



# COMPUTER SCIENCE & ENGINEERING

UNIVERSITY OF MICHIGAN



# Gender-balanced TAs from an Unbalanced Student Body

Amir Kamil, James Juett, and Andrew DeOrio

University of Michigan

SIGCSE 2019

# Context

- CS2 course at the University of Michigan
  - ~1000 students a semester, over 5 lecture sections and >30 lab sections
  - Topics: procedural and data abstraction, pointers and arrays, dynamic resource management, linked structures, recursion, trees
  - 25-30 undergraduate teaching assistants (TAs), 4-6 graduate TAs
- Focus of this work: undergraduate TAs

# The Challenge of Hiring a Gender-balanced Staff

- Fraction of overall population that is women

AP CS test-takers	23%
<b>CS2 at University of Michigan</b>	<b>25%</b>
<b>Declared CE/CS/DS majors at UM</b>	<b>20%</b>
CS degree at major research university	18%
Professional computing occupations	26%

- Teaching assistants form front line of our courses – hold lab sections, office hours, answer Piazza questions, ...
- Representation of women on staff important as role models, improving retention of women in CS

# Research Questions

- What is the gender balance at all phases of the undergraduate-TA application process?
- Do women and men perform differently in the evaluative measures used?

# Previous Hiring Process

- Hiring new TAs before Fall 2016:
  - Ad hoc process
  - Informal faculty interview
- Issues of fairness and scaling
  - >100 applicants, can't interview them all
  - Course/staff sizes becoming larger, more faculty involved

# New Hiring Process

- New process (Fall 2016+) based on that of Dr. Mary Lou Dorf in CS1
- Two-phase hiring process for new TAs
  - Applicants submit teaching videos (100-150 applicants)
  - Videos determine which candidates are interviewed in person (20-25 interviews)
  - Hiring based on in-person interviews (6-12 new TAs hired)



# Application Content

- Prior teaching experience, why the interest in teaching CS2
- Link to 5-minute teaching videos on the CS2 topic of their choice
- Academic information
- **We do not consider GPA or grade in deciding who to interview**

# Review Process

- Faculty lead watches all videos (at 2x speed), rates them on 5-point scale
- Those that score  $\geq 3.5$  get second opinion from another faculty member
- Criteria for inviting to in-person interview:
  - Video ratings (most important)
  - Experience and why they are interested
  - Recommendations by faculty
  - **We do not consider GPA or grade in CS2 in deciding who to interview**

# In-person Interviews

- Each candidate is interviewed by 2 faculty members
  - 30-minute slot (20-25 minutes + 5-10 minute buffer)
- First part of interview: standard set of questions
  - Why are you interested in teaching?
  - What do you like about the course and what do you think can be improved?
  - A diversity and inclusion question
    - e.g. How can we make the climate in our course better for underrepresented students?

# In-person Teaching Demos

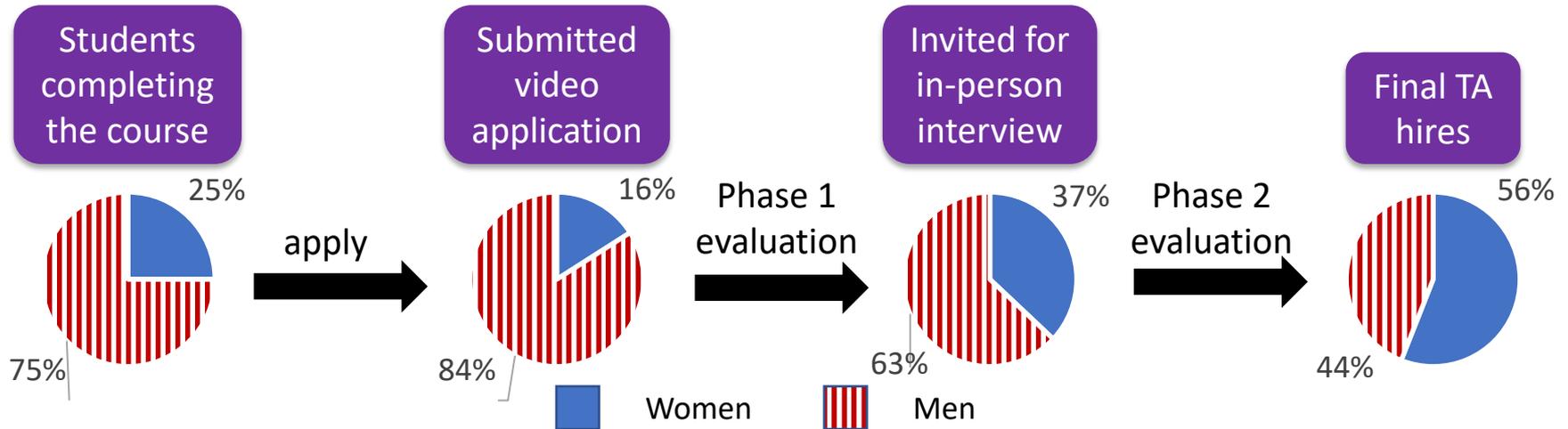
- Second part of interview: teaching demonstration
  - We tell candidates the topic in advance
  - We make it clear we're interested in teaching style, not technical knowledge
  - We ask realistic questions, based on common misconceptions
- Each faculty member rates 4 aspects of their teaching
  - Clarity
  - Technical proficiency
  - Use of whiteboard
  - Responsiveness to student questions and needs

# Data Collection and Statistical Methods

- Data sets for analysis
  - Teaching-video scores for first-time applicants
  - Interview scores for the 4 evaluated categories
  - Course evaluations collected by the university for each TA
- Demographic and academic data from university analytics system
  - Gender (system only tracks binary gender)
  - GPA at the time of application and grade in CS2
- 2-sided Student's t-tests for statistical significance ( $p < 0.05$ )
- Pearson for correlation, followed by t-test for significance

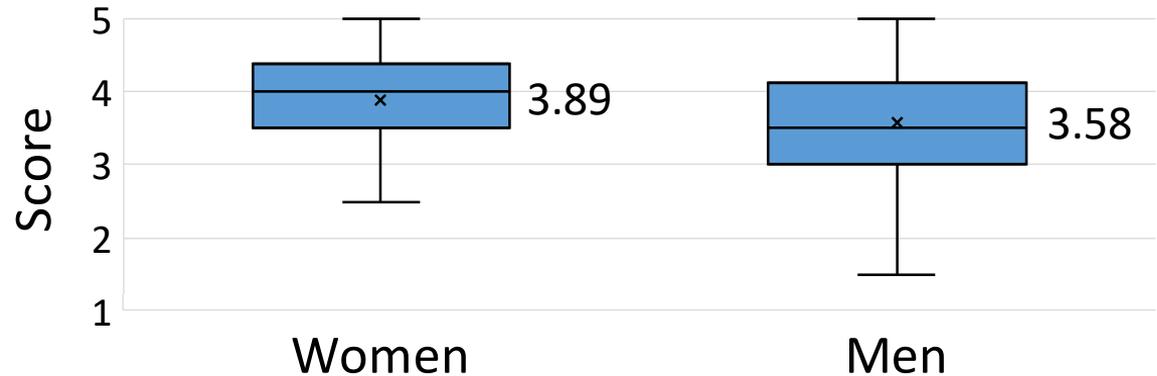
# Gender Balance at Each Step

- Women underrepresented in applicant pool (16.5%) compared to population in course (25%)
- Representation increases significantly at each subsequent step (37% of candidates interviewed, 56% of those hired)



# Evaluation of Teaching Videos

- Average video score for women is 9% higher than men
  - Statistically significant  $p = 0.0001$

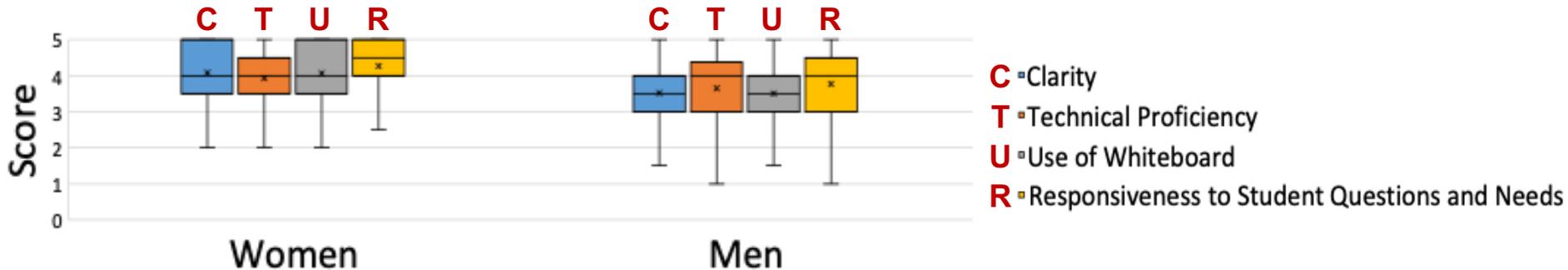


- No significant difference in GPA and grade in CS2 between women and men applicants (average  $\sim 3.65$  GPA for both, A- in CS2)

# Evaluation of In-person Teaching Demonstrations

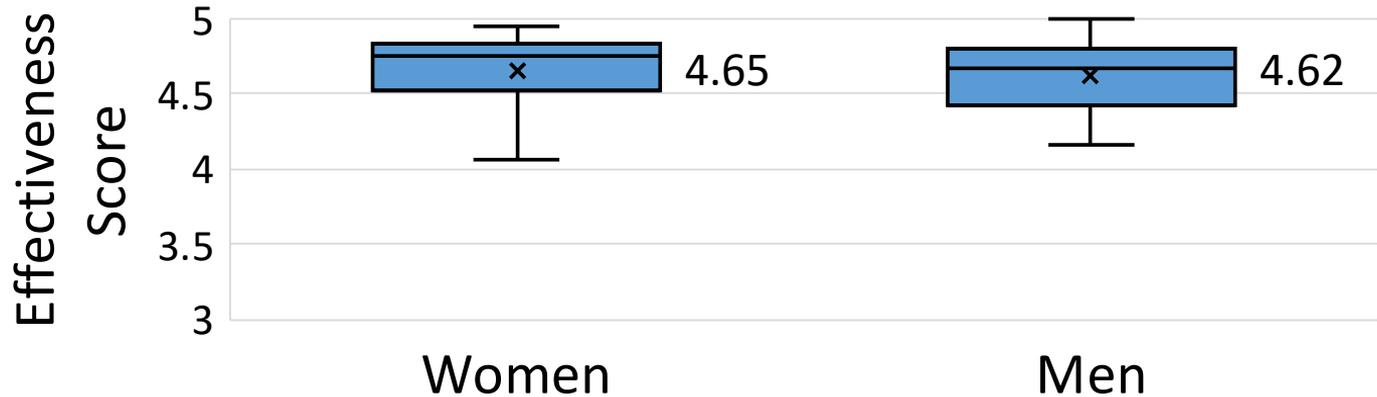
- Women rate significantly better than men in 3 of the 4 categories

Average Score	Women	Men	P-Value
<b>Clarity</b>	4.01	3.52	0.0029
Technical	3.93	3.65	0.091
<b>Use of Whiteboard</b>	4.07	3.51	0.0026
<b>Responsiveness</b>	4.27	3.77	0.011



# Course Evaluations

- No significant difference between women and men ( $p = 0.584$ )
  - **Women TAs are as effective as men**



- No significant difference between new and old processes ( $p = 0.781$ )
  - **Gender balance does not come at the cost of effectiveness**

# Qualitative Observations

- Application videos the most critical component of initial applications
  - Demonstrate applicant's ability to
    - Communicate clearly
    - Use effective visual aids
    - Choose appropriate pacing and detail level
  - Efficient: assess 100-150 candidates in a few days
- In-person teaching demo the most valuable part of the interview
  - Showcases candidate's abilities in an interactive setting

# Gender Differences in Applications

- 75% of videos from women applicants score  $\geq 3.5$  (threshold for second view), compared to 50% from men
- Women also appear to perform better on qualitative parts of the application
  - Prior teaching experience, answers to free-form questions, etc.
- Possible explanations
  - Self-selection, perhaps due to lower confidence levels
    - But not GPA or grade – our data show no difference
  - Lower confidence may lead to more time and effort on video

# Gender Differences in In-person Interviews

- Our data show women do better in in-person teaching demos
- Anecdotally, women also seem to do better in the question/answer part of the interview
  
- **Women do better than men even after filtering everyone through application videos**
  - In-person interviews are important for gender balance

# Challenges

- Getting women to apply is a challenge
  - 25% of students in CS2 are women, but only 16.5% of applicants
- Anecdotal experience: can take significant individual encouragement to convince women to apply
  - TAs can provide more effective encouragement than faculty
- 16% of men apply more than once vs. only 4% of women
  - Takeaway: we should encourage promising applicants to apply again

# Alternative: Hiring Based on GPA or Grade

- Given the same applicant pool, hiring based on GPA or grade would result in a very unbalanced staff
- Just GPA: 17-24% for cutoffs  $\geq 3.6$
- Just grade: 14-18% for cutoffs  $\geq B+$
- Most applicants have a high GPA and grade, so need some other factor for hiring

# Correlation between GPA or Grade and Performance

- No significant correlation between GPA or grade and performance on any metric
- **Validates our decision to not consider GPA or grade**

	GPA		CS2 Grade	
	Correlation	P-Value	Correlation	P-Value
Video	0.0620	0.218	0.0796	0.114
Clarity	0.0431	0.678	0.0747	0.472
Technical	0.107	0.303	0.129	0.214
Use of Whiteboard	-0.0329	0.752	-0.00180	0.986
Responsiveness	-0.00439	0.966	0.0985	0.342
Course Evals	-0.0806	0.523	0.0566	0.654

# Limitations

- Teaching videos can be a barrier to entry
- Unclear whether results would be applicable to upper-level courses
  - More time for students to improve after CS2 than upper-level course
- May be implicit bias in our evaluation process
  - Mitigations
    - Opinions from multiple faculty members
    - Multiple criteria for evaluation
  - Course evaluations show no evidence for favoritism

# Conclusions

- In our experience in a CS2, women do better than men in both teaching-demonstration videos and in-person teaching demos
  - Two-step process has led to a gender-balanced staff without sacrificing teaching effectiveness
  - GPA and grade show no correlation with performance
- The two-step process scales to a large number of applicants
  - ~6-8 hours from each faculty member in our course
  - Well-defined evaluation metrics allow the process to be parallelized
- Explicit consideration of gender was not necessary to achieve a gender-balanced and effective teaching staff