You Want Me to Work with *Who*? Stakeholder Perceptions of Automated Team Formation in Project-based Courses

Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, Brian Bailey
University of Illinois at Urbana
{fjahan4, wfu, kkarahal, marinov, bpbailey}@illinois.edu

ABSTRACT
Instructors are increasingly using algorithmic tools for team formation, yet little is known about how these tools are applied or how students and instructors perceive their use. We studied a representative team formation tool (CATME) in eight project-based courses. An instructor uses the tool to form teams by surveying students' working styles, skills, and demographics−then configuring the criteria as input into an algorithm that assigns teams. We surveyed students (N=277) in the courses to gauge their perceptions of the strengths and weaknesses of the tool and ideas for improving it. We also interviewed instructors (N=13) different from those who taught the eight courses to learn about their criteria selections and perceptions of the tool. Students valued the rational basis for forming teams but desired a stronger voice in criteria selection and explanations as to why they were assigned to a particular team. Instructors appreciated the efficiency of team formation but wanted to view exemplars of criteria used in similar courses. This work contributes recommendations for deploying team formation tools in educational settings and for better satisfying the goals of all stakeholders.

Author Keywords
Algorithms; CATME; Education; Team formation.

ACM Classification Keywords
H.5.3 [Information Interface and Presentation]: Group and Organization Interfaces--Collaborative computing.

INTRODUCTION
Teamwork can provide the multiple perspectives and diverse skills needed for solving complex problems [4, 18]. Industry rates the ability to work in teams as one of the most desirable soft skills in prospective employees [33]. The Accreditation Board for Engineering and Technology requires graduates to demonstrate "an ability to function effectively on teams to accomplish a common goal" [2]. Given its importance, more instructors are emphasizing teamwork in their courses.

CHI 2017, May 06-11, 2017, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/3025453.3026011

A challenge faced by instructors who want to incorporate teamwork into their courses is how to form student teams, especially in courses with diverse student makeup. To meet this challenge, instructors can now use tools that automate the team formation workflow [27]. An advantage is that these tools make it easier for instructors to compose student teams, and ideally "good" teams by applying findings from studies relating attributes of team composition to team performance and satisfaction [3, 6-8, 15, 17-21, 28, 30, 31, 34, 39].

However, despite knowing much about how teams *should* be formed, there is little knowledge of how instructors apply team formation tools in practice, or how students perceive their use. For example, what criteria do instructors apply and why? How do students and instructors perceive the tradeoffs of an automated approach for team formation? How could the approach be improved? Answering these questions is important because team assignments can affect students' project outcomes, learning, and course grades [35]. It is also important for instructors because by adopting these tools they are assuming the responsibility for team formation.

We report on a field study of a representative team formation tool (CATME [1]) deployed in eight project-based courses over two semesters. An instructor uses the tool to form teams by first surveying students' working styles, skills, and demographics, among other criteria; then configuring these criteria as input to an algorithm that optimizes the team assignments. We surveyed students (N=277) in the courses to gauge their perceptions of the strengths and weaknesses of the tool and ideas for improving it. In addition to collecting the criteria used in these courses, we interviewed instructors (N=13) outside of these eight courses to gain further insight into the criteria selections and experiences of using the tool.

We found that students appreciated the rational basis for grouping them, but they wanted a stronger voice in the criteria selection phase because the assignments would affect them and wanted explanations as to why they were assigned to a given team. Despite these suggestions, students reported being reasonably satisfied with the teams assigned by the tool (μ=4.0 on a 5-point scale). Instructors valued the increased efficiency of team formation and the ability to blame the tool in cases where students were dissatisfied with their team assignment. However, instructors also reported wanting to browse exemplars of successful criteria applied in similar courses and ways to explore how various criteria configurations in the tool would affect the team assignments.

The contributions of this work are: 1) recommendations for deploying team formation tools in educational settings and for better satisfying the goals of students and instructors; and 2) results creating better awareness of how students and instructors perceive the tradeoffs of this type of tool.

RELATED WORK

We situate our work in the context of prior studies of team formation tools, how the use of team formation tools can support effective team composition, and how a tool-based approach compares to other approaches for team formation.

Studies of Team Formation Tools

Despite the increased use of team formation tools, only a few studies of such tools have been performed in course settings. One study showed that the use of a team formation tool can optimize team assignments with respect to instructor-defined criteria better than the instructor [27]. Redmond deployed a team formation tool that emphasized schedule compatibility and reported that students perceived the team assignments as fair [32]. However, this finding was based on the instructor having received few complaints from students rather than an explicit measure of their perceptions. Tafliovich et al. used a team formation tool to assign students to teams and studied how students wanted their teams to be evaluated [38].

Our work shares the goal of studying a team formation tool in academic courses. However, our work is novel because we are studying perceptions of the team formation experience, including ratings of student satisfaction with the assignments and open-ended responses to interpret those ratings. We also analyze the criteria applied by instructors and how the automated team formation workflow can be improved.

Achieving Effective Team Composition

Team composition is known to affect team performance and satisfaction [35]. For example, research shows that teams with a balance of personality types perform better and have higher satisfaction than teams with a surplus of leader-type personalities [28]. Bear and Woolley found that the percent of women on a team correlates with higher collective intelligence [7] while Jehn et al. reported that gender balanced teams have higher satisfaction with their team experiences [24]. Team performance is also improved when the team has diverse skills relevant to their chosen project and a team size commensurate with the project scope [19].

The benefit of a criteria-based team formation tool is that an instructor can potentially use it to create teams by applying findings from the corpus of prior studies. The benefit is especially desired in courses where students have diverse demographics, prior performance, and skills. However, the number of possible configurations in a team formation tool leaves many gaps relative to the configurations that were tested in the prior studies. Also, the instructors using these tools may not be experts on the topic of team formation. It is therefore unclear how instructors would approach selecting the criteria in practice. Our study addresses this gap by examining the criteria selected by instructors in fourteen courses using one representative tool and by sharing how instructors feel these decisions can be better supported.

Approaches to Team Formation

Three common approaches to team formation in courses include self-selection, random, and criteria-based. Self-selection requires students to form their own teams. The key advantage of this approach is that the teams can experience increased initial cohesiveness [37], and the approach is often preferred by those who have potential teammates in mind [30]. The disadvantage is that students typically select teammates similar to themselves [22]. Homogenous teams typically lack the skill diversity needed for the work [6] and can be susceptible to groupthink [23]. This approach is also problematic in courses where students are less familiar with one another (e.g., inter-disciplinary courses) and are unable to join a team. The instructor must then develop a strategy for how to assign the remaining students to teams. Random assignment ensures that all students are placed on a team, but like self-selection, this approach is unlikely to provide teams with the necessary resources for performing the work [10].

To overcome the limitations of random assignment and self-selection, instructors are increasingly applying criteria-based approaches. The literature favors this decision. For example, empirical studies have shown that teams formed using a criteria-based approach outperform self-selected teams [8] and randomly assigned teams [41], assuming the criteria are grounded in team composition theory (e.g., see [19, 35]). However, teams formed with criteria-based approaches can still experience teamwork issues such as social loafing, lack of trust, and decision conflicts [9], but these issues should be no different than in teams formed with other approaches.

In these prior studies of criteria-based team formation, the criteria were selected and controlled for experimentation. In practice, less is known as to how instructors choose the criteria or why, what criteria they choose, and how students perceive an algorithmic intervention into the team formation workflow. Our study contributes to filling this gap.

THE TEAM FORMATION TOOL

We studied the Comprehensive Assessment for Team-Member Effectiveness (CATME) tool that automates the team formation workflow [1]. CATME is representative of the class of criteria-based team formation tools. We chose CATME because it is grounded in team formation theory [27], is an online tool and freely available to instructors upon request, and is used in courses at many universities.

To begin team formation, the instructor selects from a set of criteria available in the tool, based on what s/he believes is most appropriate for her/his course. The tool defines 27 criteria, including schedules, skills, academic performance, working styles, and demographics. The instructor can revise and extend the criteria and how the associated questions are phrased in the survey. Once finalized, the tool generates an online survey with questions relating to the selected criteria and distributes it by email to the students in the course.

After students respond to the survey, the instructor reviews the aggregated data, specifies the desired team size, and assigns a weight to each criterion in the range [-5, +5]. Selecting a negative weight for a criterion prioritizes groups with differing responses to the associated survey question, whereas a positive weight prioritizes groups with similar responses. For example, assigning -5 for a criterion related to skills (e.g., are you best at visual design, programming, or writing?) prefers skill diversity in groups. The magnitude of the weight for a criterion indicates its impact relative to the other criteria. A weight of 0 instructs the algorithm to ignore the criterion. The tool defaults some of the criteria to non-zero weights, but without explanation, and the instructors in our study almost always changed them.

The instructor can review the team assignments, adjust the criteria configuration, and re-run the algorithm. Given its use of a randomized greedy algorithm to satisfy the constraints, the tool may produce different team compositions for each run. Instructors have the option to accept the initial team assignments or generate alternatives and select the one that they believe to be most satisfactory. An instructor may also specify which students should (not) be placed on the same team in the tool. Once finalized, the system notifies students of the team assignments and shares their contact information.

RESEARCH QUESTIONS

Our study aspires to understand automated team formation from the perspectives of its key stakeholders: students and instructors. We focus on two sets of questions:

SQ1: Having been assigned to a team via an automated method, how satisfied are students with their assignments?

SQ2: What strengths and weaknesses do students perceive in an automated approach to team formation?

SQ3: What do students suggest for improving the approach?

IQ1: What strengths and weaknesses do instructors perceive of this approach to team formation?

IQ2: What are the implications of the instructor and student perceptions for improving automated team formation?

Answering these questions contributes to the base of knowledge for team composition from perspectives beyond team performance. It also highlights some of the unexpected consequences of algorithmic intervention in team formation.

METHOD

To answer these questions, we conducted a mixed methods study. It consisted of a survey with both structured and open-ended questions distributed to students and semi-structured interviews with instructors who had used the tool (CATME). The study was approved by our Institutional Review Board.

Courses Deploying CATME

The team formation tool was deployed for the first time in eight computer science courses at the University of Illinois. The deployment was part of an internal educational initiative to improve the consistency, efficiency, and effectiveness of team formation in engineering courses. The instructors of the eight courses agreed to deploy the tool, as each was grappling with the challenges of forming student teams in a large course. Prior to this deployment, the most common method of team formation in these courses was self-formation. The research team assisted with the tool in these courses, but the choice of the criteria was left to the instructors. The eight courses listed in the left-hand half of Table 1 display the instructor's selected criteria for that course.

Three courses (Software Engineering I, UI Design, and HCI) were offered in Fall 2015 and the remaining five (Software Engineering II, Social Visualization, The Art and Science of Web Programming, Mobile Design, and User Interface Design – a second time) were offered in Spring 2016.

The total enrollment in these courses was 869 students. The majority came from Computer Science and Electrical and Computer Engineering (80%). Most were upper-level undergraduates or first-year graduate students (96%). For the design-oriented courses, a fraction of the students came from non-engineering majors such as Psychology, Information Science, and Art and Design. All of the courses required the students to complete between one and five group projects during the semester. Some projects lasted a few weeks; others lasted the majority of the semester. Team size was course dependent, and varied from 3 to 8 students.

In all courses, there was a very high response rate to the team formation surveys, ranging from 83% to 97%, indicating that students were vested in the team formation process.

Student Survey

To gauge the students' perspectives on the use of a team formation tool, we developed an online survey. The survey contained two structured questions (5-point scale): 1) How satisfied were you with your team members? (1=Not satisfied, 5=Very satisfied) 2) Do you recommend using CATME for team formation in future courses? (1=Don't recommend, 5=Strongly recommend), followed by three open-ended questions: 3) If you prefer another method of team formation, please describe; 4) What do you see as the strengths of using CATME to form teams?; and 5) What do you see as the weaknesses of using CATME to form teams?

This survey was announced in the courses during final exam week or just prior to it. It was made clear that the students' participation in the survey would have no impact on their grades and that the course staff would not know who chose to participate until after the grades were finalized. There was no compensation for filling out the survey. The survey was completed online and the responses were anonymized. The students also filled out consent forms to allow us to use their data for research. A total of 277 students provided responses to the study survey and gave consent. Of those who reported gender on the survey, there were 35 females and 200 males.

Some respondents were enrolled in two or more of these courses. However, given the size of the dataset and the focus of our research questions, we believe this issue is inconsequential to the findings of the work.

Interviews with Instructors

We conducted semi-structured interviews with 13 instructors who were using CATME on their own initiative. Instructors involved in the initial deployment were not interviewed since our planning interactions may have biased their attitudes. Seven interviewees had used the tool in one course and the others had used it more than once. Interviewees were offered $10 for participation. Of these instructors, nine reported reading team formation literature prior to using CATME.

The interview focused on the instructors' motivations for using the team formation tool in their courses, the criteria they used, and the perceived strengths and weaknesses of the tool. The interview questions are shown in Table 2. Each interview lasted 30 minutes and was audio recorded. The recordings were transcribed and the originals destroyed.

Interviewees had deployed the tool in five lower- and seven upper-level undergraduate courses in engineering. Eight interviewees shared the criteria configurations used in their courses; listed in the rightmost six columns of Table 1. Two pairs of the interviewees had taught two courses jointly and had decided on the criteria together.

Qualitative Data Analysis

To analyze the free-form responses from the student surveys, we partitioned the responses into idea units. An idea unit is a coherent unit of thought [36]. The partitioning resulted in a total of 824 units from 258 students who had provided a response to at least one of the open-ended survey questions.

To develop the coding taxonomy, a member of the research team conducted a first pass over the idea units and assigned preliminary categories. Subsequent passes were performed to revise the categories such that they were reasonably exclusive and relevant to our research focus. Each category was given a label, definition, and example idea unit.

Using the taxonomy, the same member of the research team categorized all of the idea units. To test inter-rater reliability, another coder was trained on the categories. A sample of the idea units (about 11%) were labeled by the coder. Both the training and test samples were selected such that they covered all of the categories with the same distribution as the full dataset. Cohen's Kappa was 0.75 and exceeded the recommended threshold for accepting the results [26].

The interview data was similarly partitioned into idea units. We focused on units related to the instructors' motivation for adopting the tool, perceived advantages and disadvantages, and thoughts for improvement. This data set contained 114 idea units. Given the similarity of questions and responses, we were able to label the instructor idea units using the taxonomy derived from the student data with only minimal revision. Table 3 shows the full taxonomy. Each category includes the number of students and instructors whose responses had at least one idea unit referencing that category. Following the prior procedure, Cohen's Kappa for inter-rater reliability for labeling the instructor idea units was 0.82.

Criteria	Topics in HCI	Software Engr. I	Software Engr. II	UI Design (first)	UI Design (second)	Web Develop.	Social Viz.	Mobile Design	Project Mgmt (I1)	Thermal Behavr (I2)	Comp. Eng. I (I4, I5)	Hydraulics (I3)	Mech. Design II (I7)	Conservation (I10, I12)
Schedule		5	5	5	5		5	5	5		5		5	4
Big-Picture vs Detail-Oriented		-4		-5			-3	-2	-3	-2	-1		-2	-2
Gender	5	4		5			-1	5	-3	-5	5		2	
GPA	-4	-4		-5				-5	-3	-5	0		-2	4
Leadership Role		-4		-5			-4	-2	-3	0	-1			-2
Race		4		-3			0		-3	-5	5		-3	
Writing Skills				-1			-3	-3	-3		-1	-3		-3
Weekend Meetings		5	2				4				2	2		5
Leadership Preference		-4		-2			3	-2	-3	-2				
Commitment Level		-4		3		0	3		-3					2
English Skills		-3					-2	-2			-1			
Previous Course Grade										-5	-4	-1		5
Software Skills		-4							-3			-1		-3
Class Year				-3			3					-4		
Course skills (programming, graphic design, communication, etc.)				-5	-5		-2							
Possession of Mturk Account	-5													
Potential Roles in Web Dev						-5								
Coding Experience								-5						

Table 1. A sample of the most selected default and custom criteria in the courses in our deployment and taught by the instructors interviewed who shared their data. For each criterion, a negative weight groups students with differing responses; positive weights group by similar responses. A weight of 0 ignores the criterion. A criterion is left blank if the instructor did not collect the data on the CATME survey. Criteria available in the tool are shown above the heavy border; a sample of the custom criteria is below it.

RESULTS

Table 1 shows the criteria configurations for the courses involved in the deployment and the courses taught by the instructors interviewed who shared their data. The criteria were not uniform between courses. Our study revealed that many factors affect the selections, including the instructor's goals for the team work and for the course, prior experiences with team formation, and preference. For example, one novel use of the tool by an instructor was to form learning groups:

> *"Our goals were not only the projects but also there were learning teams. So learning teams meet weekly. They help each other with homework. That's basically the idea. They're like study groups."* (I4)

The table shows interesting patterns in the selections. First, the most frequently selected criterion and with the most common weight was *schedule*, indicating that instructors felt finding common free time was important for team formation. The next most frequent criterion was *big-picture / detailed-oriented*; instructors agreed that teams should have a mix of these working styles, as signified by the consistent use of a negative weight. A third pattern was related to *gender* as the weights spanned the full range of values. A positive weight favors teams with at least as many females as males, and is recommended for courses where women are the minority [31]. Reasons for the negative weights could range from uninformed choices to usability flaws; yet these differences highlight how the decisions could have a large impact on student experiences and need better support in the interface.

Differences in the criteria also reflect different perspectives on how teams should be created. For example, some instructors felt that teams should mix academic performance (where self-reported *GPA* serves as a proxy), as this would allow the weaker students to learn from the stronger ones:

> "*If you group students with similar academic ability… the people who may not be strong academically, don't have as much opportunity to be brought up and the people who are very strong academically never have to work with anyone who's not strong academically. If the goal is a learning experience which mimics what they will find when they get a job… they're gonna have to work with people of all different backgrounds and abilities.*" (I7, set GPA to '-2')

Other instructors disagreed, believing that teams with only weaker students need to elevate their project performance:

> "*…people with similar grades were mixed together,… The students [with lower GPAs who are grouped together] are used to maybe not working as hard. So they're still relying on someone else to pull the weight. But there's a transition that happens in every single one of those groups. At some point in the semester they each start becoming more self-reliant. They start pulling each other. That group almost always finishes at the very top in terms of their final project and in terms of their overall semester grades.*" (I10, set GPA to '4')

1. Can you characterize the course(s) in which you have used CATME?
2. Can you characterize the students in those courses?
3. What motivated you to use CATME for team formation in the course? What methods did you use prior to CATME and why did you switch?
4. Can you describe the criteria configuration you typically select?
5. What policies do you have for team formation? Can students switch teams once assigned? Can students give input for team members?
6. What are the advantages of using CATME for team formation?
7. What are the disadvantages of using CATME for team formation?
8. How could the use of CATME for team formation be improved?

Table 2. Questions for the instructor interviews.

SQ1: Student Perspectives

Students were mostly satisfied with the team assignments (μ=4.0, s=1.0) with 75.1% rating their satisfaction a 4 or 5. Only 9.4% of the students rated it unsatisfactory (rating of 1 or 2). The degree to which all students recommended the tool was lower (μ=3.58, s=1.52). Just over half (54.7%) would recommend or strongly recommend the approach in future courses. A smaller fraction (15.0%) would not recommend it, and 30.3% were unsure (rating of 3). Gender did not affect ratings of team satisfaction ($t(235)=1.00$, $p=0.32$) or recommendations ($t(232)=-0.68$, $p=0.50$). The patterns were generally consistent across the courses, despite the use of different criteria. However, students were less certain whether the positive team experience could be replicated:

> *"Maybe this time we got lucky that we had a great team, since I heard other team[s] did not do well."* (S427020);

> *"…CATME does not always work 100% of the time. Most of the time it seemed to match by schedule, but whether or not you will end up in a good team is still based on luck..."* (S427123).

Note that students are anonymously identified with the string "S" + course number + three-digit student identifier. The notation "S=n" for a category indicates that n number of students have cited that category in their responses. Each category was counted at most once per student so as not to skew the counts toward the response of any one student. Table 3 summarizes the categories of perceived strengths and weaknesses and ideas for improvement reported by students.

SQ2: Perceived Strengths by Students

We elaborate on some of the strengths based on how often they were cited and our interpretation of their importance.

Appreciate the use of rational criteria (S=177). The team formation tool organizes students into teams using specified criteria. Students valued the fact that there was rationale for the assignments, but only if the rationale matched their own interpretation of the "right" criteria: *"…CATME asked important questions when forming teams…"* (S427015). One implication is that instructors may want to show students how the criteria are set and how teams are formed.

Of the specific criteria cited in students' responses, the two most desired criteria were schedule compatibility (S=78) and diversity within teams (S=83). By grouping students who have similar schedules, teams should find it easier to meet: *"I believe matching the time availability is the biggest strength. This way, even with a big group, we can most likely meet at our desirable time"* (S427083). This issue was highlighted primarily by students in courses where teams consisted of at least five members. Students also believe their team could tackle difficult projects if comprised of diverse demographics, disciplinary skills, and leadership roles:

> *"Tries to match strengths and weaknesses of a group so that groups have all the resources they need"* (S467015).

> *"as our four team members come from three different discipline[s] (HCI, Software Engineering, Systems and Networking), in the brainstorming sessions, each of us attack the problem from our own domain. Thus the brainstorming sessions were full of interesting discussions"* (S565020)

Since prior work has shown that perceived diversity can reduce ratings of group satisfaction [34], the fact that many students reported diversity as a strength was surprising.

Learn to work with unfamiliar people (S=43). The tool's team formation algorithm considers criteria such as working styles and demographics, but not prior interaction or social relations between students. As a result, teams will likely be comprised of students who are unfamiliar with one another. The majority of students who referenced this issue saw this as a strength, though some perceived it as a weakness (S=20).

By being grouped with individuals they did not know, students reported having to learn team management and collaboration skills: *"This method also requires students to learn team management skills more similar to what is done in industry."* (S498rk014). They also view the situation as an opportunity to meet and befriend new people and become better prepared for industry jobs: *"Because people have to be used to working in different teams. It is very rare that they can choose their teammates."* (S427061).

Reduce stress and burden (S=37). With self-formation, the most common prior approach to team formation in the courses studied, the requirement to find teammates can be stressful. The team formation tool reduces this stress and simplifies the process. Three representative statements were:

> *"It's extremely beneficial for students from departments other than the dominants (CS, ECE) of this source who have few acquaintances."* (S565007)

> *"If you do not have friends in the class, CATME assures everyone has a group without feeling left out."* (S498bb009).

> *"No effort is required. You don't need to meet other people in the class or search for good partners, it is automatic."* (S498bb013).

SQ2: Perceived Weaknesses by Students

We elaborate on two of the common weaknesses cited by students. The full range of cited weaknesses are in Table 3.

Mismatch between student and instructor preferred criteria (S=108). A frequently cited weakness was when students' interpretation of what criteria should have been selected did not match the instructor selections: "*Some questions should not have been in the form [survey]"* (S465002). Fifty-five students wanted the instructor to consider personality traits, motivation, and personal work habits for team formation. As some of these criteria are already in the tool, the instructor could partially address this issue by adapting their selections.

Other issues would be difficult to capture in the tool because they only surface after teams begin working together: *"It does not talk about the characters of the teammates - if someone gets really frustrated when work is not done according to their style, or their deadline etc."* (S467002). One way to reduce such conflicts is to incorporate team building activities soon after teams are formed (e.g., [11]).

There was also disagreement around some criteria, such as schedules. While many students valued schedule matching when forming teams, others did not (S=15). These students reasoned that their schedules change as they add or drop courses or get involved in activities and this criterion should be given less weight: *"Scheduling is hardly a good metric of assigning teammates. Student schedules are far too volatile for them to stick with the given schedule"* (S467014).

Students sometimes agreed on the inclusion or exclusion of certain criteria, but disagreed on the weights. An example centers on language proficiency. Some students favored a mixture of proficient and not-so-proficient English speakers in a team. Their rationale is that the presence of a proficient speaker helps with essays and presentations: *"All of our teammates are non-native speakers…, so compared to others team with at least [one] native-speaker…, it is a little bit unfair"* (S565009). Others considered this to be a hindrance to effective communication within the group: *"If you are put into a group with random people, there could be a huge language barrier that prevents work from being done efficiently. For example, there are a lot of international students from China that don't speak English that well. I am not Chinese and don't understand it, so it would be difficult to communicate ideas with them..."* (S465001)

One way to resolve the issue of criteria mismatch is for the instructor to involve the students in choosing the criteria. For example, an approach taken by one of the instructors in our study was to show the tool live on the lecture screen and work through the criteria configuration with the entire class.

Lack of transparency (S=54). Students reported lack of transparency as another weakness of the tool. When notified of their assignment, teams do not receive any explanation as to why they were selected to work together. Without such explanations, for example, teams may fail to recognize the presence of a particular skill or role in their group know who

	Category	Definition	Example
Strengths	Based on rational criteria (S=177; I=3)	Criteria perceived to be important for forming teams.	*"I do like the idea of matching people with similar schedules especially for large teams."* (S427118)
	Reduce stress and burden (S=37; I=13)	Reduces anxiety / burden / discomfort for having to find teammates, or the convenience of its use.	"*It's extremely beneficial for students from departments other than the dominants (CS, ECE) of this course who have few acquaintances."* (S565007); *"And why did I switch?... Well, I didn't wanna do it by hand anymore. And I wanted to try an online system, [and to see] if it's easy, if it's gonna be good for the students and these kind of things. And maybe regarding using it in the future in a bigger class where doing it by hand would be a huge pain."* (I13)
	Learn to work with unfamiliar people (S=43; I=2)	Meeting new people as in a real world job, or developing authentic team skills.	*"it's very like the real-life scenarios, that we can't choose who to work with sometimes, and we have to deal with these situations very often."* (S565001); *"I think that our program is really designed to prepare the students for industry jobs and graduate schools where they have to work in teams so it's tough sometimes dealing with human aspects of engineering courses. So, I would like to systematize it. This is nice."* (I7)
	Level the playing field (S=12; I=5)	Giving a fair chance of being assigned to a good group.	*"The biggest advantage I see from using CATME is the fair environment it creates for a class. It prevents students who know each other and work well together from being able to form super groups."* (S427012); *"And it's nothing the students are gonna discuss with me a lot. Because they feel that since I outsourced it to something that is considered to be some programmed platform, there's less of a personal bias in it. I think that's an advantage. If I would do it on my own, they would probably say you did this and you did that, this can be done differently."* (I2)
	Free of relationship biases (S=6; I=0)	Teams do not consist of friends. Members cannot rely on friendship to refrain from doing work.	*"Working with friends can be awkward especially if one of them doesn't do their part."* (S498rk014)
	Positive comparison or view (S=64; I=1)	Strengths that did not fit in other categories.	"*CATME is great!"* (S428009)
Strengths / Weaknesses	Lack of transparency (S=54; I=3 as strength, I=6 as weakness)	Not knowing how the teams are formed, having little clue as to what the strengths and weaknesses of one's team members are.	*"I have no idea if the survey we filled is actually useful. I don't know what kind of algorithm the CATME is using to form a team so it's seems to be just random I guess."* (S427008); *"They [the students] always say "We don't have anybody that knows this topic in our group. I think we're at a disadvantage.". It's not true. They just don't know. They don't know that this person had this prerequisite. They just assume that we don't have anybody who knows this topic."* (I3, citing as a weakness); *"I like that the students think it's not us, we're not trying to manipulate them in any way although we are. "* (I6, citing it as a strength)
Weaknesses	Mismatch between student and instructor preferred criteria (S=108; I=0)	Students do not agree with the criteria selections made by the instructor.	*"In creating a schedule, CATME's schedule is not particularly useful due to changes in people's lives and events that may only happen during a given week that can throw meeting times awry."* (S427109)
	Burdensome to learn and use (S=29; I=8)	Problems with the tool's user interface or documentation); or the perceived burden of having to learn / use it.	*"The user interface is very unintuitive, causing multiple members of the team to incorrectly fill out the time they were available."* (S427121); *"The disadvantage is that it's another piece of time and the students are very busy, we're very busy. You'd definitely want to use all the tools available but you don't want to have it become a burden on the students."* (I10)
	No validation of responses (S=20; I=2)	Team assignments could be influenced by inaccurate responses since they cannot be verified. Susceptible to gaming the system.	*"Ranking your abilities on a spectrum is also very relative. One student's 'somewhat proficient' may be another student's 'very proficient'."* (S467015); *"[it] allows people to game system. Because when people realize they're in the same lab together,... they say I wanna be on the team with you so then we'll both select midnight to 4am on Wednesdays and lo and behold they're always together. Because nobody else is gonna select [that time]. So it's kind of dubious in that regard but that's ok."* (I11)
	Lack of consideration for team preferences (S=20; I=0)	Requests to be (not) grouped with others were not considered.	*"inability to add strong preferences for people that you would like to work with, although this may just be intentional, it would be nice if the course allowed for more control on how people choose to work with."*(S428013)
	Team chemistry and communication (S=13; I=0)	Team struggles to communicate, or lacks team chemistry.	*"It's also hard sometimes to communicate with new people whom you've never worked with before because you don't know the best way to get a hold of them is."* (S498bb019)
	Cold start (S=8; I=0)	Need to break the ice with the assigned team.	*"team synergy can be slow to build."* (S498bb003)
	Immeasurability of criteria (S=0; I=1)	There are criteria that surveys cannot measure.	*"you're not measuring certain features that are important like management and the ability to meet the deadlines."* (I3)
	Negative comparison or view (S=54; I=0)	Weakness that cannot fit in the other categories.	*"Only rating it low because of negative experience with it. Not sure if self-formed teams would work better, but I don't see any advantage of using this system."* (S427101)

Ideas for Improvement	Integrate students' team preferences (S=21; I=1)	Mixing teams created via the tool and self-assembled teams	*"Random assignment teams with self-formed suggestions would be nice, that way teams wouldn't be so disparate."* (S465016); *"…making sure they can do preferred and unpreferred"* (I9)
	Expand criteria and aid criteria decisions (S=4; I=10)	Ranged from having better defaults, expanded criteria, and other UI improvements.	*"It would be nice to add metrics for matching students based on how likely they are to attend class and keep up with the work."* (S465015); *"…getting more feedback from others that have used team formation about what would be good criteria and good weights to use, when they've been successful at using it.* (I9)
	Support iteration (S=2; I=0)	Iterating until a good team is formed	*"It would be interesting to see what the result would be if the steps are: filling survey in CATME -> assigning groups ->group meet and know each other -> confirm or decline assignment -> reassign groups."* (S565010)
	Add a help wanted forum (S=5; I=0)	Create place for students to exchange skills and strengths for recruitment.	*"Perhaps we could get people to shout out ideas and get people to join their idea and form a team instead. That's what CS 198 does."* (S427009)
	Other ideas (S=3; I=2)	Ideas not fitting in any of the other categories.	"*Instead of self rank, they should be an individual project at the start of the semester. The individual would then submit their work to highlight their strengths, weaknesses, and work ethic. Their peers will then decide who they want to work with and would create a balanced team."* (S467021)

Table 3. The full taxonomy used to categorize idea units from the student surveys and instructor interviews. It contains strengths, weaknesses, and ideas for improvement as top-level categories, and a set of lower-level categories within each of them. In the lower level categories, 'S=' and 'I=' are the numbers of students and instructors whose responses contained that category.

possesses it: *"I don't know any of my partners[‘] skills and backgrounds, so it's hard to assess what they can be capable of when dividing up tasks"* (S427127).

Unaware of how the tool's algorithm forms teams or why, some students formed their own hypotheses: *"[CATME is] discriminating (why does it need to know my ethnicity, GPA, etc.)"* (S465011); This phenomenon is similar to how users hypothesize how the invisible Facebook curation algorithm chooses content for their News Feed [13, 14].

To address the issue, an instructor could share the team members' survey responses or have the team work through a self-assessment of their skills and working styles. The tool could also be enhanced to report how the team scores on the various criteria and highlight their strengths as a team. Future research is needed to test how increasing transparency of the team formation process would affect team effectiveness.

SQ3: Ideas for Improvement from Students

We elaborate on two suggestions for improvement from students. All of the themes for improvement are in Table 3.

Integrate student preferences with the tool (S=21). One idea for improvement was to let students select part of their team*: "Might want to have an option where you can specify one teammate you want and if the other person reciprocates then there's a good chance they get teamed up?"* (S498rk010); *"The win-win situation would be to let students choose between [a] self-formed team or CATME. So those who knows [sic] whom to team up with can choose their own teammates. And the other students who do not know anyone (or do not want to form their own teams) can sign up to be matched through CATME."* (S465004).

Use help wanted forums (S=5). A few students felt that different projects required different skills and working styles. They suggested to use a forum in which students could post what attributes they are looking for in team members, or what they could bring to a team. This way, they can recruit team members, or be recruited by others*: "I tend to prefer picking teammates through something of a posting board. It allows individuals to match themselves based on what they value personally to find best fits for themselves"* (S498rk001).

IQ1 and IQ2: Instructor Perspectives

We describe strengths and weaknesses of the team formation tool from the perspective of instructors. In most cases, the instructors' opinions matched with students'. We elaborate on issues that we believe are insightful and unique relative to what has been discussed from the student perspectives.

Strengths. Instructors identified reducing their burden and student stress (I=13), leveling the playing field to give all teams an equal chance of success (I=5), criteria-based matching (I=3), students' needing to learn to work with unfamiliar people (I=2), and the lack of transparency of the algorithm (I=3) as strengths of the tool. All but the last were also cited by students. Table 3 describes all these strengths.

All of the instructors interviewed cited increased efficiency of the team formation process as a key strength of the tool (I=13). The tool allows for quickly revising the criteria and reviewing the resulting teams, which would not be feasible if the teams were assigned manually: *"The efficiency of distributing students with respect to various criteria is very attractive to me because all I've to do is click a button. I don't have to sit there and do it for hours and hours."* (I7). The instructors also viewed the use of the tool as an anxiety reducer for students: "*I often tell the students at the beginning of 201 when they're graduating from U of I, they're leaving with a professional degree. So their experience in that first year is similar to the experience of a first-year law student or first year medical students, which tends to be incredibly stressful. So I feel like this CATME approach, basically having teams, is a great way to reduce the stress and make all a little bit more manageable*" (I10).

Some instructors (I=3) felt the lack of transparency of the process was desirable for at least two reasons. First, they believed students should not be conscious of how criteria such as race or gender were used: *"We certainly wanted to do what we could based on some literature that we should for example group under-represented minorities for example women together and we just wanted to make sure that some of those things were artificially emplaced and not that the students sort of feel that we were targeting them in any way. So CATME provides a little bit of anonymity in that sense because we just tell them it's an algorithm and based on what you fill in it'll optimize your group preferences but we don't tell them what the algorithm behind the scenes did"* (I6).

Second, instructors felt the tool allows them to deflect blame to the software if some students are dissatisfied with their team*: "Theoretically you can blame the software if things don't go right. If the students aren't allowed to self-assemble and if the faculty members were going to assemble it just randomly, there's always some who are disgruntled"* (I9).

Weaknesses. Instructors also perceived weaknesses of the tool: it was burdensome to learn and use (I=8), there are no means to validate student responses to the survey (I=2), and certain criteria are not measurable (I=1). Some instructors also viewed the lack of transparency as a weakness, whereas others saw it as a strength. All except the immeasurability of criteria were also identified as weaknesses by the students.

Instructors (I=6) who considered lack of transparency as a weakness deemed it important for students to know why they are placed on a given team. They believed such knowledge would eliminate many doubts about one's team assignment and would spare instructors the need for explanation*: "One comment I was getting from some students is that they don't understand why they're put together with certain students in a team; because we don't wanna release the other students' information to them for privacy reasons. But CATME could generate an anonymized summary for me. So a lot of people keep asking "why am I with this person in a team"* (I3).

Instructors believed that effective team formation requires matching along certain criteria, but that some criteria have no means of proper self-measurement (I=1). One example is commitment, which is part of the built-in criteria in the tool and often selected by instructors (see Table 1): "*I don't know if you can measure that. I don't think anybody would ever say I have a low level of commitment. At the first deliverable people put in their hours. Second one other courses started to take their time and some people just stopped doing anything. So everybody can commit at a high level if they have nothing to do. When you have a lot of stuff to do, that's when the question of how committed you are comes in"* (I3).

Ideas for Improvement. A unique suggestion from the instructors was to include better guidelines for configuring the criteria (I=3). This would make the user interface more usable and the process more effective since instructors are not always well versed on the latest literature for team formation. In particular, they expressed wanting the tool to offer in-situ explanations for the available criteria*: "I know that there's been a lot of research on some of these approaches to team formation. It would be nice if as you're choosing which things you would use for team formation you could just kinda click and see why [it] is important to group things similarly or these things dissimilarly. That way you just have all the information available to you"* (I10).

DISCUSSION AND FUTURE WORK

In this study, we gauged perceived strengths and weaknesses of a team formation tool from the perspectives of students and instructors. For students, the strengths of the tool were: it is based on rational criteria, removes the stress of finding teammates, and promotes learning to work with new people. Instructors identified similar strengths, but also noted the efficiency of forming teams and the ability to deflect blame when students were dissatisfied with their team assignment.

For key weaknesses, students identified mismatches between their preferred criteria and the instructor's selections, not knowing why they were assigned to a team, the inability to specify preferred team members, and a cold start phase due to team members' typically not knowing one another prior to the course. Instructors also identified the inability to properly measure some of the criteria as a weakness.

From our results, we identified several recommendations for instructors to address these weaknesses and improve the deployment of a team formation tool. One recommendation is to engage students in selecting which criteria are selected as input to the tool's algorithm. As mentioned in our study, instructors could poll students live during lecture and use the results to make selections in the tool or focus student input on a narrower set of criteria while explaining the rationale for the criteria already selected. Instructors could also allow students to rank the criteria as part of the survey procedure performed at the onset of team formation. Instructors could use the rankings to assign the criteria weights in the tool. The rankings could be collected separately, or tool designers could implement this feature in the team formation survey.

Another recommendation is to provide each team with an explanation for its formation and why it is "good." This could take the form of an anonymized summary of the team members' responses to the survey, how the team scored on the criteria, or how the team' scores relate to other teams as a way to highlight their strengths and weaknesses. As stated by one instructor: *"If CATME could give us an anonymized report for the team that these are the strengths; these are the weaknesses of your team. That lets the teams think the system is fair to them" (I3).* Instructors could also create and give teams a rubric based on the selected criteria to self-assess their qualities. This could also serve as an activity to address the cold start phase experienced when teams first meet.

A final recommendation for instructors is to give students more agency in the selection of their teammates. To support preferences, instructors could allow students to form partial teams and use the tool to complete the teams. This approach echoes prior work advocating for constrained self-formation [5], while also achieving the benefits of a criteria-based approach (e.g., skill diversity in the teams). Instructors could also use the tool to generate multiple team assignments, have the teams meet and rank their preferences, and include this data as a criterion in the tool [29]. These approaches could yield higher team satisfaction, but knowing the effects on team performance would require future work.

An interesting aspect of the criteria configurations reported in Table 1 is that no two courses were the same. This can be explained in part by the courses having different learning goals and student makeup. Some instructors targeted the experience of learning; others the importance of project outcomes and grades. These differences highlight the need for team formation tools to better support instructors. One implication for tool designers is to provide more effective defaults and in-situ explanations in the user interface for criteria selection. A second implication is to offer a dynamic visualization that shows how modifying the criteria affects the assignments. Last, tool designers should create an online catalogue of exemplars of configurations used by instructors that could be browsed by course topic and student makeup.

Some criteria configurations could also be due to irrational or uninformed choices. Despite instructors having good intentions, the operationalizations of the intent may be flawed. For example, eight of the instructors discussed their rationale for the gender criteria weight, but also expressed uncertainty whether it was "correct." These sentiments echo the instructors' need for guidance within the interface. While some instructors were aware of the tool's algorithm, with the number of possible criteria, predicting the team assignments is a challenge. We see this challenge only growing, as this genre of tool will continue to consider more criteria (e.g., imagine scanning students' social media profiles to infer personality traits) and apply more sophisticated modeling. How does one assign weights to an expansive set of criteria and when the relations between the criteria are difficult to grasp? Similarly, some students were briefed on the rationale for the team formation tool during lecture. Might they have been swayed by the framing of systems like CATME, believing in the promise of the tool [25]?

We used Spearman's ρ to test how student ratings of team satisfaction (scale was 1-5) related to their project scores (normalized to 0-100). This was tested in five of our courses. The others were not included due to limitations of our data set. The correlations were small to modest, and inconsistent in direction. The values were positive for three courses ($\rho=0.21$, $p=0.45$; $\rho=0.38$, $p=0.09$; $\rho=0.19$, $p=0.02$), and negative for the others ($\rho=-0.57$, $p<0.01$; $\rho=-0.11$, $p=0.60$). We interpret this pattern to mean that one's satisfaction with his or her team assignment is only loosely related to the project grade. Additional work is needed to better understand what factors influence students' ratings of team assignments.

We see several additional directions for future work. First, team composition is important, but it is only one factor affecting team outcomes. Future work should compare how team composition affects performance relative to enhancing team dynamics such as psychological safety [12, 40]. The outcome would help instructors know whether to allocate more attention to team formation or to creating activities that strengthen team member relations. Second, future work could study the inclusion of additional criteria such as social intelligence [16] as inputs to the matching algorithm, which would enable balancing teams along additional axes. Third, this work examined the use of a team formation tool in the context of courses. Future work should study such tools in other teamwork settings such as in hackathons and design competitions. Finally, future work is needed to test the generalizability of our findings in courses with different student makeup, criteria selections and matching algorithms.

LIMITATIONS

One limitation is that the student perceptions reported in this study were largely based on students enrolled in computer science courses. Since these students are knowledgeable of algorithms, their perspectives may not be representative of students in other disciplines. A second limitation of our study is that student perceptions of the team formation process could be affected by how instructors configured the tool. Given the scale of all possible combinations of the criteria available in the tool, it was not possible to analyze this relationship. Third, our study was conducted in courses taught in a single academic unit at the same institution, and may not be representative of different teaching cultures.

CONCLUSION

Instructors are increasingly leveraging algorithmic tools to organize students into teams in educational settings. We reported on student and instructor experiences with team formation as defined by one tool (CATME) that is widely used in university courses. Students reported being satisfied ($\mu=4.0$ on 5-point scale) with the teams assigned by the tool, but were less certain about its use in the future. To improve student satisfaction, instructors can give students a stronger voice in the criteria configuration process and make them aware as to *why* they were assigned to a particular team (e.g., give anonymized summaries showing how the teams scored on the criteria). Instructors appreciated increased efficiency of team formation but also wanted the tool to offer aids such as exemplars and in-situ explanations for selecting the most appropriate criteria for their course and the ability to explore how various configurations affect team assignments.

ACKNOWLEDGMENTS

This work was partially funded by the Strategic Instructional Innovations Program at the University of Illinois and by NSF award CCF-1439957. We thank professors Tao Xie, Ranjitha Kumar, and Alex Kirlik for helping with data collection. We also thank Helen Wauck and Sanorita Dey for helping with data analysis, and Yiwen Xu and Yi-Chieh Lee for assisting with the initial deployment of the team formation tool.

REFERENCES

1. CATME Smarter Teamwork. Available from: http://info.catme.org/.
2. ABET. Criteria for Accrediting Computing Programs: Effective for Reviews During the 2014-2015 Accreditation Cycle. Available from: http://www.abet.org/wp-content/uploads/2015/04/E001-14-15-EAC-Criteria.pdf.
3. Anthony Lising Antonio, Mitchell J Chang, Kenji Hakuta, David A Kenny, Shana Levin and Jeffrey F Milem. 2004. Effects of racial diversity on complex thinking in college students. *Psychological Science*, 15 (8): 507-510. http://dx.doi.org/10.1111/j.0956-7976.2004.00710.x
4. Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman and Eric Scharff. 2000. Transcending the individual human mind - creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction*, 7 (1): 84-113. http://dx.doi.org/10.1145/344949.345015
5. Donald R Bacon, Kim A Stewart and William S Silver. 1999. Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, 23 (5): 467-488. http://dx.doi.org/10.1177/105256299902300503
6. Donald R Bacon, Kim A Stewart and Sue Stewart-Belle. 1998. Exploring predictors of student team project performance. *Journal of Marketing Education*, 20 (1): 63-71. http://dx.doi.org/10.1177/027347539802000108
7. Julia B Bear and Anita Williams Woolley. 2011. The role of gender in team collaboration and performance. *Interdisciplinary Science Reviews*, 36 (2): 146-153. http://dx.doi.org/10.1179/030801811X13013181961473
8. James L Brickell, David B Porter, Michael F Reynolds and Richard D Cosgrove. 1994. Assigning students to groups for engineering design projects: A comparison of five methods. *Journal of Engineering Education*, 83 (3): 259-262. http://dx.doi.org/10.1002/j.2168-9830.1994.tb01113.x
9. Rupert Brown. 2000. *Group Processes: Dynamics Within and Between Groups*. Blackwell Publishing.
10. Kenneth J Chapman, Matthew Meuter, Dan Toy and Lauren Wright. 2006. Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. *Journal of Management Education*, 30 (4): 557-569. http://dx.doi.org/10.1177/1052562905284872
11. Dawn R Deeter-Schmelz, Karen Norman Kennedy and Rosemary P Ramsey. 2002. Enriching our understanding of student team effectiveness. *Journal of Marketing Education*, 24 (2): 114-124. http://dx.doi.org/10.1177/0273475302242004
12. Amy Edmondson. 1999. Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, 44 (2): 350-383. http://dx.doi.org/10.2307/2666999
13. Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, ACM, 2371-2382. http://dx.doi.org/10.1145/2858036.2858494
14. Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton and Christian Sandvig. 2015. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, ACM, 153-162. http://dx.doi.org/10.1145/2702123.2702556
15. Graham D Fenwick and Derrick J Neal. 2001. Effect of gender composition on group performance. *Gender, Work & Organization*, 8 (2): 205-225. http://dx.doi.org/10.1111/1468-0432.00129
16. Daniel Goleman. 1998. *Working with emotional intelligence*. Bantam.
17. Narasimhaiah Gorla and Yan Wah Lam. 2004. Who should work with whom?: building effective software project teams. *Communications of the ACM*, 47 (6): 79-82. http://dx.doi.org/10.1145/990680.990684
18. J. Richard Hackman (ed.). 1990. *Groups That Work (and Those That Don't): Creating Conditions for Effective Teamwork*. Jossey-Bass Publishers, San Francisco.
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): 987-1015. http://dx.doi.org/10.1177/0149206307308587
20. Roland Hübscher. 2010. Assigning students to groups using general and context-specific criteria. *IEEE Transactions on Learning Technologies*, 3 (3): 178-189. http://dx.doi.org/10.1109/TLT.2010.17
21. David Hunkeler and Julie E Sharp. 1997. Assigning functional groups: The influence of group size, academic record, practical experience, and learning style. *Journal of Engineering Education*, 86 (4): 321-332. http://dx.doi.org/10.1002/j.2168-9830.1997.tb00305.x

22. David S Jalajas and Robert I Sutton. 1984. Feuds in student groups: Coping with whiners, martyrs, saboteurs, bullies, and deadbeats. *Journal of Management Education*, 9 (4): 94-102. http://dx.doi.org/10.1177/105256298400900413

23. Irving Lester Janis. 1982. *Groupthink: Psychological studies of policy decisions and fiascoes*. Houghton Mifflin Boston.

24. Karen A Jehn, Gregory B Northcraft and Margaret A Neale. 1999. Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, 44 (4): 741-763. http://dx.doi.org/10.2307/2667054

25. George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

26. J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33 (1): 159-174. http://dx.doi.org/10.1145/2675133.267516610.2307/2529310

27. 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): 1-28.

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. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ACM, 260-273. http://dx.doi.org/10.1145/2818048.2819979

29. Ioanna Lykourentzou, Robert Kraut and Steven P. Dow. 2017. Team Dating Leads to Better Online Ad Hoc Collaborations. *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*, to appear.

30. Sakthi Mahenthiran and Pamela J Rouse. 2000. The impact of group selection on student performance and satisfaction. *International Journal of Educational Management*, 14 (6): 255-265. http://dx.doi.org/10.1108/09513540010348043

31. Barbara Oakley, Richard M Felder, Rebecca Brent and Imad Elhajj. 2004. Turning student groups into effective teams. *Journal of Student Centered Learning*, 2 (1): 9-34.

32. Michael A Redmond. 2001. A computer program to aid assignment of student project groups. *ACM SIGCSE Bulletin*, 134-138. http://dx.doi.org/10.1145/366413.364562

33. Marcel M Robles. 2012. Executive perceptions of the top 10 soft skills needed in today's workplace. *Business Communication Quarterly*, 75 (4): 453-465. http://dx.doi.org/10.1177/1080569912460400

34. Timothy S Schoenecker, Kathryn D Martell and Joseph F Michlitsch. 1997. Diversity, performance, and satisfaction in student group projects: An empirical study. *Research in Higher Education*, 38 (4): 479-495. http://dx.doi.org/10.1023/A:1024966627400

35. Greg L Stewart. 2006. A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32 (1): 29-55. http://dx.doi.org/10.1177/0149206305277792

36. A.L. Strauss. 1987. *Qualitative Analysis for Social Scientists*. Cambridge University Press.

37. James T Strong and Rolph E Anderson. 1990. Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, 12 (2): 61-67. http://dx.doi.org/10.1177/027347539001200208

38. Anya Tafliovich, Andrew Petersen and Jennifer Campbell. 2016. Evaluating Student Teams: Do Educators Know What Students Think? *Proceedings of the ACM Technical Symposium on Computing Science Education*, ACM, 181-186. http://dx.doi.org/10.1145/2839509.2844647

39. Martin Wessner and Hans-Rüdiger Pfister. 2001. Group formation in computer-supported collaborative learning. *Proceedings of the ACM Conference on Supporting Group Work*, ACM, 24-31. http://dx.doi.org/10.1145/500286.500293

40. Anita Williams Woolley, I. Aggarwal and Thomas W. Malone. 2015. Collective intelligence in teams and organizations. Malone, T.W. and Bernstein, M.S. eds. *The handbook of collective intelligence*, MIT Press, Cambridge, MA, in press.

41. 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): 686-688. http://dx.doi.org/10.1126/science.1193147