Inclusive Study Group Formation At Scale

Sumer Kohli*
sumer.kohli@berkeley.edu
UC Berkeley

Neelesh Ramachandran*
neelesh.r@berkeley.edu
UC Berkeley

Ana Tudor*
anamtudor@berkeley.edu
UC Berkeley

Gloria Tumushabe*
gloriatumushabe@berkeley.edu
UC Berkeley

Olivia Hsu
owhsu@stanford.edu
Stanford University

Gireeja Ranade
ranade@eecs.berkeley.edu
UC Berkeley

ABSTRACT

The student peer-group is one of the most important influences on student development. Group work is essential for creating positive learning experiences, especially in remote-learning where student interactions are more challenging. While the benefits of study groups are established, students from underrepresented communities often face challenges in finding social support for their education when compared with those from majority groups. We present a system for flexible and inclusive study group formation that can scale to thousands of students.

Our focus is on long-term study groups that persist throughout the semester and beyond. Students are periodically provided opportunities to obtain a new study group if they feel their current group is not a good fit. In contrast to prior work that generates single-use groups, our process enables continuous refinement of groups for each student, which in turn informs our algorithm for future iterations.

We trialed our approach in a 1000+ student introductory Electrical Engineering and Computer Science course that was conducted entirely online during the COVID-19 pandemic. We found that of all students matched to study groups through our algorithm, a large majority felt comfortable asking questions (78%) and sharing ideas (74%) with their group. Students from underrepresented backgrounds were more likely to request software-matching for study groups when compared with students from majority groups. However, underrepresented students that we did match into study groups had group experiences that were just as successful as students from' majority groups. Students in engaged, regularly participating study groups had more positive results across all other indicators of the study group experience, and certain positive group experiences were associated with higher exam scores overall. Furthermore, students performing at a B-level on the first class midterm, who participated in high-quality software-matched study groups, demonstrated higher final exam scores than students in lower-quality groups.

L@S '22, June 01–03, 2022, Roosevelt Island, NY

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

KEYWORDS

education, matching algorithms, group formation, remote learning

ACM Reference Format:
Sumer Kohli, Neelesh Ramachandran, Ana Tudor, Gloria Tumushabe, Olivia Hsu, and Gireeja Ranade. 2022. Inclusive Study Group Formation At Scale. In *Proceedings of Learning @ Scale 2022 (L@S '22).* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Group work is an important component of engineering, where many projects (both at the university level and in real-life) are team-based. Students often form informal study groups to work on homework problems, study for exams, and do course projects. Astin's influential paper calls out the student peer-group as the most important environmental influence on student development [4]. Since learning is a social experience [60], peer-learning is essential [3, 9, 50, 53], and peer-groups and social networks can have a huge impact on the success of individuals [11, 12, 21, 35, 42, 58], providing social support is critical for students. The positive impacts and importance of study groups has been extensively studied [7, 15, 17, 18, 24, 31, 33, 34, 38, 41, 44, 52].

However, the move to online learning due to the COVID-19 pandemic made it significantly more challenging for students to meet each other and work together [20, 62, 64]. And even under pre-pandemic circumstances, study groups have not been equally available to all students. Students from underrepresented communities struggle to make essential social connections and study-groups and may feel excluded [3, 8, 10, 13, 40, 45, 49, 57]. This is particularly true at UC Berkeley where our introductory classes often have 1000's of students, and very small numbers of students from certain groups (e.g. less than 10 Black students in a class of 1000).

For instance, back in 2013, a news outlet ran a story about UC Berkeley [2], with a quote from the then director of the African American Student Development office:

> *"A black student might be in a science course, ... and the professor says, 'Okay, everybody has to have a study group.' Nobody picks them for a study group. They first have to show that they can get an A before they get selected."*

Unfortunately, this problem still persists today. In fact, the president of the Black Graduate Engineering and Science Students at UC Berkeley recently wrote [6]:

> *"... I've stood by black CS undergrads who were constantly ostracized in group projects, often the last ones to be picked. Their peers regularly perceive them to*

be less knowledgeable and capable in carrying coding projects. I have had to console crying teenagers scarred from the experience, questioning, for the first time in their life, whether being black was worth it. Imagine that. Their only path to recognition in that environment was through a bogus burden of excellence. What ever happened to the opportunity of being average yet respected?"

To address these challenges, we designed a flexible, inclusive, and scalable custom software solution to automatically group students based on their preferences while enforcing inclusive group formation. Our focus here is on developing an end-to-end system that can be used to develop long-term study groups (for one course or more). We distinguish study-groups (which are informal, fluid, open-ended, and student-driven) from project-groups (which are officially tied to grades, fixed, and have a predetermined focus). Our goal is to have long-term (one semester or more) study groups that evolve into friendships and support systems, not just groups to work on a particular assignment or project. With this perspective, group reassignment based on feedback from students is a key feature of our system, and this distinguishes our work from other approaches [22, 32].

We first trialed our system in an introductory Electrical Engineering and Computer Science (EECS) course at UC Berkeley, consisting of over 1,000 students, and it has subsequently been used in multiple other lower-division and upper-division courses in the EECS department. The online semester provided a unique opportunity to study the impact of intentional and structured study groups for students, since the natural in-person social process by which students form study groups was disrupted.

1.1 Main contributions

This paper discusses the development of a scalable survey and software system to generate long-term study groups. A key feature of our system allows students to 'try out' a study group for a few weeks and then request a reassignment to a new group if their first group is not working out for them. They may do this multiple times throughout the semester. This gives the system and course instructors the opportunity to improve group matches over time and improve the student experience, and sets up a system by which our system can *learn* better groups over time.

Our matching system can considers multiple concrete factors (e.g. student schedules, demographics) as well as factors such as student motivation in generating the groups. We build on previous works such as [22, 32]. These works identified that solo/singleton status (students who are the only representative of a given demographic in their study group) in a group may harm students. However, the optimization-based frameworks of [22, 32] are likely to avoid solo status, but cannot guarantee it. Our partitioning based system (see Sec. 2.3) can provide guarantees for this. In addition, our matching system and code can scale to thousands of students, unlike [22]. Our code will also be available open-source at [29]. Our focus is on a system for voluntary long-term student interactions, without any connections to grades, and we build other infrastructure (e.g. guidelines for study-group operation, discussion-based homework problems, and more in Sec. 2.1) to support this. This contrasts with the shorter-term project focus that other systems may have.

In evaluating our system, we offer a case study of the impact of study groups during the first exclusively remote semester at UC Berkeley. We find that most (over 50%) of the students in groups we assigned had positive experiences across four of five quality indicators (see Sec. 3.1). Notably, more than 74% were comfortable asking questions and sharing ideas in their groups (see Sec. 3.1). We find that our system was particularly useful to students from underrepresented demographic groups — these students were more likely to request a study group via our system than form their own study groups (see Sec. 3.1.2). Furthermore, for those students who were matched into groups by our system, we find that students of underrepresented race and gender demographics had similarly positive study group experiences to majority groups (see Sec. 3.2). In analyzing group associations with academic performance, feeling comfortable sharing ideas and asking questions in study groups was significantly associated with increased midterm scores (see Sec. 3.3).

1.2 Background

We are building on extensive work to understand the importance of social support for students, as well as on work to construct high-quality teams and evaluate their impact. Here, we discuss representative works in different areas.

Social support in the classroom: Social support matters everywhere, and especially in the classroom. It has been shown that socially-connected employees are less likely to quit or change their jobs [36, 37]. Social-networks can encourage individuals to persist in challenging environments where they might otherwise drop out [37]. Social integration, institutional commitment, and intention matter in students' retention, and students with broader, well-connected networks are more likely to persist in college [55]. Similarly, students' friends' retention was found to have a large impact on their own retention [19], and their integration into the social network matters as well [65]. Simple psychological interventions that improved social connectivity in the classroom have also increased retention [58]. Anxiety in the classroom can be mitigated to some extent by feelings of interpersonal familiarity and acceptance by peers [59].

Most students entering university in Fall 2020 did not have access to the usual meetups, orientations, or events that allow them to develop these social support networks. This widens the divide between those who come to Berkeley from feeder schools (where multiple students from the same school/community college are in university together), versus those who are the only representative from their school/area. Based on student interviews, we know that students from underrepresented groups are less likely to have extensive peer networks when they come to Berkeley. This made it imperative that we institutionally help students find social support and peer groups. For many first-year students, the study groups were one of the few ways they connected with others. This also made it important to indicate to students that these groups are not limited to working on specific projects, but are really opportunities for them to develop relationships with their peers.

Group composition: The details of the social support matter very much, and the specific composition of a study groups can also have a great impact on their dynamics. Individuals experience "solo status" when they are the only members of their social category (e.g., gender or race) present in an otherwise homogeneous group. Field studies and surveys indicate that members of socially disadvantaged groups, such as women and racial minorities, have more negative experiences as solos than members of majority groups, such as white and male-identifying students [40]. Solo status can lead to lower public performance, which hurts retention of women and racial minority students [57]. Tokenism, or perfunctory/symbolic efforts to be inclusive to members of underrepresented groups [25], can lead to poor performance of students from the underrepresented groups [49]. [56] highlights that retaining women in undergraduate computing can be facilitated through collaborative learning in classrooms by creating a sense of community. In fact, a study about improving the persistence of first-year undergraduate women in computer science showed that women appreciated all-women support groups because they provided opportunities to discuss academic and social issues concerning the major [43]. In another study, women who were randomly assigned to male majority teams were less willing to become team leaders than women assigned to women majority teams [8]. It is further important to consider intersectional identities in this [46–48]. This literature leads us to strongly believe that we should ensure that no students experience solo or singleton status in study groups.

At the same time, homogeneity in groups is not necessarily ideal. We know that broader networks that go beyond self-identified student affinity groups are better for student retention and success [55]. Furthermore, the work of Hong and Page [23] tells us that diverse groups of problem solvers can outperform high-ability groups. Many other works also emphasize the importance of diversity in groups [54, 61]. Thus it is important that students not get siloed into interacting only with students from familiar racial, ethnic and gender identities and learn to effectively interact across these groups.

A variety of approaches have been explored by researchers to form impactful student groups. For example, the at times complementary topic-master or skills may be helpful [14, 16]. Another approach has been to create groups of students who approach problems very differently [1]. This system also grouped students with complementary learning styles to avoid conflicts stemming from incompatible learning types. Yet another work used parameters such as laboratory sections and out-of-class schedules to form student groups using software [63]. We do not focus on skills/topic-mastery in our context, since our goal is not the completion of a particular project, however our system does try to check for compatible schedules. While learning styles have not been considered in this version of our algorithm, we hope to consider them in the future.

Group formation tools: There have been many studies for understanding the impact and construction of student teams [7, 15, 17, 18, 22, 24, 31–34, 38, 41, 44, 52]. For instance, [17] considers teams for specific projects, and discusses the balance between skills across team-members, practical constraints such as schedules and also factors such as motivation. [18] presents an algorithmic approach and tool to ease instructor burden in group formation, which is similar to our goals. However, this tool is not open-source, and does not incorporate any machine learning components that we aim to bring in. [38] specifically considers teams for in-lecture exercises that must be rapidly formed. [39] includes a simulation-based approach for adaptively learning better groups. [51] focuses on real-time groupings for Massive Online Open Courses (MOOCs) that are formed and dissolved on very short time-scales using an optimization based approach. [52] also provides guidelines for successful teams in the MOOCs, taking into account the high dropout rates in MOOCs.

The most notable tool for student team formation is the CATME Teammaker [32]. CATME is a flexible and powerful tool, which focuses on student-teams for projects, and can incorporate demographic constraints, schedule constraints, course-matching constraints and more. Instructors have the option to indicate how important each of these constraints are to them. In addition, this framework offers a suite of tools that can evaluate the performance of teams through peer-evaluation, for example as in [5]. Unfortunately, the full version of their work is not available open source. CATME uses a non-convex optimization problem on the backend to identify the groups, with a hill-climbing approach to finding the optimum. This approach likely only leads to the discovery of a local optimum, and while constraint violations may lead to a decrease in a team-score, and thus make the team less-likely to be chosen, the optimization framework cannot guarantee this. Our approach avoids this by using a partitioning approach. Their framework does not account for the added value of having diversity across race and gender in teams, which we are building explicit checks for.

Another tool of note is Gruepr, which builds on CATME but provides an open-source implementation [22]. However, their algorithm only works for partitioning up to 300students at a time, while ours scales to 1000's of students. It uses a genetic algorithm to explore the optimization landscape — with a large number of random seeds it has a chance to avoid local optima. However, one still cannot provide guarantees of constraint non-violation. Our partitioning based approach allows us to prioritize certain constraints over others as deemed important by the instructor, with is different from the optimization approach of CATME and Gruepr.

2 METHODS

2.1 Course setup and timeline

We trialed our group-matching algorithm in a first-year introductory electrical engineering and computer sciences class at UC Berkeley in Fall 2020. This was the first entirely-remote semester due to the COVID-19 pandemic, and there was a great deal of uncertainty among both students and faculty around learning and teaching. The course had one lecture and two discussion sections, as well as a homework and lab assignment, all on a weekly basis.

In keeping with our goal to create study groups that would last for the semester and beyond, students opted in to use our study-group matching system — which was completely optional with no grade incentives — and were also allowed to form their own groups and simply indicate that to us. If students found that their assigned or self-formed study groups were not working well for them, they could request a different study group during a "reassignment round" which we periodically conducted throughout the semester. We executed the same matching algorithm for the students who requested

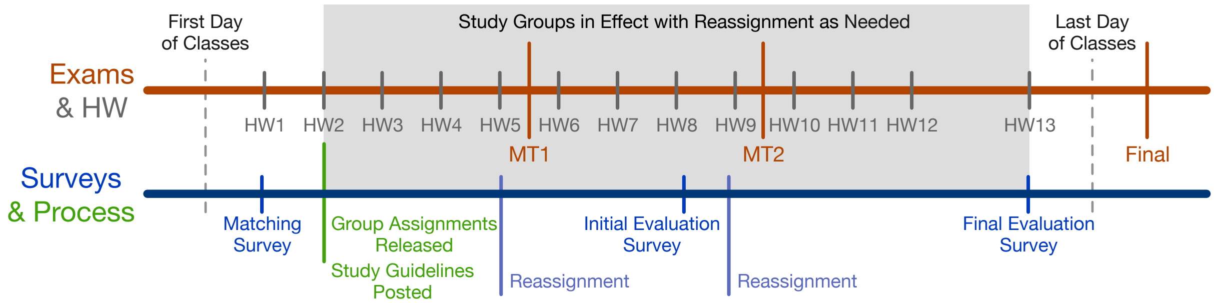

Figure 1: The timeline of the group assignment process, from initial surveys to the final evaluation, in the context of the homeworks and exams in the class.

reassignment and formed new study groups from those students. Since very few reassignments were requested in the third round of reassignment, we did not conduct further rounds.

A timeline of the process in the context of our class is shown in Fig. 1. Matching surveys were sent out at the very beginning of the semester as part of the first homework assignment (see [26, 30]), and groups were released to students online a week later. In subsequent semesters, we automatically emailed each student group together in order to open a line of communication between them. Feedback[1] on the study groups was collected through a mid-semester and a final evaluation survey (see [27]). In this paper, we present the results based on the final evaluations of students who consented after all rounds of reassignment were completed.

To encourage better communication within the groups, we released a set of guidelines for effective group work that are available at [28]. We additionally wrote new homework questions that were open-ended to encourage discussion and sharing of ideas and opinions (with no "correct" answers). Since most groups were only able to meet virtually, these questions were designed as coursework-related "ice-breakers".

2.2 Collecting student preferences

To understand students' needs in study groups, we first conducted informational interviews with students from various cultural and identity-based STEM student organizations as well as the Director of Student Diversity in EECS. We tailored our process based on their feedback, and had our final released surveys vetted by those students and staff members.

In order to help students meet suitable study partners, we asked a series of questions in the initial surveys released with the first homework. This included information about their current timezone, preferred meeting days and times, their coursework background, their year, as well as optionally, their race and gender identity. We also asked students how important the course was to them, what other courses they are currently taking, and their preferred discussion section attendance times (to allow students to attend discussion section together if possible).

[1]We obtained student consent and IRB approval of our human research protocol titled *Creating positive experiences for first-year underrepresented students through scalable automated inclusive study group formation.*, with Protocol ID '2020-08-13526' and Approval Date 'September 25, 2020.'

Our aim was to create a structured process for supporting and creating informal study groups, all while ensuring inclusion of students from underrepresented groups. Since our goal was not to form student teams to complete particular projects, we did not consider particular skills or GPA type information in making groups.

2.3 Group matching algorithm

At a high-level, our algorithm partitions the class repeatedly into subdivisions based on student constraints until the divisions are small enough to become study groups, and then ensures that our inclusivity and diversity constraints are met. Our goal was to form study groups sized between four to eight students — we allowed this flexibility in size for greater matching freedom. The general algorithm and process is outlined in Fig. 3.

In more detail, our algorithm has three main phases: (1) Multi-partitioning, (2) Group division, and (3) Postprocessing. Each phase of the algorithm is explained as follows.

Multi-partitioning: The algorithm is configured with a priority ordering for constraints and matching preferences, which are used to partition the students into subdivisions. We repeatedly partition the class to maximally match student preferences until either the subdivisions are too small or until we have exhausted the matching criteria. Fig. 2 shows an example of this process, using student timezone as the first partitioning criterion and then using scheduling constraints as the second criterion.

While this partitioning approach would ensure that no constraints are ever violated, strict partitioning can be too restrictive when it comes to select-all-that-apply questions, which is why we use multi-partitioning. For example, consider a scheduling constraint, where students are asked to choose days of the week when they would prefer to meet. In implementing multi-partitions for this criteria, students are included in *all* of the subdivisions where they could feasibly meet. Thus, a student who indicated either Monday or Wednesday would work for them would initially be placed in both subdivisions simultaneously. We then iterate over subdivisions and prune other copies of the students in order to achieve well-balanced partitions. For example, if the Monday partition is substantially larger than the Wednesday partition, then students who indicated both availabilities would preferentially be maintained in the smaller Wednesday partition.

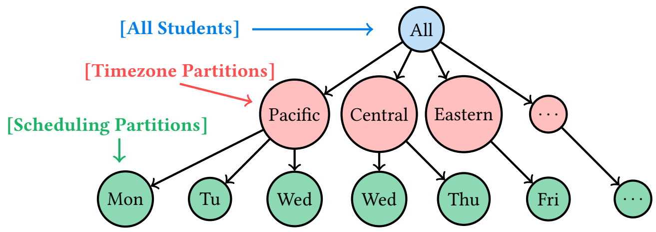

Figure 2: An example of class partitions modeled in a tree-like figure. In the first stage of this partitioning process, we separate students by which timezone they belonged to (for example). Then, we further partition by which days they are available. Note that in each subgroup, if a student *could* be part of multiple of the groups below them, they are placed into *all of these partitions.* Note that the diagram is symbolic, and each criteria had more layers and subgroups formed.

For Fall 2020, we used the following priority-ordered partitioning sequence: (1) Timezone, (2) Meeting day, (3) Meeting time, (4) Class year, (5) Background classes.

Group division: Within each subdivision, students are clustered into affinity groups based on their indicated preferences. We sequentially form "soft" clusters of students based on criteria not used for partitioning such as demographics, other enrolled courses, motivation in the course and so on. Students with similar responses are clustered together. These clusters are then split into study groups that obey the sizing constraints, although groups may be formed across cluster boundaries if necessary.

Singletons and homogeneity: Since the earlier clustering is done in a soft manner, it does not provide any hard guarantees. In order to guarantee that no student has solo status, we further iterate over all groups and identify students who are singletons. Such students are partnered into pairs or triples and then combined into larger study groups. To the extent possible, intersectional identities are taken into account in generating these pairs. If suitable partners cannot not be found within the same subdivision, subdivisions are allowed to be violated.

Since homogeneity of a single gender or race within a group was also undesirable, we checked to ensure that there were no completely racially-homogeneous groups during our first algorithm execution. Subsequently this check has been incorporated into the algorithm. However, given the demographics of the student body — where there are a large number of men and Asian students — it was not possible to guarantee non-homogeneity for all groups, and some groups from our majority demographics were homogeneous.

3 ANALYSIS AND EVALUATION

3.1 Overview of student impact

In our analysis below, we only consider students who consented to our use of their data with a total of 477 consenting students. The demographic distribution across categories can be seen in column A of Fig. 6. Our final evaluation survey (see [27]) had five questions related to the quality of the study group experience for each student: (1) the frequency of interaction of the study group, (2) the number

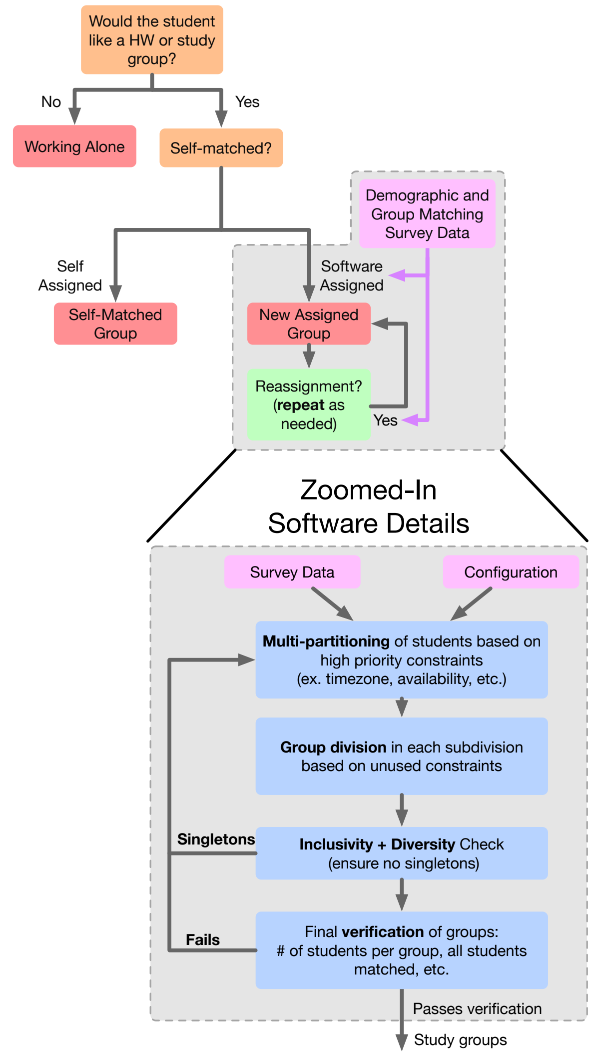

Figure 3: Group matching process flowchart along with a detailed software diagram (in gray). First, students are routed based on whether they want a study group and whether they already have a self-matched group in mind. For the remainder of students, our software uses the student preferences to match them into groups. The algorithm first partitions students based on the most important criteria, forms groups within subdivisions to maximize matches on secondary criteria, and finally then performs checks on the demographic and preference-based breakdown within each group. After ensuring that no singletons are present and performing additional verification, the resulting groups are finalized and displayed.

of students in the study group which regularly participated in interactions, (3) the comfort of the student in sharing ideas with their group, (4) the comfort of the student in asking questions in their group, (5) whether the student wants to take future courses with their group. We were also interested in understanding associations

Group Quality Indicator	Software-Assigned	Self-Formed
Future courses	52.4% (47.0%, 57.8%)	97.2% (94.5%, 99.9%)
Group interaction	56.6% (51.3%, 61.9%)	95.8% (92.5%, 99.1%)
Group participation	62.3% (57.1%, 67.5%)	95.1% (91.6%, 98.6%)
Comfort asking questions	74.2% (69.6%, 78.9%)	93.7% (89.7%, 97.7%)
Comfort sharing ideas	78.4% (74.0%, 82.9%)	95.1% (91.7%, 98.6%)

Figure 4: Percentages (with confidence intervals) of students who reported positive responses to the five group quality indicators. First column: Students in software-matched groups. Second Column: Students in self-formed groups.

with how many times a student requested reassignment to a new group.

When analyzing responses in these factors, we classify "positive" responses, i.e. indicators that the student had a good experience in their study group, as follows:

- Future courses: students state they hope they can, or definitely will, take future courses with their group
- Group interaction: students interacted with their group once a week or more
- Group participation: Some, most, or all members participated in group interactions
- Comfort asking questions: Students agreed or strongly agreed that they feel comfortable asking questions in their group
- Comfort asking questions: Students agreed or strongly agreed that they feel comfortable sharing ideas in their group

When looking at all the students who received groups from us, we see evidence of overall positive group experiences (see Fig. 4 for response percentages and confidence intervals). For example, 74% of students report feeling comfortable asking questions in their groups, 78% of students report feeling comfortable sharing ideas. We performed significance tests in comparison to the null hypothesis that 50% of students would have positive study group experiences and 50% of students would have negative study group experiences. For all quality indicators, over 50% of our sample had positive experiences. Performing analyses with proportion z-tests at $\alpha = 0.05$, over 50% of all software-matched students had positive experiences with group interaction, group activity, and comfort asking questions and sharing ideas, as can be noted in the second column of Fig. 4.

3.1.1 Correlations in group quality indicators. To preface the upcoming analysis, we note that positive responses to certain study group quality indicators tended to correlate with some more than others, using the Pearson correlation coefficient (see Fig. 5). Specifically, positive responses to the group interaction and group activity questions correlated highly with each other, and at moderate levels with the other three indicators. Additionally, the comfort indicators correlated highly with each other.

Figure 5: Pearson correlation matrix of positive responses to study group quality indicators.

Demographic group	(A)	(B)	(C)	(D)
Women	139	103	74.1%	(66.8%, 81.4%)
Men	326	221	67.8%	(62.7%, 72.9%)
Gender non-conforming/ Genderqueer	2	2	100%	-
Other/Prefer not to answer	10	10	100%	-
Black/ African American	7	6	85.7%	(59.8%, 100%)
Hispanic	39	28	71.8%	(57.7%, 85.9%)
Native American/ Alaska Native/ Hawaiian Native	9	6	66.7%	(35.9%, 97.5%)
White	86	66	76.7%	(67.8%, 85.7%)
Asian	345	233	67.5%	(62.6%, 72.4%)
Other/Prefer not to answer	27	20	74.1%	(57.5%, 90.6%)
Freshman	323	214	66.2%	(61.1%, 71.4%)
Junior or Senior Transfer	66	53	80.3%	(70.7%, 89.9%)

Figure 6: Demographic distribution of students in our sample. Column (A) shows the total counts of students in each demographic subgroup who either self-formed or requested software-assigned groups. Column (B) shows the count of students of each demographic subgroup who requested software-assigned groups, and column (C) shows percentages of students in subgroup who requested software-assigned study groups. Column (D) shows population-level confidence intervals on percentages in column (C).

Whether a student indicated a desire to take future courses with their group did not strongly correlate with a student's comfort asking questions and sharing ideas in the group. This indicates that the comfort in a study group may not be fully aligned with whether a student wants to take future courses with a group. This may be because future academic logistics impact the desire to take future courses, which is independent of the quality of the study group.

In general, we consider these differing correlations to point towards different forms of well-functioning study groups, and do not define any particular combination to be an ideal study group. We therefore will conduct analysis of study group quality across individual indicators.

3.1.2 Comparison to self-formed groups. In our dataset, 143 students reported having self-formed groups and 334 students asked for software-matched groups. Self-formed groups often were a result of students knowing each other prior to the start of the course. These may be students who are not first-years, or are coming from feeder high-schools, i.e. generally students who are more connected in the social network. Self-formed groups reported extremely positive results across all the metrics, as can be observed in the rightmost column of Fig. 4. One perspective is that self-formed groups can be viewed as a gold standard, where students already know they feel comfortable and productive when working with students they choose.

We notice that students of underrepresented demographics were more likely than majority students to request study groups. For example we observe, in the second section of column C of Fig. 6, that Black and Hispanic student proportions of requesting software-matched study groups are higher than Asian student proportions. Similarly, proportions of women and gender non-conforming / genderqueer students requesting software-matched study groups are higher than proportions of men, observable in the first three rows of Fig. 6. Therefore, even though self-matched groups seem to have more positive responses than software-matched groups, it is useful to a large population of students, and especially traditionally underrepresented students, to provide inclusive study group options.

3.1.3 Requesting reassignments. A small but sizeable proportion of students requested reassignment to new groups at least once, with 27% of students requesting only one reassignment, 2% of students requesting two reassignments, and <.5% of students requesting three reassignments. Using 2-sample proportion z tests, Hispanic students (and White students) requested reassignments significantly more often than non-Hispanic (non-White) students, respectively indicating that initial group matches did not work out as well for Hispanic and White students. Asian students requested reassignment significantly less often than non-Asian and non-White students. These results support the decision to offer at least one reassignment along the course of a course term, as otherwise there may be students, occasionally disproportionately of underrepresented demographics, left with lower-quality groups from the outset. We also note in Section 3.3 that never requesting reassignments can be associated with higher performance on exams. There may be different reasons for this and we do not have any insight into possibly hidden variables, so it is hard to establish any causality. However, we believe it is important to have a quick turnaround for reassigning groups to dissatisfied students.

3.2 Student impact within demographic groups

To evaluate if our system worked well for students from underrepresented groups, we compare student responses across demographic groups — column A in Fig. 6 provides the count breakdown across these demographic groups.

In general, students who identify as being from an underrepresented racial or gender demographic did not demonstrate significant differences in study group quality compared to majority demographic groups, with a few exceptions. We interpret this positively — that with high confidence, the process of forming groups without racial or gender singletons enables underrepresented groups to have study group experiences on par with other groups, which was a major goal for us. We did, however, notice some differences in study group outcomes based on student year; there are likely many reasons for this.

We were not able to control against courses forming study groups through other methods, so we focus our analysis on normative statements to be made about experiences in our pilot group. We additionally do not draw any gender or race comparison conclusions about students who preferred not to state their gender or racial identity.

3.2.1 Group quality vs student gender. Using 2-sample proportion z-tests, we found no significant difference between students in the majority group (men) and those who identified as any other gender (the yellow (men) and green (women) bars (representing percentage positive responses for each question) are of similar heights in Fig. 7, and not that different from the other two groups either). Additionally, confidently more than half of all women and men students in software-assigned groups responded positively to 4 of 5 group quality indicators, aligned with our general population results (see plots 2-5 of group quality indicators in the student gender section, in Fig. 7). One student identifying as gender non-conforming / genderqueer (GNC) had positive group experiences in all categories, and one other GNC-identifying student had negative group experiences in all categories. Due to the very small count of GNC-identifying students, we are not able to generalize these results.

3.2.2 Group quality vs student race. Within differing racial subgroups, we find some slight differences in student group experiences. All statistical significance tests for smaller subgroups (Hispanic, Black, and Native American / Alaska Native / Native Hawaiian students) were conducted using the Fisher exact test, and for larger subgroups they were conducted using 2-sample proportion z-tests. Most notably, no significant differences were found between all students of majority racial groups (White or Asian) and all students of underrepresented racial groups (non-White and non-Asian) — all analyses below on individual racial subgroups support this finding.

Hispanic-identifying students did not have statistically significant differences in their responses, indicating that final group assignments were comparable in quality to other demographic subgroups. We also note that we can say with 95% confidence that the majority of Hispanic students felt comfortable asking questions and sharing ideas in their final groups. These findings can be observed in the second row of Fig. 7, when qualitatively comparing the height of the percentage bars of Hispanic students to other subgroup bars.

Students identifying as Black / African-American (AA) did not show statistically significant differences in any study group indicators, in comparison to non-Black/AA students. For the six Black/AA students who participated in software-matched study groups, there were very positive experiences across all indicators. All six students felt comfortable asking questions, and five of six felt comfortable sharing ideas in their groups (Fig. 7). The small sample size limits our interpretations here, but this is promising.

Students identifying as Native American, Alaska Native, or Native Hawaiian (referred to shorthand as Native American), we observed four of the six students indicating positive responses

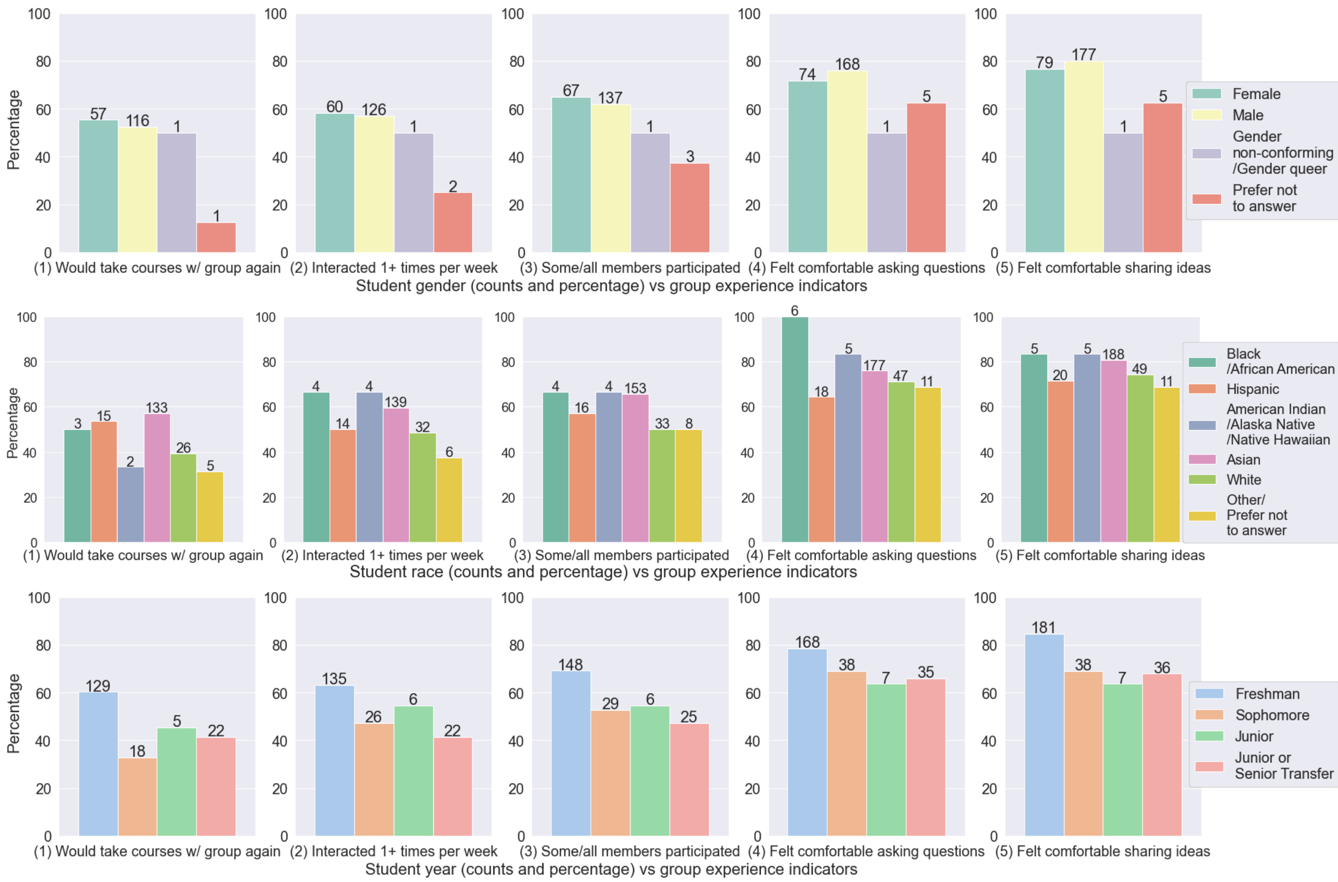

Figure 7: Comparisons between students of differing demographics, in percentages (y-axis) answering positively to key study group quality questions, and counts (displayed on bars) answering positively to key study group quality questions. For student race, students of mixed race were counted in each of the racial categories they indicated they identified with.

across all study group indicators, and five of the six feeling comfortable asking questions and sharing ideas in their groups (Fig. 7). Again, due to small sample sizes, generalizing statements about their experiences or comparisons to other groups cannot be made with confidence.

White-identifying students had significantly less positive study group experiences than non-White students in a few categories, namely: wanting to take courses with their study groups less often, number of students who participated in their groups (seen in the second section of Fig. 7).

Asian-identifying students had significantly more positive study group experiences than non-Asian students (Fig. 7). However, when performing tests of significant difference between Asian students and all other non-White or non-Asian students, we found no significant differences in study group quality. Similarly, we found no significant differences in study group experiences between White students and all other non-White and non-Asian students. This may indicate that there are different social and cultural experiences in study groups between two demographics that may be considered majority groups in engineering classrooms at UC Berkeley, but non-majority groups do not face disproportionately different study group experiences in comparison to either.

3.2.3 Comparisons in group quality vs student year. When considering a student's year in school at the time of the course, it became clear there are extremely different study group outcomes depending on a student's year. In many ways, our group matching software was geared towards first-year students, who are the least connected socially.As such, it is promising to note that Freshman students had overwhelmingly positive study group experiences, with significantly higher proportions of positive responses to all five study group indicators in comparison to non-freshmen (Fig. 7).

Transfer students, however, had significantly less positive responses in the indicators of: group group interaction, group activity, and comfort sharing ideas in their group (Fig. 7). This is an additional possible indicator that post-processing matching to ensure non-singletons of certain groups may actually be quite important, our software matcher in Fall 2020 did not fully ensure against student year singletons[2]. Although we partition on student year to ensure groups predominantly comprised of the same year, transfer

[2]Later iterations of our software allow this.

students may need to be considered an underrepresented group in themselves, and other as of yet unaddressed factors may impact their study group success.

3.3 Association of high quality study groups and student grades

Although there is clear self-contained benefits to having a high-quality study group, we conducted analyses to verify whether indicators of higher-quality study groups might independently correlate with higher test scores. We find that feeling comfortable asking questions and sharing ideas in a study group is a key indicator of student success, and that study group environments which are likely to encourage these feelings are also highly effective academic supports.

At the classroom-level, all software-matched students demonstrated significant associations between feeling comfortable sharing ideas in their groups, and increases in both midterm and final exam scores, as seen in the top section of Fig. 8. For example, in the "Final score" column of the second row in Fig. 8, we observe that students who felt comfortable asking questions averaged 72.22 on the final, versus students who didn't feel comfortable asking questions averaging near 66.08, with a confidence of 98.9% of the first score being higher. Similar results were seen around comfort sharing ideas, as well as not requesting reassignments. All software-matched students also demonstrated significant associations between feeling comfortable sharing ideas in their groups, and increases in Midterm 2 (MT2) and Final scores.

These findings might be explained by the external factors of a student's academic proactiveness and general social comfort to be associated with their better performance as a student, and we cannot draw any firm conclusions. However, given many students feel comfortable in groups and are not requesting reassignments, it could also be inferred that groups afford them the socio-academic support to exercise their academic proactiveness.

Specializing to freshmen, our largest but also one of our target demographics, we find even larger gains on exam scores for those who did not request reassignments. These results are especially promising given that over 75% of freshmen never requested regroup assignments.

Finally, we explored the impact of study groups on students in different grade ranges. The most impact was seen in the B-range students. Students with B-range MT1 scores (between 68 and 89 percent, based on the class grading scale) saw significantly higher Final exam scores associated with high group interaction groups, and with feeling comfortable asking questions in the group. Additionally, knowing that the general distribution of Final scores was much lower than MT1 scores, students at the B range demonstrated significantly lower decreases from MT1 to the Final associated with several factors: frequently interacting groups, multiple group members participating, and feeling comfortable asking questions in their groups (Fig. 8). These results suggest that for students who entered the class performing at a mid-level, being part of active and comfortable study groups benefited their grade. This range of grades constitutes a significant portion of the student population.

We also considered students who received A, and C to below C-level, grades on MT1, but found no significant associations between their scores and group quality indicators. This may be because students scoring at an A-level are well-prepared to succeed independently, and students with C-level grades may have external factors affecting performance and their ability to engage effectively with groups.

However, our findings above differed significantly when considering transfer students — this demographic group had generally lower exam performance in general. In addition, students desiring to take courses with their group again, having higher group interaction in group, and having higher group member activity, tended to be associated with having lower exam scores. It is difficult to reason about this phenomenon, due to the large variability of transfer student backgrounds and personal situations.

3.4 Student anecdotes

While we were unable to conduct systematic student interviews to understand the student experience beyond the surveys, some students voluntarily let us know about the impact the groups had on them, through written feedback regarding study group experiences after the end of the semester. Some notable anecdotes are below:

- *Starting my freshman year of college virtually due to the pandemic made me feel really disconnected from the Berkeley community, but luckily being a part of a study group for EECS16A in my first semester really changed that. As a woman and underrepresented minority in STEM, my study group gave me the support system necessary to succeed in an extremely rigorous and minority-isolating major. This group was essential when it came to having people to complete assignments, learn concepts, and study for exams with, but it was also very helpful in me finding life-long friends. Through my study group I met my first and closest friends at Berkeley who would help me adjust to college in an online world and who later became my supportive and caring roommates in my sophomore year of college.*
- *As a new transfer student beginning remote learning at Berkeley, my 16A study group had a massive impact on how I carried about my studies. Not only did meeting up with other students remind me how important team-work is within the field of engineering, but it was also a constant reminder that it is OK to struggle on problems before arriving at solutions. Our group worked really well together because we had shared experiences. 5/6 of us are transfer students, so we understand each other's strengths and weaknesses when learning new material. Some of my few, closest friends at Berkeley are my EECS16A/16B study group peers.*

Independent of any grade or learning improvements, just making even a few students have better and happier college experiences through study groups is valuable. We hope that at scale this can have a massive impact, which may or may not be measurable through numerical data.

4 CONCLUSION & FUTURE WORK

The results presented in this paper stem from the Fall 2020 offering of EECS 16A at UC Berkeley. Since then, our research group has executed the matching process in numerous courses totalling over 3,000 students throughout the Spring 2021, Summer 2021, Fall 2021, and Spring 2022 semesters, and have had positive results based on preliminary looks at the data. A detailed analysis is yet to be

Student group	Indicated positive response to question / did not indicate positive response	MT1 Score (Max 50)	MT2 Score (Max 50)	Final Score (Max 100)	Misc. diff
All software-matched students	Did not request reassignment / did request reassignment 1-3 times	40.27 / 37.71 (p=0.003)	38.72 / 36.42 (p=0.025)	72.01 / 67.35 (p=0.046)	-
	Felt comfortable asking questions / did not	-	38.68 / 36.22 (p=0.022)	72.22 / 66.08 (p=0.011)	-
	Felt comfortable sharing ideas / did not	40.04 / 37.60 (p=0.011)	38.81 / 35.27 (p=0.002)	71.96 / 65.86 (p=0.046)	-
Freshmen	Did not request reassignment / did request reassignment 1-3 times	41.40 / 37.60 (p=0.0006)	40.44 / 37.17 (p=0.007)	74.56 / 68.55 (p=0.046)	-
	Felt comfortable sharing ideas / did not	40.96 / 37.41 (p=0.009)	-	-	-
Transfer	Would take future courses with group / would not	-	32.00 / 39.57 (p=0.001)	61.02 / 73.71 (p=0.012)	MT1-MT2: -0.075 / 0.012 (p=0.018)
	Any group interaction / no interactions with group	34.96 / 39.43 (p=0.013)	32.45 / 38.98 (p=0.005	61.65 / 72.81 (p=0.030)	
	Some to most members participated / no members participated	35.64 / 39.31 (p=0.024)	33.08 / 39.16 (p=0.009)	-	-
Students receiving a B-, B, B+ on MT1	Any group interaction / no interactions with group	-	-	72.71 / 65.78 (p=0.008)	MT1-Final: -0.075 / -0.14 (p=0.003)
	Some to most members participated / no members participated	-	-	-	-0.084 / -0.130 (p=0.047)
	Felt comfortable asking questions / did not	-	38.99 / 36.53 (p=0.044)	72.18 / 63.91 (p=0.004)	-0.082 / -0.154 (p=0.004)

Figure 8: Student's t-tests on difference in sample means, between exam grades within demographic groups, split on positive/negative responses to group experience indicators in the group survey. Only cells with significant differences in sample means are highlighted, with p-values provided. The Miscellaneous Differences column contains percentage-score changes for a student group, either from Midterm 1 to the Final, or from Midterm 2 to the Final, as indicated. Final scores were overall lower than midterm scores, so lower decreases in percent scores are interpreted as positive results.

conducted. Despite Fall 2021 and later semesters being in-person, we have found that study group formation continues to be in demand and well-used by courses. Student feedback has iteratively refined our process, and are experimenting with machine learning and reinforcement learning techniques to improve groups. We hope to soon release an open-source web-based privacy preserving version of our improved algorithm soon.

ACKNOWLEDGMENTS

Thank you to all UC Berkeley EECS 16A Fall 2020 students and staff for their incredible support and participation that made this project possible. Thanks to Armando Fox, Anant Sahai, Tesha Sengupta-Irving, Thomas Philip, Austin Patel, Drew Kaul, and Rahul Arya for helpful discussions and software assistance.

REFERENCES

[1] Juan Manuel Adán-Coello, Carlos Miguel Tobar, Eustáquio São José de Faria, Wiris Serafim de Menezes, and Ricardo Luís de Freitas. 2011. Forming groups for collaborative learning of introductory computer programming based on students' programming skills and learning styles. *International Journal of Information and Communication Technology Education (IJICTE)* 7, 4 (2011), 34–46.
[2] Douglas Allen-Taylor. 2013. Why Black Students Are Avoiding UC Berkeley. https://eastbayexpress.com/why-black-students-are-avoiding-uc-berkeley-1/
[3] CT Amelink, K Davis, BG Ryder, and MO Ellis. 2018. Exploring Factors Influencing the Continued Interest in a Computer Science Major. In *2018 ASEE Annual Conference & Exposition*. ASEE Conferences, Salt Lake City, Utah.
[4] Alexander W Astin. 1997. *What matters in college?* JB, Los Angeles, CA.
[5] Behzad Beigpourian, Daniel M. Ferguson, Frederick C. Berry, Matthew W. Ohland, and Siqing Wei. 2019. Using CATME to Document and Improve the Effectiveness of Teamwork in Capstone Courses. In *2019 ASEE Annual Conference & Exposition*. ASEE Conferences, Tampa, Florida. https://peer.asee.org/33497.
[6] Carlos Biaou. 2020. Our Culture. https://docs.google.com/document/d/1ErFyG2TboAJ3kHyd5SMxA3M1zGm7g2Qf0EMfEXSkoOc/edit
[7] Simone Borges, Riichiro Mizoguchi, Ig Ibert Bittencourt, and Seiji Isotani. 2017. Group formation in CSCL: A review of the state of the art. In *Researcher Links Workshop: Higher Education for All*. Springer, Springer, New York, NY, 71–88.
[8] Andreas Born, Eva Ranehill, and Anna Sandberg. 2020. Gender and willingness to lead: Does the gender composition of teams matter? *Review of Economics and Statistics* (2020), 1–46.
[9] David Boud, Ruth Cohen, and Jane Sampson. 2014. *Peer learning in higher education: Learning from and with each other.* Routledge, Oxfordshire, England.
[10] Nicole A Buzzetto-More, Ojiabo Ukoha, and Narendra Rustagi. 2010. Unlocking the barriers to women and minorities in computer science and information systems studies: Results from a multi-methodolical study conducted at two minority serving institutions. *Journal of Information Technology Education: Research* 9, 1 (2010), 115–131.
[11] Antoni Calvó-Armengol. 2004. Job contact networks. *Journal of economic Theory* 115, 1 (2004), 191–206.
[12] Antoni Calvo-Armengol and Matthew O Jackson. 2004. The effects of social networks on employment and inequality. *American economic review* 94, 3 (2004), 426–454.
[13] Jingnan Chen and Daniel Houser. 2019. When are women willing to lead? The effect of team gender composition and gendered tasks. *The Leadership Quarterly* 30, 6 (2019), 101340.
[14] J Clarke. 1994. Pieces of the puzzle: The jigsaw method,‖ Handbook of cooperative learning methods.

[15] Wilmax Marreiro Cruz and Seiji Isotani. 2014. Group formation algorithms in collaborative learning contexts: A systematic mapping of the literature. In *CYTED-RITOS International Workshop on Groupware*. Springer, Springer, New York, NY, 199–214.

[16] Katherine Deibel. 2005. Team formation methods for increasing interaction during in-class group work. In *Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education*. Association for Computing Machinery, New York, NY, USA, 291–295.

[17] Dora Dzvonyar, Lukas Alperowitz, Dominic Henze, and Bernd Bruegge. 2018. Team composition in software engineering project courses. In *2018 IEEE/ACM International Workshop on Software Engineering Education for Millennials (SEEM)*. IEEE, IEEE, New York, NY, 16–23.

[18] Dora Dzvonyar, Dominic Henze, Lukas Alperowitz, and Bernd Bruegge. 2018. Algorithmically supported team composition for software engineering project courses. In *2018 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, IEEE, New York, NY, 1753–1760.

[19] James E Eckles and Eric G Stradley. 2012. A social network analysis of student retention using archival data. *Social Psychology of Education* 15, 2 (2012), 165–180.

[20] Per Engzell, Arun Frey, and Mark D. Verhagen. 2021. Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences* 118, 17 (2021). https://doi.org/10.1073/pnas.2022376118 arXiv:https://www.pnas.org/content/118/17/e2022376118.full.pdf

[21] Daniel T Flynn. 2016. STEM field persistence: The impact of engagement on postsecondary STEM persistence for underrepresented minority students. *Journal of Educational Issues* 2, 1 (2016), 185–214.

[22] Joshua L Hertz. 2022. GRUEPR, a software tool for optimally partitioning students onto teams. https://coed.asee.org/2021/07/31/gruepr-a-software-tool-for-optimally-partitioning-students-onto-teams/

[23] Lu Hong and Scott E Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101, 46 (2004), 16385–16389.

[24] An Ju, Xiao Fu, Joshua Zeitsoff, Adnan Hemani, Yannis Dimitriadis, and Armando Fox. 2018. Scalable Team-Based Software Engineering Education via Automated Systems. In *2018 Learning With MOOCS (LWMOOCS)*. IEEE, IEEE, New York, NY, 144–146.

[25] Rosabeth Moss Kanter. 1987. Men and women of the corporation revisited. *Management Review* 76, 3 (1987), 14.

[26] Sumer Kohli, Neelesh Ramachandran, Ana M Tudor, Olivia Hsu, Gloria Tumushabe, and Gireeja Ranade. 2020. 16A Group Matching Form Fall 2020. https://drive.google.com/file/d/1p9xD_WzxQJIef2roTpNJctQzfY7-LU3K/view?usp=sharing.

[27] Sumer Kohli, Neelesh Ramachandran, Ana M Tudor, Olivia Hsu, Gloria Tumushabe, and Gireeja Ranade. 2020. 16A Study Group Final Survey Evaluations Fall 2020. https://drive.google.com/file/d/11volHEjTpMOvaxl_XYvvhqabsPg3BVyR/view?usp=sharing.

[28] Sumer Kohli, Neelesh Ramachandran, Ana M Tudor, Olivia Hsu, Gloria Tumushabe, and Gireeja Ranade. 2020. 16A Study Group Guidelines Fall 2020. https://drive.google.com/file/d/1HkuWLCkNT2nwQMeCpMcyXhHqu9ud6F6g/view?usp=sharing.

[29] Sumer Kohli, Neelesh Ramachandran, Ana M Tudor, Olivia Hsu, Gloria Tumushabe, and Gireeja Ranade. 2022. Scalable, Inclusive Matched Study Groups. https://github.com/sim-sg/group-matcher.

[30] Sumer Kohli, Neelesh Ramachandran, Ana M Tudor, Gloria Tumushabe, Olivia Hsu, and Gireeja Ranade. 2020. 16A Demographics Form Fall 2020. https://drive.google.com/file/d/1SVTEnRjCNQMNJ2BYTCG0mZVrGwBWYXl6/view?usp=sharing.

[31] Dragan Lambić, Bojana Lazović, Aleksandar Djenić, and Miroslav Marić. 2018. A novel metaheuristic approach for collaborative learning group formation. *Journal of Computer Assisted Learning* 34, 6 (2018), 907–916.

[32] Richard A Layton, Misty L Loughry, Matthew W Ohland, and George D Ricco. 2010. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education* 2, 1 (2010), n1.

[33] Henrik Hillestad Løvold, Yngve Lindsjørn, and Viktoria Stray. 2020. Forming and assessing student teams in Software Engineering courses. In *International Conference on Agile Software Development*. Springer, Springer, New York, NY, 298–306.

[34] Naseebah Maqtary, Abdulqader Mohsen, and Kamal Bechkoum. 2019. Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning* 24, 2 (2019), 169–190.

[35] Shweta Mishra. 2020. Social networks, social capital, social support and academic success in higher education: A systematic review with a special focus on 'underrepresented'students. *Educational Research Review* 29 (2020), 100307.

[36] Kevin W Mossholder, Randall P Settoon, and Stephanie C Henagan. 2005. A relational perspective on turnover: Examining structural, attitudinal, and behavioral predictors. *Academy of Management Journal* 48, 4 (2005), 607–618.

[37] Donald P Moynihan and Sanjay K Pandey. 2008. The ties that bind: Social networks, person-organization value fit, and turnover intention. *Journal of public administration research and theory* 18, 2 (2008), 205–227.

[38] Amir Mujkanovic and Andreas Bollin. 2016. Improving learning outcomes through systematic group reformation-the role of skills and personality in software engineering education. In *2016 IEEE/ACM Cooperative and Human Aspects of Software Engineering (CHASE)*. IEEE, IEEE, New York, NY, 97–103.

[39] Amir Mujkanovic, David Lowe, Keith Willey, and Christian Guetl. 2012. Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories. In *International Conference on Information Society (i-Society 2012)*. IEEE, Springer, New York, NY, 50–57.

[40] Yolanda Flores Niemann and John F Dovidio. 1998. Relationship of solo status, academic rank, and perceived distinctiveness to job satisfaction of racial/ethnic minorities. *Journal of Applied Psychology* 83, 1 (1998), 55.

[41] Chinasa Odo, Judith Masthoff, and Nigel Beacham. 2019. Group formation for collaborative learning. In *International Conference on Artificial Intelligence in Education*. Springer, Springer, New York, NY, 206–212.

[42] Robert T Palmer, Dina C Maramba, and T Elon Dancy. 2011. A qualitative investigation of factors promoting the retention and persistence of students of color in STEM. *The Journal of Negro Education* 80, 4 (2011), 491–504.

[43] Rita Manco Powell. 2008. Improving the persistence of first-year undergraduate women in computer science. *ACM SIGCSE Bulletin* 40, 1 (2008), 518–522.

[44] Budi Laksono Putro, Yusep Rosmansyah, et al. 2018. Group Formation in Smart Learning Environment: A Literature Review. In *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, IEEE, New York, NY, 381–385.

[45] Isabel J. Raabe. 2019. Social Exclusion and School Achievement: Children of Immigrants and Children of Natives in Three European Countries. *Child Indicators Research* 12, 3 (01 Jun 2019), 1003–1022. https://doi.org/10.1007/s12187-018-9565-0

[46] Yolanda A. Rankin and Jakita O. Thomas. 2020. *The Intersectional Experiences of Black Women in Computing*. Association for Computing Machinery, New York, NY, USA, 199–205. https://doi.org/10.1145/3328778.3366873

[47] Yolanda A. Rankin, Jakita Owensby Thomas, and Sheena Lewis Erete. 2021. *Real Talk: Saturated Sites of Violence in CS Education*. Association for Computing Machinery, New York, NY, USA, 802–808.

[48] Amber Solomon, Dekita Moon, Armisha L. Roberts, and Juan E. Gilbert. 2018. Not Just Black and Not Just a Woman: Black Women Belonging in Computing, In 2018 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT). *2018 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*, 1–5.

[49] Eve Spangler, Marsha A Gordon, and Ronald M Pipkin. 1978. Token women: An empirical test of Kanter's hypothesis. *Amer. J. Sociology* 84, 1 (1978), 160–170.

[50] Leonard Springer, Mary Elizabeth Stanne, and Samuel S Donovan. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research* 69, 1 (1999), 21–51.

[51] Ivan Srba and Maria Bielikova. 2014. Dynamic group formation as an approach to collaborative learning support. *IEEE transactions on learning technologies* 8, 2 (2014), 173–186.

[52] Thomas Staubitz and Christoph Meinel. 2019. Graded team assignments in MOOCs: Effects of team composition and further factors on team dropout rates and performance. In *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*. Association for Computing Machinery, New York, NY, USA, 1–10.

[53] Paris S Strom, Kelli L Hendon, Robert D Strom, and Chih-Hsuan Wang. 2019. How Peers Support and Inhibit Learning in the Classroom: Assessment of High School Students in Collaborative Groups. *School Community Journal* 29, 2 (2019), 183–202.

[54] Sachiko Takeda and Fabian Homberg. 2014. The effects of gender on group work process and achievement: an analysis through self- and peer-assessment. *British Educational Research Journal* 40, 2 (2014), 373–396. https://doi.org/10.1002/berj.3088 arXiv:https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/berj.3088

[55] Scott L Thomas. 2000. Ties that bind: A social network approach to understanding student integration and persistence. *The journal of higher education* 71, 5 (2000), 591–615.

[56] Leisa Thompson. 2018. A Systemic Approach to Recruiting and Retaining Women in Undergraduate Computing.

[57] Mischa Thompson and Denise Sekaquaptewa. 2002. When being different is detrimental: Solo status and the performance of women and racial minorities. *Analyses of Social Issues and Public Policy* 2, 1 (2002), 183–203.

[58] Kate M Turetsky, Valerie Purdie-Greenaway, Jonathan E Cook, James P Curley, and Geoffrey L Cohen. 2020. A psychological intervention strengthens students' peer social networks and promotes persistence in STEM. *Science advances* 6, 45 (2020), eaba9221.

[59] Roli Varma. 2006. Making computer science minority-friendly. *Commun. ACM* 49, 2 (2006), 129–134.

[60] Julita Vassileva. 2008. Toward social learning environments. *IEEE transactions on learning technologies* 1, 4 (2008), 199–214.

[61] Warren E Watson, Lynn Johnson, and George D Zgourides. 2002. The influence of ethnic diversity on leadership, group process, and performance: An examination of learning teams. *International Journal of Intercultural Relations* 26, 1 (2002), 1–16.

[62] Jess Whitley, Miriam H. Beauchamp, and Curtis Brown. 2021. The impact of COVID-19 on the learning and achievement of vulnerable Canadian children and youth. *FACETS* 6 (2021), 1693–1713. https://doi.org/10.1139/facets-2021-0096 arXiv:https://doi.org/10.1139/facets-2021-0096

[63] Craig E Wills, David Finkel, Michael A Gennert, and Matthew O Ward. 1994. Peer learning in an introductory computer science course. In *Proceedings of the twenty-fifth SIGCSE symposium on Computer science education* (Phoenix, Arizona, USA), Vol. 26. Association for Computing Machinery, New York, NY, USA, 309–313.

[64] Tai-ming Wut and Jing Xu. 2021. Person-to-person interactions in online classroom settings under the impact of COVID-19: a social presence theory perspective. *Asia Pacific Education Review* (2021), 1–13.

[65] Justyna P Zwolak, Remy Dou, Eric A Williams, and Eric Brewe. 2017. Students' network integration as a predictor of persistence in introductory physics courses. *Physical Review Physics Education Research* 13, 1 (2017), 010113.