

A Learner-Centered Technique for Collectively Configuring Inputs for an Algorithmic Team Formation Tool

Emily M. Hastings, Sneha R. Krishna Kumaran, Karrie Karahalios, Brian P. Bailey*

[ehstngs2,srkrish2,kkarahal,bpbailey]@illinois.edu

University of Illinois at Urbana-Champaign

Urbana, Illinois, USA

ABSTRACT

The configuration that an instructor enters into an algorithmic team formation tool determines how students are grouped into teams, impacting their learning experiences. One way to decide the configuration is to solicit input from the students. Prior work has investigated the criteria students prefer for team formation, but has not studied how students prioritize the criteria or to what degree students agree with each other. This paper describes a workflow for gathering student preferences for how to weight the criteria entered into a team formation tool, and presents the results of a study in which the workflow was implemented in four semesters of the same project-based design course. In the most recent semester, the workflow was supplemented with an online peer discussion to learn about students' rationale for their selections. Our results show that students want to be grouped with other students who share the same course commitment and compatible schedules the most. Students prioritize demographic attributes next, and then task skills such as programming needed for the project work. We found these outcomes to be consistent in each instance of the course. Instructors can use our results to guide team formation in their own project-based design courses and replicate our workflow to gather student preferences for team formation in any course.

CCS CONCEPTS

- Human-centered computing → Human computer interaction (HCI); Collaborative and social computing;
- Applied computing → Collaborative learning.

KEYWORDS

Team formation, CATME, Algorithm, Learning, Team Composition

ACM Reference Format:

Emily M. Hastings, Sneha R. Krishna Kumaran, Karrie Karahalios, Brian P. Bailey. 2022. A Learner-Centered Technique for Collectively Configuring Inputs for an Algorithmic Team Formation Tool. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2022), March 3–5, 2022, Providence, RI, USA*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3478431.3499331>

*The first two authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGCSE 2022, March 3–5, 2022, Providence, RI, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9070-5/22/03...\$15.00

<https://doi.org/10.1145/3478431.3499331>

1 INTRODUCTION

Teamwork is a foundational skill and is becoming an integral part of computing courses at all levels (e.g., [7, 9, 24, 34, 35, 38]). With the growing necessity of programming skills in many disciplines, instructors of computing courses must now determine how to best form teams from an increasingly large and diverse pool of students.

Instructors can utilize criteria-based team formation tools such as CATME [25] to keep pace with growing enrollments. When using a tool, however, an instructor must decide how to prioritize the large collection of criteria available in the tool for team formation. The instructor could consult the literature to aid this decision and make selections in the tool to diversify groups with respect to task skills [19], gender [4, 44], or personality [28]. However, these choices may not align with the preferences of those with the most at stake in the process – *the students* – and the number of criteria available in a tool is more expansive than what is covered in the literature.

Another technique, and the one we contribute to in this paper, is for the instructor to gather student preferences for the composition of a team, as prior work shows that increasing student agency and ownership can improve learning outcomes [8, 10, 12, 18, 31]. An open question is which criteria do students prioritize for team formation, to what degree do students agree with each other and the literature, and how do students perceive taking ownership over the team formation process. To address this gap, we build on theories of crowdsourcing to gather and aggregate student opinions for the magnitudes (priorities) and directions (grouping by dissimilar vs. similar attributes) for criteria available in a team formation tool.

In our study, we implemented a learner-centered workflow and deployed it in four instances of the same project-based user interface design course taught in a computer science department. Students selected a weight for each of 13 criteria nominated by the instructor that might be valuable for team formation in the course. These preferences were gathered through an online survey and aggregated to form a final configuration of the weights that was entered into the tool. In the most recent semester, students additionally participated in an online peer discussion about their preferred weights and could change their selections during the discussion period.

We present the criteria weights that were chosen by students in each semester and analyze where students agree and disagree on the criteria weights, how students prioritize the criteria, and what students reported learning from the team formation process. Of the 13 criteria nominated by the instructor for this course, students wanted to be grouped with other students who shared similar levels of course commitment and had compatible schedules for the project work the most; by dissimilar demographic attributes next, and by complementary task skills third. We also found that students disagreed most on the weights for criteria such as GPA and Work

	Criterion	Avg	SP19 (N=32)	FA19 (N=108)	FA20 (N=106)	SP21 (N=65)
Logistic	Schedule	3.8 (3.0)	4.1 (1.5)	3.8 (2.2)	3.9 (2.4)	3.8 (3.1)
	Commitment	3.1 (4.0)	3.3 (4.4)	2.9 (3.4)	3.4 (2.8)	2.7 (2.1)
	Leadership Pref.	2.1 (6.2)	0.8 (4.8)	1.9 (5.7)	2.3 (5.0)	2.7 (2.5)
	Work Style	0 (10.7)	0.8 (8.6)	-0.2 (7.6)	0.3 (8.8)	-0.6 (8.2)
	Leadership Role	-1.5 (6.4)	-0.4 (4.0)	-1.5 (5.6)	-1.7 (4.7)	-1.8 (4.4)
Academic	English Proficiency	0.8 (8.5)	0.2 (8.9)	0.6 (7.1)	1.3 (6.4)	0.9 (4.9)
	GPA	0.9 (6.5)	2.0 (5.4)	0.4 (6.2)	1.1 (5.0)	0.9 (2.2)
	Writing Experience	-0.8 (6.4)	-0.5 (8.6)	-1.2 (7.0)	0.6 (7.0)	-1.0 (4.4)
	Programming Exp.	-0.8 (8.2)	-	-1.1 (6.0)	-0.2 (6.8)	-1.4 (5.9)
	Design Exp.	-1.5 (6.0)	-	-1.7 (4.8)	-1.3 (4.8)	-1.6 (4.2)
	Area of Study	-1.2 (7.6)	-	-1.4 (7.2)	-0.8 (6.1)	-1.3 (3.5)
Demographic	Ethnicity/Race	-0.9 (6.5)	0.2 (3.7)	-1.4 (6.0)	-0.9 (4.8)	-0.5 (3.9)
	Gender	-1.3 (5.5)	-0.9 (3.1)	-2.0 (3.8)	-1.6 (4.1)	-0.1 (3.8)

Table 1: Mean (variance) of the student weights in four semesters of the user interface design course. The "Avg" column shows the overall mean and variance. See [1] for more information about the criteria. The scale is from -5 (group by dissimilar, shown in red) to 5 (group by similar, shown in blue), with 0 meaning to ignore the criterion. Due to a smaller size, SP19 had a slightly different course format (same instructor) and set of criteria.

Style and found that the online peer discussion helped to increase agreement for these criteria. We found the patterns in student preferences to be consistent across the four instances of the course in our study. The aggregated weights determined by the students in each semester of the course are summarized in Table 1.

Our results make several contributions to computer science education. First, we contribute a longitudinal dataset that instructors can use as a starting point for configuring a team formation tool in their own project-based design courses. Researchers can also use the dataset to motivate future experiments to test how the criteria preferred by students impact team effectiveness compared to the criteria reported in the literature. Second, we propose a workflow that allows students the opportunity to think and learn more deeply about team formation and reach a consensus with their classmates about the weights assigned to the criteria used by the algorithm. Instructors can replicate our workflow to gather student preferences for team formation in any course. Lastly, we offer implications for instructors wishing to implement a similar workflow for gathering student input for team formation in their courses. For example, we suggest which criteria instructors should ask students to discuss with each other and which criteria instructors could prioritize on their own. We also discuss a few situations where an instructor might intervene to revise the students' choices or the tool's output.

2 RELATED WORK

We situate our work with respect to the existing literature on team formation and crowdsourcing in learning environments.

2.1 Team Composition and Criteria-based Team Formation

Team composition has been shown in prior work to have important impacts on team outcomes. For example, teams with more women members have been found to have a higher collective intelligence than those with fewer women [4, 44]. Among other factors, skill diversity [19] and a balance of personality types [5, 28] have also been shown to increase team performance. For this reason, criteria-based approaches grounded in this literature form teams based on characteristics of potential members, including skills, demographics, and work habits and preferences. These teams have been shown to outperform those formed by other methods such as self-selection [6] and random assignment [44].

Algorithmic tools such as the Comprehensive Assessment for Team-Member Effectiveness (CATME) Team-Maker [25], DIANA [39], and groupformation.org [17], have become a popular way for instructors to automate criteria-based team formation as course enrollments grow and move online. These tools are grounded in the large body of literature on team composition. However, the existing literature often focuses on isolated experiments examining the effects of a single criterion (or small combinations of criteria) at a time, and often in organizations or simulated environments such as controlled in-lab studies or crowdsourcing platforms rather than authentic course settings [16, 23].

Some prior work has investigated incorporating preferences of potential team members regarding their peers, or aspects of their prior interactions, into the team formation process. For example, previous studies have explored forming teams based on pairwise transactivity in discussion and deliberation activities [41, 42], tie strength among members [33], and "speed dating" approaches

where potential teammates meet and rate each other before teams are formed [29, 30]. However, most of these techniques do not consider which criteria for team formation are most meaningful to students. Students have diverse opinions on what criteria should be included in the team formation algorithm [11, 15]. Even if students agree that a criterion is important enough to include, they may still be divided on how important it is relative to other criteria or whether similar or dissimilar students should be placed on the same team [15, 20, 22]. Due to this variety of differing opinions, when an instructor attempts to combine all student's voices (e.g., by averaging the criteria weights suggested by students), the result may not accurately reflect the configurations meaningful to students.

Our work reports a longitudinal dataset with the weights students prefer for 13 criteria for team formation, along with rationales. These results create research opportunities to study gaps between what the prior literature, instructors, and students prioritize. It also reports a workflow that instructors can implement to gather student input for prioritizing criteria for team formation and to nurture consensus amongst the students.

2.2 Crowdsourcing in Learning Environments

Crowdsourcing aggregates individual work and opinions based on the assumption that the combined localized knowledge of contributors rivals that of experts [37]. Prior work shows successful applications of crowdsourcing techniques in learning environments (learnersourcing), where students can participate in meaningful learning experiences for themselves and develop materials for other learners [21], including feedback on design projects [13, 14, 26, 43] and problem solving advice [32, 40].

Some research has investigated how learnersourcing can be used to select and prioritize criteria for team formation. This approach can enable students to have a stronger voice in the team formation process and help instructors to learn about what students prefer for team formation and why. For example, Connerley and Mael surveyed students to identify formation criteria that they deemed important and assess the perceived invasiveness of each [11]. Hastings et al. had students propose, discuss, and vote on criteria that would be used to form them into project teams, and then compared team outcomes and perceptions of the team formation process to those experienced by students formed using the traditional instructor-led process [15]. However, these studies provide limited insight into why students prioritize certain criteria over others and why students want to be grouped with other students who share similar or dissimilar values for these criteria (equating to positive or negative weights in a tool like CATME).

In this paper, we contribute a workflow to learn about student preferences and rationales regarding criteria weights for team formation. We also contribute a longitudinal dataset showing student preferences over four semesters of a course. Finally, we provide implications for instructors and designers of team formation tools regarding how to best elicit and incorporate student priorities into the algorithmic team formation process.

3 RESEARCH QUESTIONS

In this study, we aim to answer the following research questions:

- (1) What are the criteria and weights that are most valued by students for team formation, and why?
- (2) Which criteria do students agree on the most and least?
- (3) How does incorporating an online peer discussion impact student selections for the criteria weights?
- (4) What do students learn from the team formation discussion and what are their experiences with the process overall?

4 TEAM FORMATION TOOL

The course used the team formation tool CATME Team-Maker [2], a representative criteria-based tool used in many courses at our university. The tool can be configured to form teams using any subset of the 27 predefined criteria, as well as instructor-defined criteria. Prior to team formation, students complete an "About Me" survey in relation to the criteria selected by the instructor in the tool, including demographics and skill sets. The instructor then reviews the survey responses to designate a weight (-5 to 5) for each criterion indicating its relative importance (magnitude), as well as whether students with similar (positive number) or dissimilar (negative number) responses should be grouped together. For example, assigning a weight of 5 to the "Leadership Role" criterion strongly prefers teams where students have similar reported preferences for leading or following, whereas a weight of -3 would moderately prefer teams where students have different preferences for leadership role. A value of 0 tells the algorithm to ignore the criterion.

Once the instructor has configured the weights, the tool uses a greedy randomized algorithm to generate the teams [1]. The instructor can rerun the algorithm until they are satisfied with the teams, at which time the tool will notify students via email of the identities of their teammates and how to contact them.

5 METHOD

To answer our research questions, we developed a workflow that allowed students to collectively determine the weights for a subset of the criteria in CATME. The workflow consists of a student criteria weight survey and an online peer discussion. These two elements are not part of the CATME team formation tool. This study was performed in a project-based design course in the computer science department at our university and was approved by the IRB.

5.1 Team Formation Workflow

After completing the "About Me" survey in CATME, students had one week to complete an online survey (a Google Form) where they entered the weights they preferred for 13 criteria (see Table 1) the instructor nominated for team formation. The survey explained the scale for the weights by providing examples of how different weights influence the algorithm. This information was also provided in a lecture. Students were encouraged to think about what makes a good team for the course based on the lecture content, their own prior teamwork experience, and online research.

In the latest instance of the course, students followed the same workflow, but also participated in a week-long online peer discussion about the weights. Using a separate discussion thread for each criterion, students explained the rationale for their selections for three or more of the criteria that they cared most about. In each post, students included their selected weight and their reasoning

for the selection. During this time, they were allowed to revise their preferred weights and view a visual summary of the class' current selections. Students were instructed to be attentive to criteria with minimal peer agreement and those where their own selection differed from the selections of their peers. This workflow is based on the Delphi method for building consensus [3, 27, 36].

5.2 Participants and Course

Students in the course were mostly 3rd and 4th year undergraduate students in the CS department at the university. In the most recent offering of the course (SP21), there were 98 students, of which 70 consented to allow their data to be used in this study. Sixty-five (65) of these students completed all of the team formation activities. For the prior course offerings, data was collected for instructional purposes and is reported in aggregate here, with IRB approval. Students in all course offerings earned course credit for participating in the team formation activities.

In the course, students worked in small teams (3-6 students) to complete a 12-week user interface design project of their choice. The project consisted of nine project deliverables, including a project proposal, paper prototype, functional implementation, and user evaluation. One deliverable was due each week of the project. For example, one team developed a mobile application to facilitate planning social events for people who are remote or have recently changed social contexts (e.g., people who recently graduated, moved, or travel frequently). The project was worth approximately 50% of a student's final course grade.

5.3 Procedure

At the beginning of the semester, students had several days to complete the "About Me" survey in CATME. Students then participated in the weight selection survey (and online peer discussion in SP21). Once the weight selection period ended, the course staff determined the final weights by calculating the median of student responses. In some cases, the instructor made minor changes to reflect minority voices or eliminate perceived redundancy between two criteria with similar weights. For example, in SP21, he adjusted the weights of Gender, Race/Ethnicity, and Major by one unit. This final configuration was shared with the students, and the instructor explained the changes he had made to the student chosen configuration.

In the most recent offering of the course, students also submitted a Team Formation Experience Survey following the announcement of the configuration. In this survey, students responded to Likert items on a scale of 1 (Strongly Disagree) to 7 (Strongly Agree) assessing their satisfaction with the team formation process (e.g. "The final configuration...represented my own preferences"), how much agency they felt (e.g., "I was able to affect the final configuration of the criteria weights entered into the team formation tool."), and how much they valued this agency ("I would have preferred that the instructor determine the final configuration of the weights to enter into the team formation tool without gathering my opinion.").

Students additionally wrote open responses explaining what they learned from the team formation process and any changes they made to their preferred weights and why.

5.4 Measures

The study had the following measures: student-generated weights for selected criteria over 4 semesters, student rationales from the discussion in SP21, and perceptions of the team formation process and what students reported learning from the process in SP21.

The first two authors of the paper performed open-coding on the student discussions and the open-ended responses from the Team Formation Experience Survey to identify major themes.

6 RESULTS

In the next sections, we report the median of student responses in the survey as " $M=<\text{val}>$ ", the number of participants whose comments aligned with a theme as " $N=<\text{val}>$ ", and identify individual participants with " $P<\text{x}>$ ". Students reported strongly caring about how teams were formed in the course ($M=7$) and reported preferring to weight the criteria in the tool themselves rather than the instructor ($M=2$, inverted scale). Students perceived they affected the final criteria weights ($M=6$), the weights represented their preferences ($M=6$), and would result in good teams ($M=6$).

6.1 Patterns in Weights Over Time (RQ1&2)

A longitudinal dataset over four semesters of the user interface design course is summarized in Table 1.

Schedule and Commitment were consistently the two criteria that students wanted to prioritize the most for team formation (i.e., the highest magnitude for the weights). In both cases, students wanted to be grouped with other students who reported similar course commitment and availability (positive weight). Students explained this preference by noting that commitment was something that would likely not change during the course, whereas other skills could be gained over the semester.

The criteria where students preferred to be grouped by dissimilar responses included Programming Experience, Design Experience, and Leadership Role. Students felt programming and design were important skill sets and felt *"people who don't know how to code can at least have people with [the skill] in their team so they can finish the project"* (P1 discussing Programming Experience). For Leadership Role, students cited how the criterion would influence team dynamics: *"A group should contain 1-2 leaders at most, since otherwise there can be a lot of conflict within the group over project design decisions"* (P36 discussing Leadership Role). However, the magnitude of the weights students preferred for these criteria were not as high as for Schedule or Commitment, partly because students wanted the algorithm to prioritize these other criteria.

Students ultimately did not highly prioritize Gender and Ethnicity, although they did note the potential benefits of including them. For example, P19, who identified as a woman, commented, *"I do feel more comfortable having another woman to advocate with on a team since it makes it easier to speak up and be heard. A team where women are greatly outnumbered has, in my experience, been just as ineffective as a team with all men in terms of diversity value. I also put down a score of [+1, mapped to CATME scale], since I don't think it should be an overbearing factor when there are other important criteria such as skills, schedules, and commitment, but it should definitely be considered whenever possible."* Others, who identified as part of the majority with respect to these criteria, wished to abstain

because "*the minority should have the greatest impact*" and therefore voted neutral (ignore) hoping that "*the votes that matter can make a difference*" (P45 discussing Gender).

Agreement (variance) on each criteria was similar in each semester of the course studied. Work Style, English Proficiency, Programming Experience, Area of Study, and GPA had the most disagreement (i.e. largest variance). For these criteria, some students were concerned by the possibility of being the sole owner of a major task (Work Style and Programming Experience) or by how the criterion correlates with other attributes (GPA and Area of Study). Disagreement on these criteria may indicate that students can benefit from further discussion or instruction on how the configuration may influence the team. In support of this argument, the variances for GPA and Area of Study substantially decreased in SP21 when including a discussion, relative to prior offerings of the course. Even for criteria where the configuration was not impacted by the discussion, students did have the opportunity to listen to others about their perspectives and progress towards consensus.

6.2 Discussion and Voting Behavior (RQ3)

Students made a total of 264 posts in the online discussion describing their rationale for the weight they chose for each criterion (see Figure 1). 65 students submitted the voting survey a total of 104 times. 24 students changed their preferred configuration, sometimes more than once (a total of 39 times). See Figure 1 for a breakdown of voting changes by criteria.

The most frequently discussed criterion was Programming Experience (33 posts), but the logs showed that only eight students actually changed their vote. It is not clear why students did not change their vote, but it could be that the arguments were not compelling enough to change one's mind. This interpretation is supported by the fact that Programming Experience had one of the highest levels of disagreement among the criteria (variance=5.9). The least frequently discussed criterion was Work Style (12 posts). This criterion also had the fewest configuration changes and the most disagreement (variance=8.2).

In the Team Formation Experience Survey, 33 students reported that they had changed their vote and 26 reported not changing¹. 52 students provided an optional explanation of their reason for changing or not changing their vote. The first two authors performed open coding on these explanations and found that 29 students stated they gained new perspectives from the discussion, 10 engaged in personal reflection, and 2 performed independent research that supported their changed vote. A further 4 students explained that they did not change their votes because their original opinion was reinforced by the discussion.

6.3 Learning about Team Formation (RQ4)

Students reported in the Team Formation Experience Survey that participating in the decision process for the weights prompted them to think about what makes a good team ($M=6$, on a scale of 1 (Strongly Disagree) to 7 (Strongly Agree)). Students explained that they learned about new perspectives from the discussion ($N=24$, e.g. "*By discussing with other classmates, I see their opinions and realize something I didn't consider before.*"—P10) and reflected on

¹Note, the self-reported count of changes does not match the observed count.

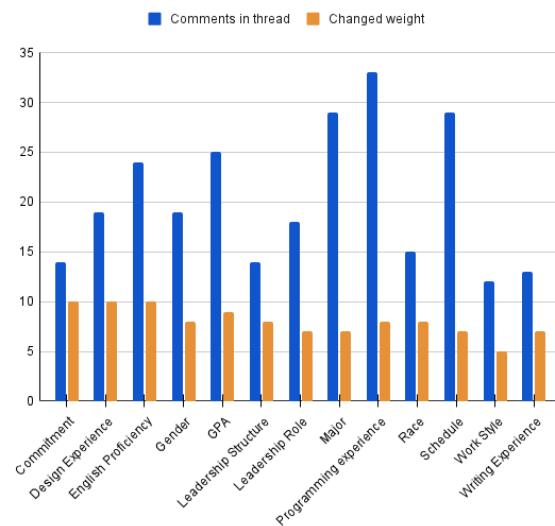


Figure 1: The number of comments students posted in the discussion thread for each criterion and the number of times a student changed the weight of a particular criterion.

their personal values for teamwork ($N=23$, e.g. "*I asked myself what is the most important thing I value when forming a team.*"—P8). They also thought about how to become a better team member ($N=15$, e.g. "*I realized that there were qualities...that I personally didn't think or care about that others thought were significant. This makes me think more about how I might be perceived as a team member.*"—P35). These results show that the discussion impacts student preferences and can help students think more deeply about team formation.

7 DISCUSSION

This paper implemented a technique for gathering student preferences for configuring the inputs to a grouping algorithm. The results extend prior work by reporting student preferences for both the magnitudes (priorities) and directions (group by dissimilar vs. similar attributes) for 13 criteria in a team formation tool [11, 15].

We find students consistently agreed on the weights for Schedule and Commitment over time. Students noted that these criteria are especially beneficial for improving team dynamics and the ability to get things done. The preference to form teams based on commitment level is interesting, because the effects of this criterion have not been studied in depth in prior literature, to the best of our knowledge, and may be difficult to emulate in a controlled study. Future field experiments are needed to test how Commitment influences team outcomes relative to other criteria.

Our results additionally show that some criteria studied in prior work such as Gender [4, 44] are prioritized lower by students (even some women) in favor of logistic criteria like Schedule. This may indicate that students are not aware of the literature or the criterion's impact on team dynamics and outcomes, or that the affected students believe that grouping by these criteria will not result in an improved team experience in this context. Instructors could provide

additional instruction on these topics or could select these weights themselves. However, in the latter case, instructors should clearly explain to students the rationale behind the decision.

Students disagreed most on the following criteria: Work Style, English Proficiency, Programming Experience, Area of Study, and GPA. For Programming Experience, there appeared to be a deeper underlying issue regarding concerns about work distribution on the team. For GPA, Area of Study, and English Proficiency, students had varying opinions on what the criterion represents or measures. Work Style was not a popular discussion thread for students (only 12 posts), so it is unclear why variance was so high for this criterion.

Our results demonstrate that including an online peer discussion improves student agreement in some cases. For example, discussing GPA and Area of Study reduced the variance in the weights for these criteria compared to semesters where there was no discussion (see Table 1). However, the discussion by itself may not be sufficient to fully resolve disagreements, and more scaffolding might be necessary. For example, for Work Style, there was still high variance in the weight for that criterion even after introducing the discussion. Instructors may wish to encourage students to focus their discussion on criteria where the variance decreased after introducing the discussion (e.g., GPA) or criteria where students continue to disagree (e.g., Work Style) to offer additional learning opportunities.

While students reported gaining new perspectives from their peers by reading other posts in the discussion threads, students did not change their preferred weights as frequently as we expected. In the Team Formation Experience Survey, 33 students reported changing their weights, but our data indicates only 24 students resubmitted the weights survey. This outcome is likely due to the fact that revising the weights in the survey was optional whereas a strict implementation of the Delphi method would require all students to revise the weights in stages [3]. Future work could study gathering student preferences with a staged workflow.

Instructors could also explore additional variations of the workflow when implementing it in their own classes. First, in our implementation of the discussion, students started with an empty discussion thread for each criterion. It might be useful to seed the discussion with comments or preferred weights from previous years or the literature to prompt responses from the current students. Next, students were required to provide a weight for all the criteria. This had the consequence that students had to provide a preference even if they didn't want their opinion to count for a specific criterion. For instance, for the criteria regarding demographic attributes, some students who identified with the majority wished to abstain so that those in the minority could influence the criteria weight the most. Students therefore elected to "ignore" the criteria (i.e. 0 weight) which affected the magnitude of the final weights rather than having the desired effect of amplifying the choices of the minority. Instructors can consider allowing students to abstain from choosing the weights for specific criteria. An instructor should also review the results for the criteria that will mostly affect the students typically under-represented in computing courses (e.g., the impact of Gender on women and the impact of Ethnicity / Race on students of color). In these cases, the instructor may need to revise the configuration to reflect the opinions of these students or discuss with these students directly which types of teams they might be most compatible with and then tweak the tool's output.

8 LIMITATIONS

This work has several limitations. First, two semesters in our dataset (FA20, SP21), the latter of which included the peer discussion, occurred during the pandemic and the course was taught using an online synchronous conferencing tool. While we did not observe a shift in the weights preferred by students, it is an open question how student opinions may change as we emerge from the pandemic. For example, students may now have a different perspective about the necessity of meeting with their team in-person. Additionally, this work was conducted in the context of a single project-based design course in CS. Further work is necessary to determine whether the dataset reported in this paper generalizes to different courses, disciplines, and modalities. Finally, we measured student preferences with the team formation process at the beginning of the semester. These preferences could change over time as students get to know their team and project. Future work can investigate how the preferences match with project outcomes or peer evaluations or by surveying students at different milestones in the course how they feel about their team and its composition.

9 CONCLUSION

When using a team formation tool, an instructor must decide the inputs for the grouping algorithm. This paper presents a workflow that allows the instructor to give students ownership over the tool's configuration by asking students to complete an online survey mimicking the same criteria weight choices the tool presents to the instructor and entering the aggregated weights in the tool. From implementing this workflow in four semesters of the same project-based course, we found that students made reasonable, consistent choices for weighting the criteria and reported preferring to weight the criteria in the tool themselves rather than the instructor. We also found that students generally wanted to be grouped with other students who shared similar course commitment and availability for project work the most; by demographic attributes next, and then by complementary task skills. Adding an online peer discussion will help students agree on the weights for some criteria and can provide an opportunity for students to think more about and learn how the composition of a team relates to its effectiveness. We hope the findings presented in this paper contribute to a future where students take ownership of the algorithmic tools used to group them into teams, are grouped into teams where they can best utilize their knowledge and skills, and have a successful learning experience.

ACKNOWLEDGMENTS

This work was partially funded by the Strategic Instructional Innovations Program (<http://ae3.engineering.illinois.edu/siip-grants/>) at the University of Illinois and by NSF awards CCF-1439957 and IIS-2016908. We thank the other course staff of CS 465 in Spring 2021—Gina Do and Simran Desai—and Prof. Emma Mercier, Prof. Darko Marinov, Wendy Shi, and Tiffany Li for their invaluable feedback.

REFERENCES

- [1] 2017. *Team-Maker Algorithm Detail*. Retrieved 4 April 2019 from <https://www.catme.org/faculty/help#TeamMakerScoring>
- [2] 2018. *CATME Smarter Teamwork*. Retrieved 31 August 2018 from <http://info.catme.org/>

[3] Maite Barrios, Georgina Guilera, Laura Nuño, and Juana Gómez-Benito. 2021. Consensus in the delphi method: What makes a decision change? *Technological Forecasting and Social Change* 163 (feb 2021), 120484. <https://doi.org/10.1016/j.techfore.2020.120484>

[4] Julia B Bear and Anita Williams Woolley. 2011. The role of gender in team collaboration and performance. *Interdisciplinary science reviews* 36, 2 (2011), 146–153.

[5] Suzanne T Bell. 2007. Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of Applied Psychology* 92, 3 (2007), 595–615.

[6] Lt Col James L Brickell, Lt Col David B Porter, Lt Col Michael F Reynolds, and Capt Richard D Cosgrove. 1994. Assigning students to groups for engineering design projects: A comparison of five methods. *Journal of Engineering Education* 83, 3 (1994), 259–262.

[7] Judy Brown and Gillian Dobbie. 1999. Supporting and evaluating team dynamics in group projects. In *The proceedings of the thirtieth SIGCSE technical symposium on Computer science education*. 281–285.

[8] Paula E Chan, Kristall J Graham-Day, Virginia A Ressa, Mary T Peters, and Moira Konrad. 2014. Beyond involvement: Promoting student ownership of learning in classrooms. *Intervention in School and Clinic* 50, 2 (2014), 105–113.

[9] Ben Coleman and Matthew Lang. 2012. Collaboration across the curriculum: a disciplined approach to developing team skills. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. 277–282.

[10] David T Conley and Elizabeth M French. 2014. Student ownership of learning as a key component of college readiness. *American Behavioral Scientist* 58, 8 (2014), 1018–1034.

[11] Mary L Connerley and Fred A Mael. 2001. The importance and invasiveness of student team selection criteria. *Journal of management education* 25, 5 (2001), 471–494.

[12] Yohanhan Eshel and Revital Kohavi. 2003. Perceived classroom control, self-regulated learning strategies, and academic achievement. *Educational psychology* 23, 3 (2003), 249–260.

[13] Elena L Glassman, Aaron Lin, Carrie J Cai, and Robert C Miller. 2016. Learnersourcing personalized hints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1626–1636.

[14] Elena L Glassman and Robert C Miller. 2016. Leveraging Learners for Teaching Programming and Hardware Design at Scale. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, 37–40.

[15] Emily M. Hastings, Albatool Alami, Andrew Kuznetsov, Christine Pisarczyk, Karris Karahalios, Darko Marinov, and Brian P. Bailey. 2020. LIFT: Integrating Stakeholder Voices into Algorithmic Team Formation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (Chi '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376797>

[16] Emily M. Hastings, Farnaz Jahanbakhsh, Karrie Karahalios, Darko Marinov, and Brian P. Bailey. 2018. Structure or Nurture? The Effects of Team-Building Activities and Team Composition on Team Outcomes. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2. ACM.

[17] Tyson R Henry. 2013. Creating effective student groups: an introduction to groupformation. org. In *Proceeding of the 44th ACM technical symposium on Computer science education*. ACM, 645–650.

[18] Susan Horwitz, Susan H Rodger, Maureen Biggers, David Binkley, C Kolin Frantz, Dawn Gundermann, Susanne Hambusch, Steven Huss-Lederman, Ethan Munson, Barbara Ryder, et al. 2009. Using peer-led team learning to increase participation and success of under-represented groups in introductory computer science. *ACM SIGCSE Bulletin* 41, 1 (2009), 163–167.

[19] Sujin K Horwitz and Irwin B Horwitz. 2007. The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of management* 33, 6 (2007), 987–1015.

[20] Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, and Brian Bailey. 2017. You want me to work with who? Stakeholder perceptions of automated team formation in project-based courses. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3201–3212.

[21] Juho Kim. 2015. *Learnersourcing: improving learning with collective learner activity*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[22] Michael S Kirkpatrick. 2017. Student perspectives of team-based learning in a CS course: Summary of qualitative findings. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 327–332.

[23] Cameron Klein, Deborah DiazGranados, Eduardo Salas, Huy Le, C Shawn Burke, Rebecca Lyons, and Gerald F Goodwin. 2009. Does team building work? *Small Group Research* 40, 2 (2009), 181–222.

[24] Celine Latulipe, N Bruce Long, and Carlos E Seminario. 2015. Structuring flipped classes with lightweight teams and gamification. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. 392–397.

[25] 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.

[26] Shang-Wen Daniel Li and Piotr Mitros. 2015. Learnersourced recommendations for remediation. In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on*. IEEE, 411–412.

[27] Weichen Liu, Sijia Xiao, Jacob T. Browne, Ming Yang, and Steven P. Dow. 2018. ConsensUs. *ACM Transactions on Social Computing* 1, 1 (feb 2018), 1–26. <https://doi.org/10.1145/3159649>

[28] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P Dow. 2016. Personality matters: Balancing for personality types leads to better outcomes for crowd teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 260–273.

[29] Ioanna Lykourentzou, Robert E Kraut, and Steven P Dow. 2017. Team Dating Leads to Better Online Ad Hoc Collaborations. In *CSCW*. 2330–2343.

[30] Ioanna Lykourentzou, Shannon Wang, Robert E Kraut, and Steven P Dow. 2016. Team dating: A self-organized team formation strategy for collaborative crowdsourcing. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1243–1249.

[31] Jeffrey A Mello. 1993. Improving individual member accountability in small work group settings. *Journal of Management Education* 17, 2 (1993), 253–259.

[32] Amy Pavel, Dan B Goldman, Björn Hartmann, and Maneesh Agrawala. 2016. VidCrit: video-based asynchronous video review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 517–528.

[33] Niloufar Salehi and Michael S Bernstein. 2018. Hive: Collective Design Through Network Rotation. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 151.

[34] Thomas J Scott, Lee H Tichenor, Ralph B Bisland Jr, and James H Cross. 1994. Team dynamics in student programming projects. *ACM SIGCSE Bulletin* 26, 1 (1994), 111–115.

[35] Debra Smarkusky, Richard Dempsey, J Ludka, and Frouke de Quillettes. 2005. Enhancing team knowledge: instruction vs. experience. In *Proceedings of the 36th SIGCSE technical symposium on Computer science education*. 460–464.

[36] Teri Spinelli. 1983. The Delphi Decision-Making Process. *The Journal of Psychology* 113, 1 (jan 1983), 73–80. <https://doi.org/10.1080/00223980.1983.9923559>

[37] James Surowiecki. 2005. *The wisdom of crowds*. Anchor.

[38] Anya Tafliovich, Andrew Petersen, and Jennifer Campbell. 2015. On the evaluation of student team software development projects. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. 494–499.

[39] Dai-Yi Wang, Sunny SJ Lin, and Chuen-Tsai Sun. 2007. DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior* 23, 4 (2007), 1997–2010.

[40] Helen Wauck, Yu-Chun Grace Yen, Wai-Tat Fu, Elizabeth Gerber, Steven P Dow, and Brian P Bailey. 2017. From in the Class or in the Wild?: Peers Provide Better Design Feedback Than External Crowds. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5580–5591.

[41] Miaomiao Wen, Keith Maki, Steven Dow, James D. Herbsleb, and Carolyn Rose. 2017. Supporting Virtual Team Formation through Community-Wide Deliberation. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 109 (Dec. 2017), 19 pages. <https://doi.org/10.1145/3134744>

[42] Miaomiao Wen, Keith Maki, Xu Wang, Steven Dow, James D Herbsleb, and Carolyn Penstein Rosé. 2016. Transactivity as a Predictor of Future Collaborative Knowledge Integration in Team-Based Learning in Online Courses.. In *EDM*. 533–538.

[43] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. 2016. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 379–388.

[44] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.