Supporting Instructor Decisions on Algorithmic Team Formation through Integrating Stakeholder Voices

Dr. Emily Hastings

Overview

- Introduction
- What is HCI?
- My research
- LIFT

Who Am I?

- Assistant Professor of CS
- Illinois native (Galesburg)
- Graduated from Knox College in 2016
  - Major: Computer Science
  - Minor: Renaissance/Medieval Studies (self-designed)
  - CS TA and research assistant
  - Costume Shop
- Recently finished PhD at University of Illinois
  - Human-computer interaction

“Human–computer interaction (**HCI**) studies the design and use of computer technology, focused on the interfaces between people (users) and computers. Researchers in the field of HCI observe the ways in which humans interact with computers and design technologies that let humans interact with computers in novel ways.”

(Wikipedia)

Human-Computer Interaction

- Multidisciplinary
  - Computer science
  - Psychology
  - Cognitive Science
  - Design
  - Human factors
- Venues
  - ACM Conference on Human Factors in Computing Systems (CHI)
  - ACM Conference on Computer Supported Cooperative Work (CSCW)
  - Many others

Learner-Centered Algorithmic Team Formation

Vision

- Every student can:
  - Have a positive team experience
  - Learn and contribute to a quality team outcome
  - Work on a high-performing team
  - Learn how to be a better teammate
- Bring together CS, the learning sciences, and other fields to design, deploy, and study **a new genre of algorithmic team formation process** that more closely considers the needs and experiences of learners

Team Formation Matters

- Teamwork is a foundational skill and integral in computing courses *(e.g., Coleman & Lang 2012; Latulipe, Long, & Seminario 2015)*
- Instructors must determine how to best form teams from increasingly large and diverse pool of students

2 Common Possibilities

- Self-selection
- Random assignment
- Strengths
  - Easy to implement
  - Some students prefer to select their own team
- Weaknesses
  - Students may struggle to find a team to join
  - Lack of skill diversity

Criteria-based Team Formation

Strategically select team members to achieve certain compositions

- Skill diversity *(e.g., Brickell et al. 1994, Horwitz and Horwitz 2007)*
- Balanced personality types *(e.g., Lykourentzou et al. 2016)*
- Balanced genders *(e.g., Jehn, Northcraft, and Neale 1999)*
- Many more

Example Algorithmic Workflow

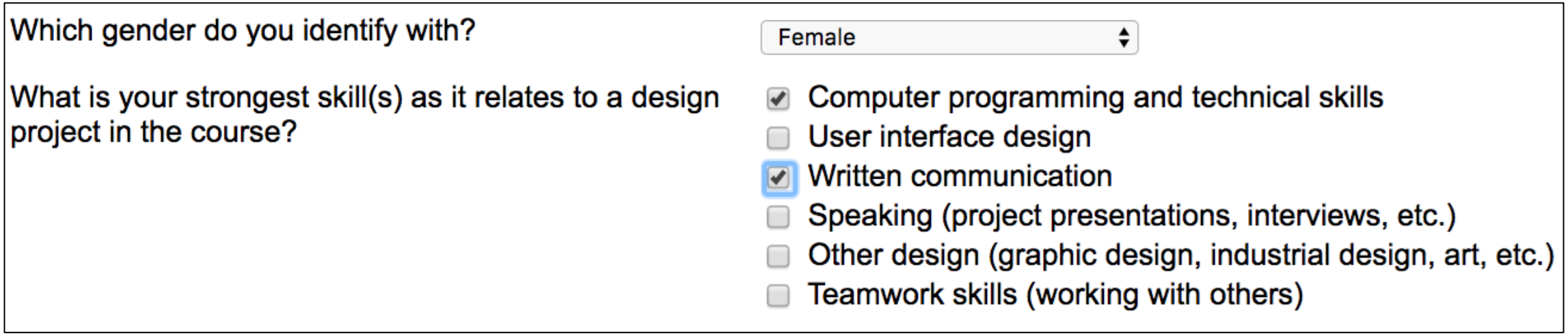
CATME
SMARTER Teamwork

1. Select criteria

Example Algorithmic Workflow

CATME SMARTER Teamwork

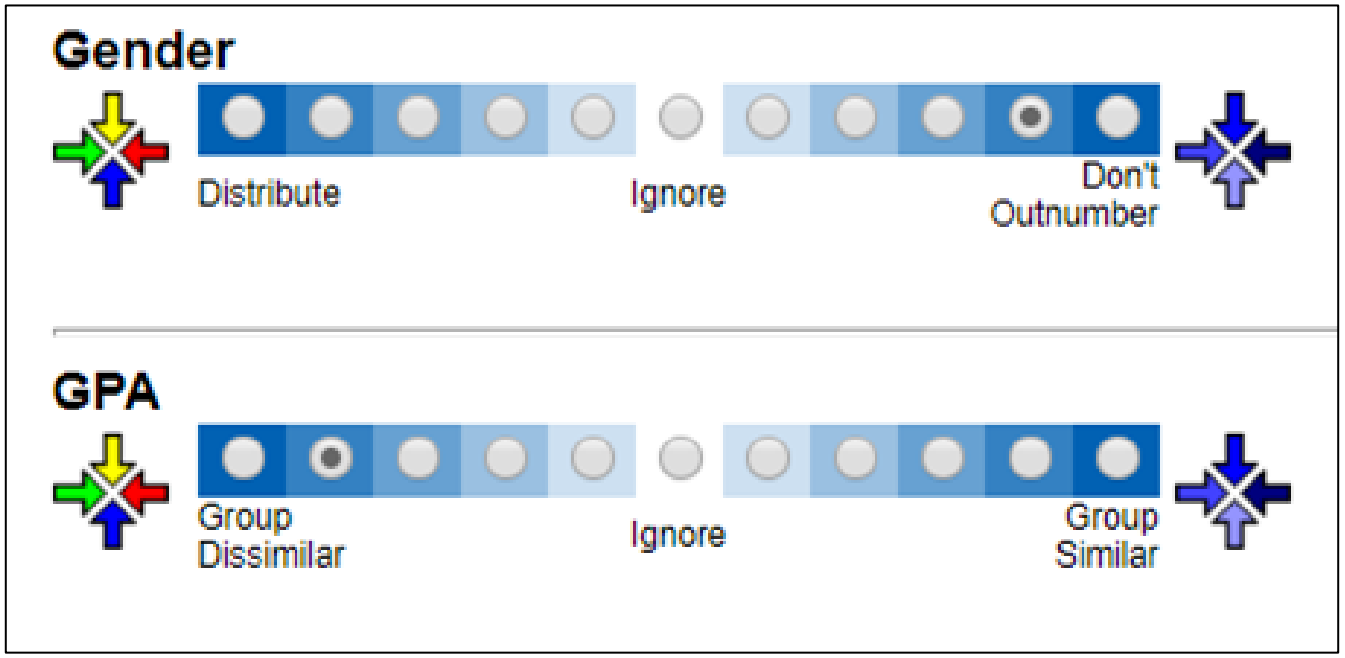
1. Select criteria
2. Gather information from students



Example Algorithmic Workflow

CATME
SMARTER Teamwork

1. Select criteria
2. Gather information from students
3. Configure weights



Example Algorithmic Workflow

CATME
SMARTER Teamwork

1. Select criteria
2. Gather information from students
3. Configure weights
4. Form teams

Example Algorithmic Workflow

CATME
SMARTER Teamwork

1. Select criteria
2. Gather information from students
3. Configure weights
4. Form teams
5. Notify students

The Problem

- Instructor must prioritize large collection of criteria available in the tool with little guidance *(Jahanbakhsh et al. 2017)*
- Could consult literature
  - Does not cover all possible combinations
  - May not align with preferences of learners
- Not clear how instructors are selecting configurations
- Teams formed algorithmically could suffer from “cold start” problems

My Approach

- New learner-centered techniques for algorithmic team formation
  - Increasing student agency and ownership can improve learning outcomes *(e.g., Chan et al. 2014, Conley & French 2014, Horwitz et al. 2009, Mello 1993)*
- Open questions:
  - How are instructors making decisions?
  - Which criteria do students prioritize?
  - To what degree do students agree with each other and the literature?
  - What impact does using student configurations have on team outcomes?
  - How to support algorithmically formed teams?

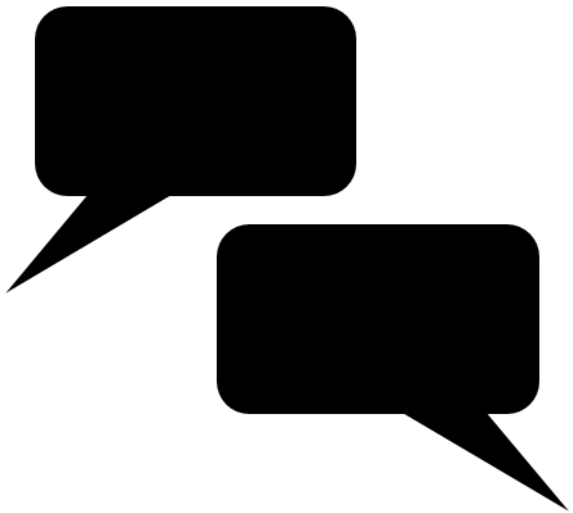
My Research

Effects of team-building and composition
(CSCW 2018)

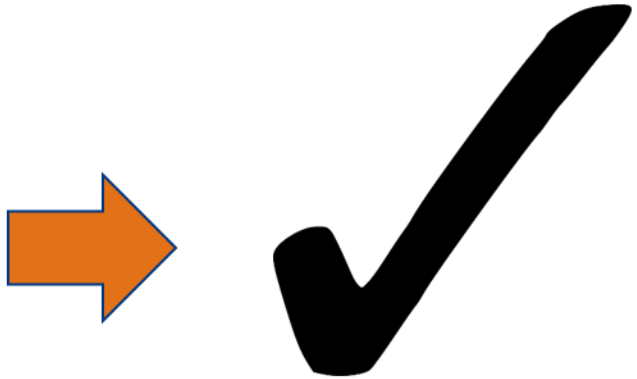
LIFT workflow and extensions
(CHI 2020, IJPHM 2021, SIGCSE 2022)

Examination of current instructor practices
(CSCW 2023)

The LIFT Workflow



Propose and discuss potential criteria



Vote on criteria

Configure weights

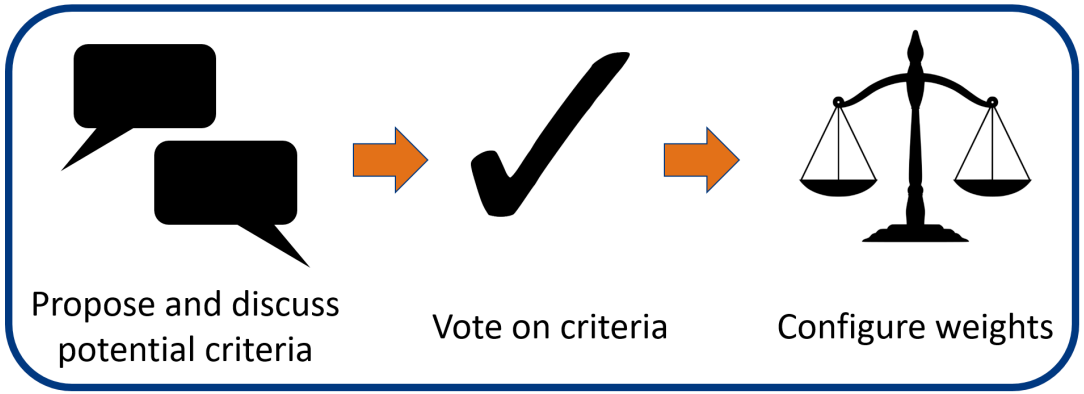
Research Questions

- RQ1: What team formation criteria do students select when given the chance? How do student and instructor choices differ?
- RQ2: How do students perceive their agency when they are allowed to have input into the team formation process?
- RQ3: How does allowing students to select criteria affect their team performance, satisfaction, and other course experiences compared to having instructors select criteria?
- RQ4: How do instructors perceive transferring agency in the team formation process to students, and what do they learn about student preferences?

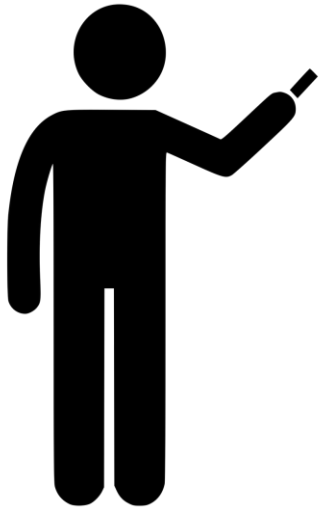
Experimental Design

- Mixed-methods between participants experiment (N=289)
  - Interviews with 18 students and 6 instructors

Condition 1: Learner (LIFT)



Condition 2: Instructor (Control)



Measures

- Project Grades
- Perceived Performance
- Satisfaction with Team Assignment
- Satisfaction with Team Formation Process
- Recommendation to Repeat Approach
- Perceived Agency
- Importance of Input

RQ1: Student Criteria Choices

- 75 criteria discussed in total, 48 (64%) newly-proposed
  - E.g., Organizational style
- 3 broad categories:
  - Team management (e.g., Leadership role, Teamwork experience)
  - Academics (e.g., GPA, Software skills)
  - Identity (e.g., Gender, Personality type)
- Voting phase eliminated all less serious criteria
- Most popular: scheduling, skills, work habits
- Least popular: aspects of past and identity not under present control

RQ1: Instructor Criteria Choices

- All instructor criteria selected from tool
- Prioritized learning and long-term success over minimizing present conflict
    - “High achievers may need to be in teams with other high achievers so that they have this sort of conflict…[and] can work through a disagreement with another student. I think it is a wonderful opportunity for growth.” (I2)

RQ2: Student Perceptions of Agency

- Students found it important to have a voice (median 6.0)
- Median agency score in Learner condition was higher (median 5.0 vs 4.0), but not statistically significant (Wald $\chi 2(1)=3.05$, B= 0.77, p=0.08)
    - Possible explanation: participation vs. choice

RQ2: Student Perceptions of Agency, cont.

- Strengths:
  - LIFT can provide insight to instructors who are disconnected from the student team experience (S=10)
  - LIFT contributed to increased sense of ownership (S=5)
- Weaknesses:
  - Students are not experts on what makes a good team (S=6)
  - Instructors more familiar with the course and what skills will be necessary (S=8)
  - Concerns of gaming (S=5)

RQ3: Effects of Criteria Selector on Outcomes

- High across conditions
- No significant effect of criteria selector
- Potential explanation:
  - Specifics of criteria configuration may not be most important factor in outcomes

RQ4: Instructor Perceptions

- Found student choices reasonable overall, including confirming personal doubts
- Some doubts about irrelevant criteria, gaming concerns, excluding important criteria
- Three instructors would adopt LIFT as-is, a fourth would integrate student criteria into his own configuration
  - Responsibility, motivation, sense of ownership
- Remaining two instructors were reluctant to adopt due to key exclusions or large course sizes

Implications

- Possible to incorporate student input into algorithmic team formation without adversely affecting grades or team experiences
- Alternatives to full LIFT workflow:
  - Adopt simplified version of LIFT for convenience
    - E.g., vote only on weights
  - Integrate student- and instructor-chosen criteria in a single configuration
    - Protect voices of minority students

Extensions

- Exploration of applying LIFT in other contexts (IJPHM 2021)
  - NIST internship: manufacturing maintenance
- **Follow-up study: extension of LIFT inspired incorporating consensus-building process (SIGCSE 2022)**

Follow-up Study Research Questions

RQ1: What are the criteria and weights that are most valued by students for team formation, and why?

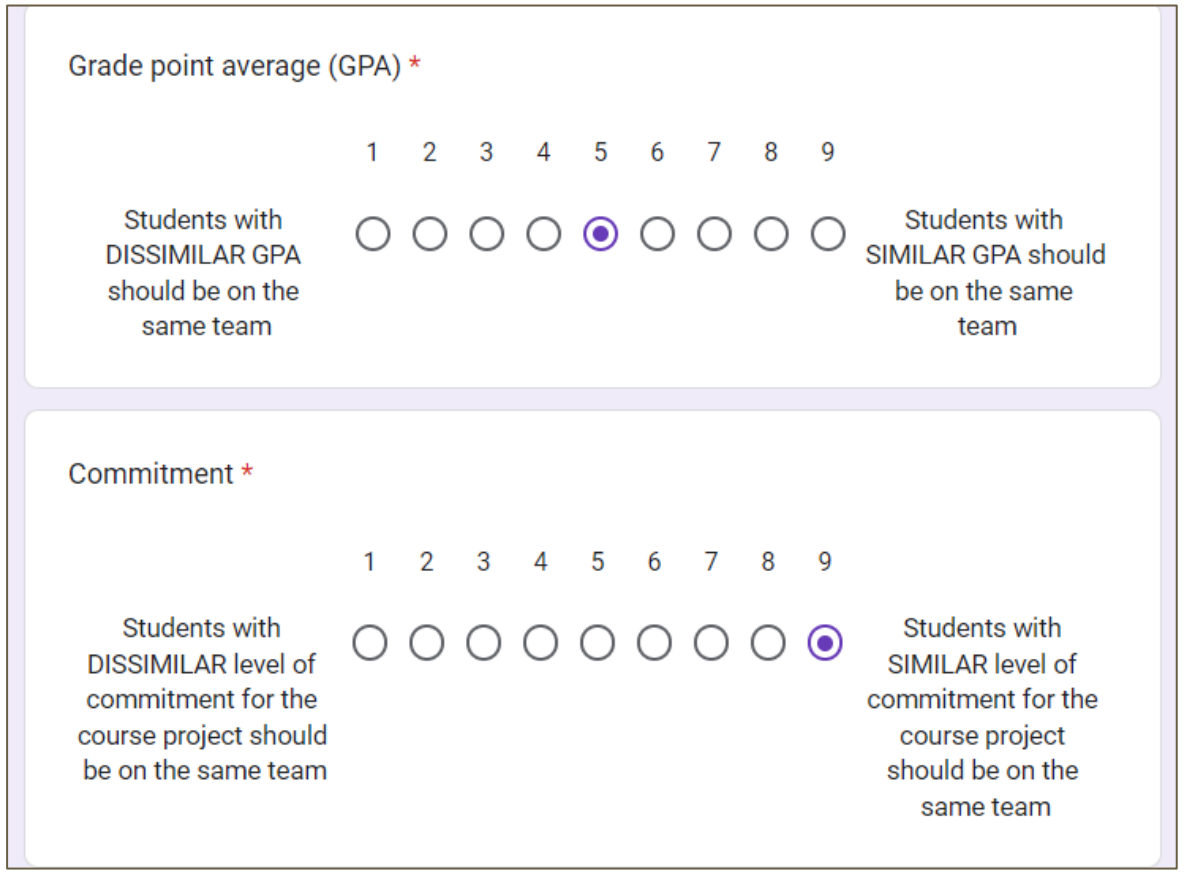
RQ2: Which criteria do students agree on the most and least?

RQ3: How does incorporating an online peer discussion impact student selections for the criteria weights?

RQ4: What do students learn from the team formation discussion and what are their experiences with the process overall?

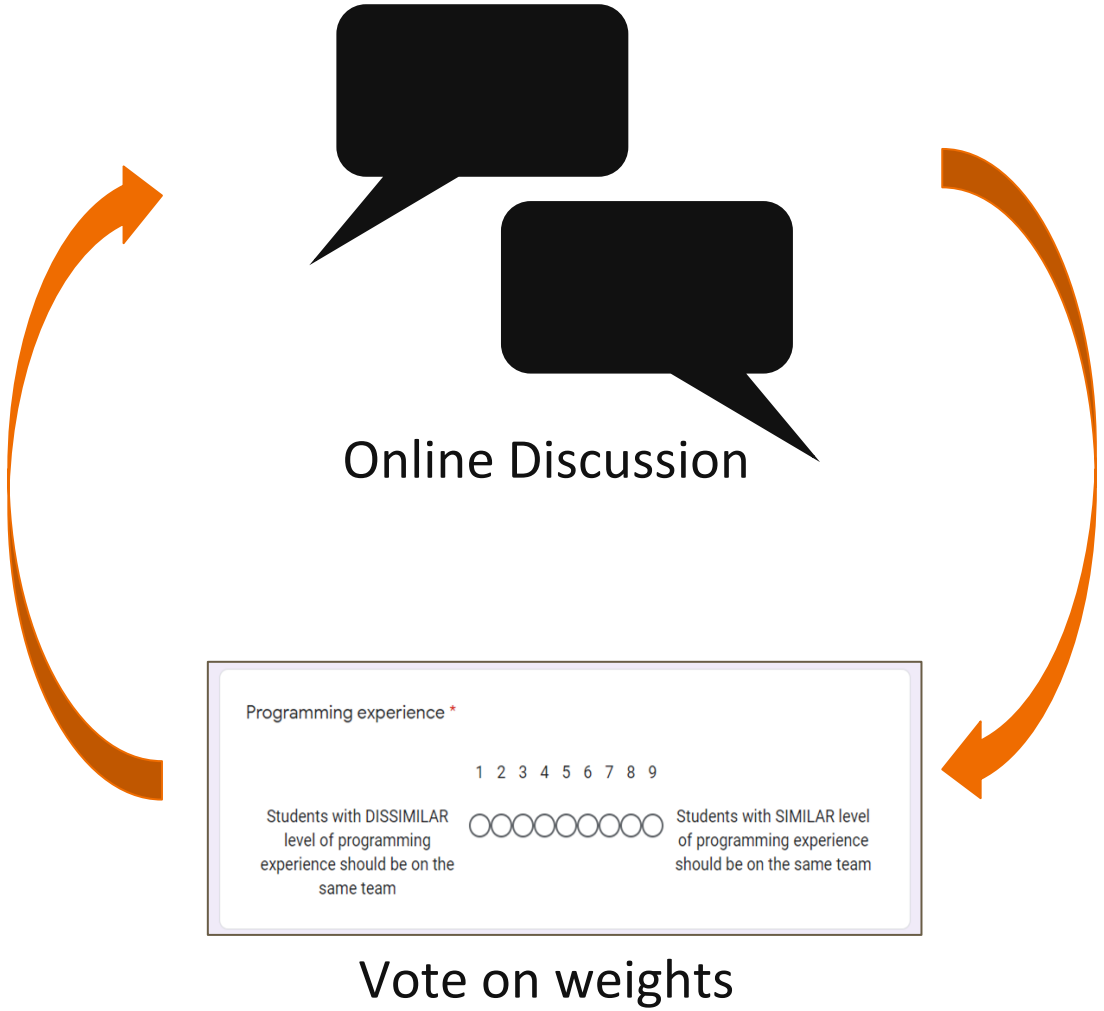
Workflow

- Project-based UI design course
- Students select weights for each of 13 candidate criteria via an online survey
- Preferences aggregated to form final configuration entered into the tool



Workflow (SP21)

- Project-based UI design course
- Students select weights for each of 13 candidate criteria via an online survey
- Preferences aggregated to form final configuration entered into the tool
- **Online peer discussion and opportunity to revise weight selections**



Measures

- Student-generated weights for selected criteria (and agreement) over 4 semesters
- Additional measures from SP21:
  - Overall perceptions of criteria voting activity (Likert-scale)
  - Student rationales from discussion (open coded)
  - What students learned from participating in activity

RQ1: Criteria and Weights Valued by Students

	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)

Each cell:
Mean (variance)

Scale:
-5 (dissimilar)
to 5 (similar)

RQ1: Criteria and Weights Valued by Students

Criterion	Avg (Var)	SP19	FA19	FA20	SP21
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 (3.4)**	**2.7 (2.1)**

"It's unfair to someone who is working really hard for the group to be paired with people who aren't going to do their jobs."
(P42, Commitment)

RQ1: Criteria and Weights Valued by Students

Criterion	Avg (Var)	SP19	FA19	FA20	SP21
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)**

"I do feel more comfortable having another woman to advocate with on a team ... 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."
(P19, female, Gender)

RQ2: Agreement

Criterion	Avg (Var)	SP19	FA19	FA20	SP21
Work Style	0 (10.7)	0.5 (8.6)	-0.2 (7.6)	0.3 (8.8)	-0.6 (8.2)
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)
Programming Experience	-0.8 (8.2)	-	-1.1 (6.0)	-0.2 (6.8)	-1.4 (5.9)
Area of Study	-1.2 (7.6)	-	-1.4 (7.2)	-0.8 (6.1)	-1.3 (3.5)

"People with dissimilar programming experience should be grouped together to make the team more well-rounded."
(P25, Programming)

"If a student who excels at writing code [is assigned] to a group where most students don't have experience... then [that student] would end up doing all the coding."
(P3, Programming)

RQ2: Agreement

Criterion	Avg (Var)	SP19	FA19	FA20	SP21
Work Style	0 (10.7)	0.5 (8.6)	-0.2 (7.6)	0.3 (8.8)	-0.6 (8.2)
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)
Programming Experience	-0.8 (8.2)	-	-1.1 (6.0)	-0.2 (6.8)	-1.4 (5.9)
Area of Study	-1.2 (7.6)	-	-1.4 (7.2)	-0.8 (6.1)	-1.3 (3.5)

More agreement (lower variance) than previous terms

RQ3: Impact of Discussion

- 24 students changed their configuration, sometimes more than once
- Rationales for (not) changing vote:
  - Gained new perspectives from the discussion (N=29)
  - Personal reflection (N=10)
  - Independent research (N=2)
  - Original opinion was reinforced by the discussion (N=4)

RQ4: Learning about Team Formation

- Thought about what makes a good team (M=6 on scale from 1 to 7)
- Explanations:
  - Learned new perspectives from discussion (N=24)
  - Reflected on personal values for teamwork (N=23)
  - Thought about how to be a better team member (N=15)

What should you do?

- Utilize the dataset as a starting point or employ workflow
- Students benefit from the discussion
- Actively review aggregated student votes
- Consider different voting and aggregation methods
- Students: advocate for yourselves!

Discussion and Implications

- Instructors desire more guidance when configuring tools
- Input from students is one way to help make decisions
  - Can incorporate student priorities without negatively affecting outcomes
  - Generally prefer schedule, commitment, etc.
  - Instructors might still include gender, race/ethnicity, etc. to benefit students underrepresented in their area
- Tool designers could incorporate features for gathering student input, facilitating knowledge sharing, team-building
  - E.g., NLP techniques to facilitate gaining insight from discussions

Future Vision: Data-Driven Teamwork Lifecycle

- Evidence-based skill assessment (e.g., prior course assignments, quizzes)
  - Addresses current tools' reliance on self-reported data of uncertain quality
  - Gaming, difficulties with self-assessment
- Teamwork profiles
  - History of team compositions, outcomes, peer evaluations, etc.
  - Can help identify learning opportunities for students (e.g., which kinds of teams would be helpful to work on before graduating)
  - Curate to show to potential employers
  - Data mining/ML possibilities
  - Generalize to different contexts

Questions?

hastinem@uwec.edu

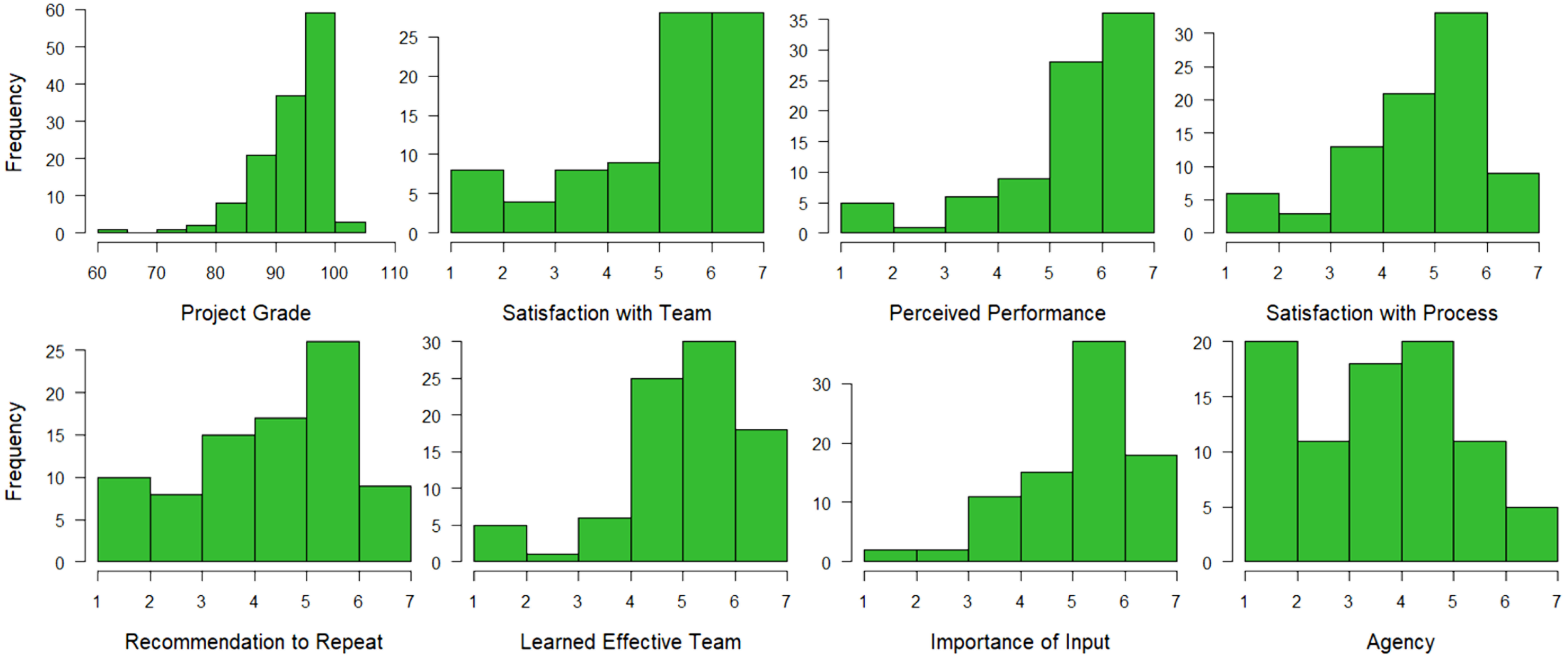
emhastings.github.io

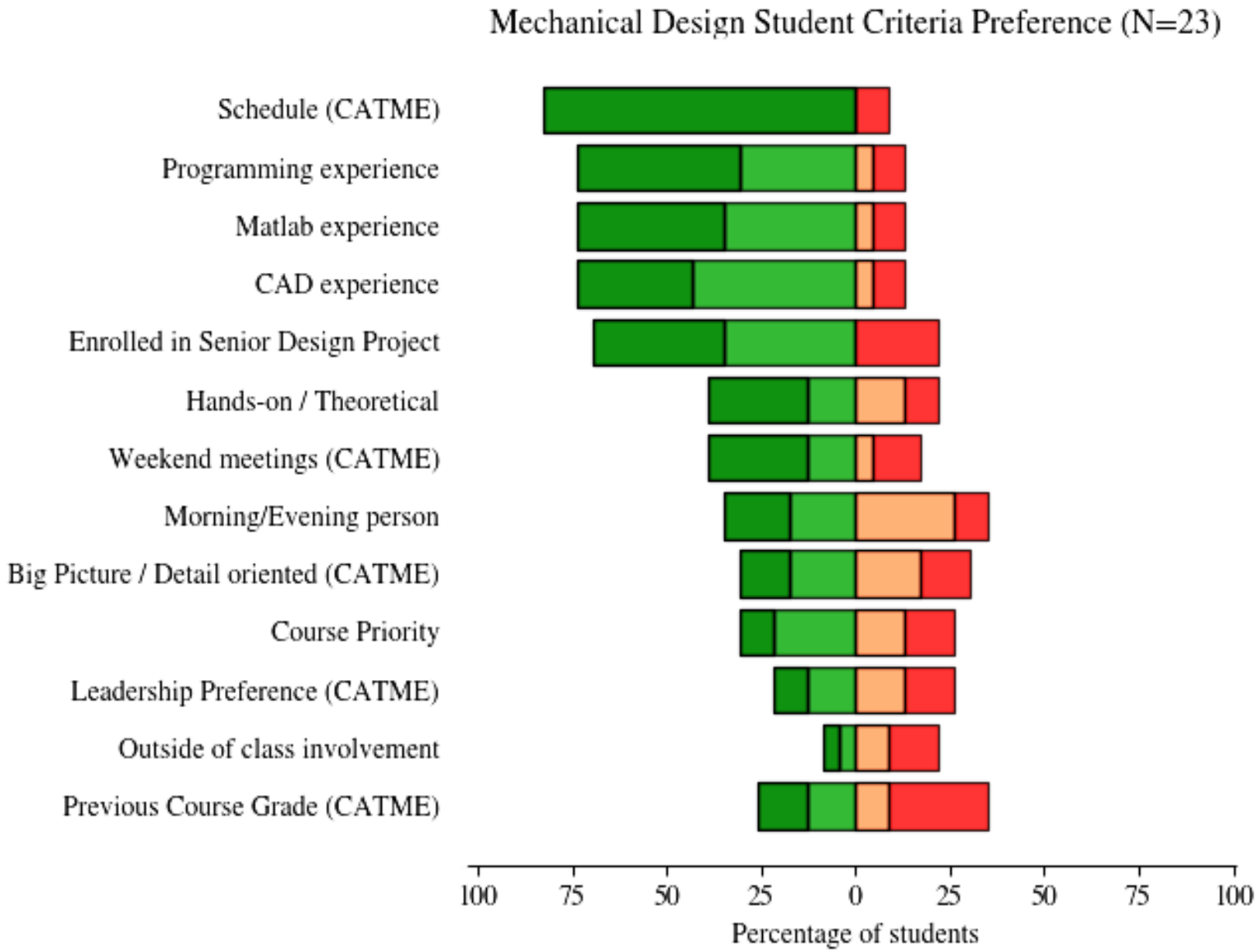
Extra Slides

LIFT

Quotes

- *“Was GPA on? See, GPA is not even on there! Gosh, see that! The students are smarter than me… See, I guess I wish [I had] heard or learned this earlier.”* (I2)
- *“That's a hard question… there's a lot of literature on gender and achievements and race, like we should really pay attention to that, but then again I don't know. I'm not the students, and I don't know what their biases are, if they have biases… all I know is literature so… I don't know. I don't know if I trust that much that they know themselves so well.”* (I3)


Frequency
Project Grade
Satisfaction with Team
Perceived Performance
Satisfaction with Process
Frequency
Recommendation to Repeat
Learned Effective Team
Importance of Input
Agency


Mechanical Design Student Criteria Preference (N=23)
Schedule (CATME)
Programming experience
Matlab experience
CAD experience
Enrolled in Senior Design Project
Hands-on / Theoretical
Weekend meetings (CATME)
Morning/Evening person
Big Picture / Detail oriented (CATME)
Course Priority
Leadership Preference (CATME)
Outside of class involvement
Previous Course Grade (CATME)
100
75
50
25
0
25
50
75
100
Percentage of students