales de Psicología/Annals of Psychology*, *30*(3), 785–791.
- Speer, P. W., Jackson, C. B., & Peterson, N. A. (2001). The relationship between social cohesion and empowerment: Support and new implications for theory. *Health Education & Behavior*, *28*(6), 716–732.
- Springer, L., Donovan, S. S., & Stanne, M. E. (1999). Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis [Information Analyses]. *Review of Educational Research*, *69*(1), 21–51.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage Publishing.
- Strong, J. T., & Anderson, R. E. (1990). Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, *12*, 61–67.
- Sullivan, A. (2020). Sex and the census: Why surveys should not conflate sex and gender identity. *International Journal of Social Research Methodology*, *23*(5), 517–524.
- Takeda, S., & Homberg, F. (2014). The effects of gender on group work process and achievement: An analysis through self-and peer-assessment. *British Educational Research Journal*, *40*(2), 373–396.
- Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE—Life Sciences Education*, *17*(3), rm2.
- Theobald, E., Eddy, S., Grunspan, D., Wiggins, B., & Crowe, A. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PloS one*, *12*(7), e0181336.
- Van der Vegt, G. S., Emans, B. J., & Van der Liert, E. (2001). Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. *Personnel Psychology*, *54*(1), 51–69.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. (pp. 1–332). Cambridge, UK: Cambridge University Press.
- Watkins, R. (2004). Groupwork and assessment. In Davies, P. (Ed.), *The Handbook for Economics Lecturers* (pp. 1–24). York, UK: Economics Network of Higher Education Academy. Retrieved from [www.economicsnetwork.ac.uk/handbook/printable/groupwork.pdf](http://www.economicsnetwork.ac.uk/handbook/printable/groupwork.pdf)
- Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, *36*(3), 590–602.
- Wilson, K. J., Brickman, P., & Brame, C. J. (2018). Group work. *CBE—Life Sciences Education*, *17*(1), fe1. <https://doi.org/10.1187/cbe.17-12-0258>