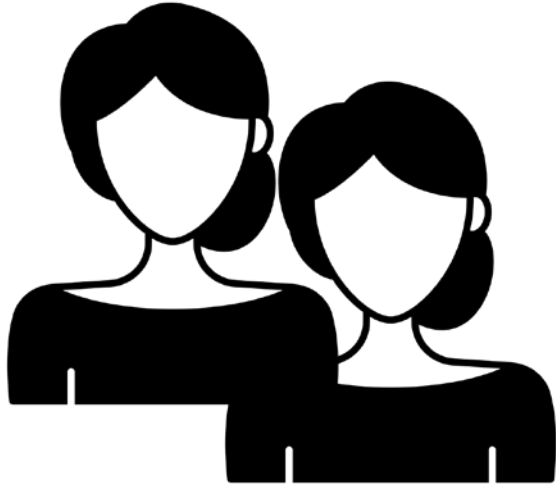


# Investigating Gender Bias in Medical Student Evaluations

Maren Loe, Arghavan Salles

November 16, 2019

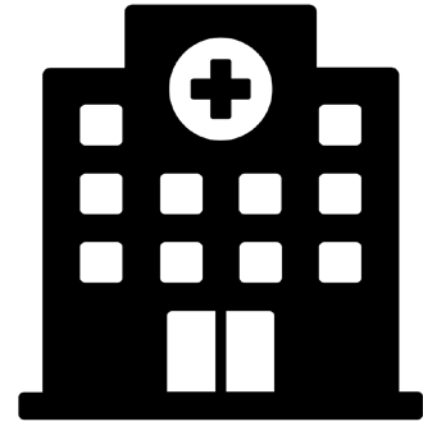
# Patient Care



Similarity

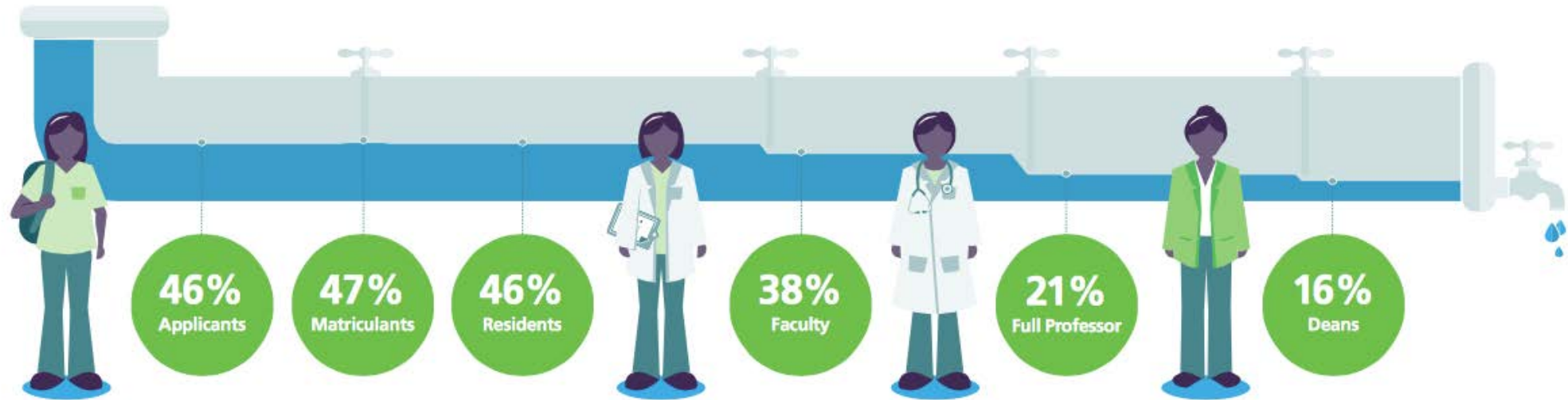


Heart disease



Hospitalized

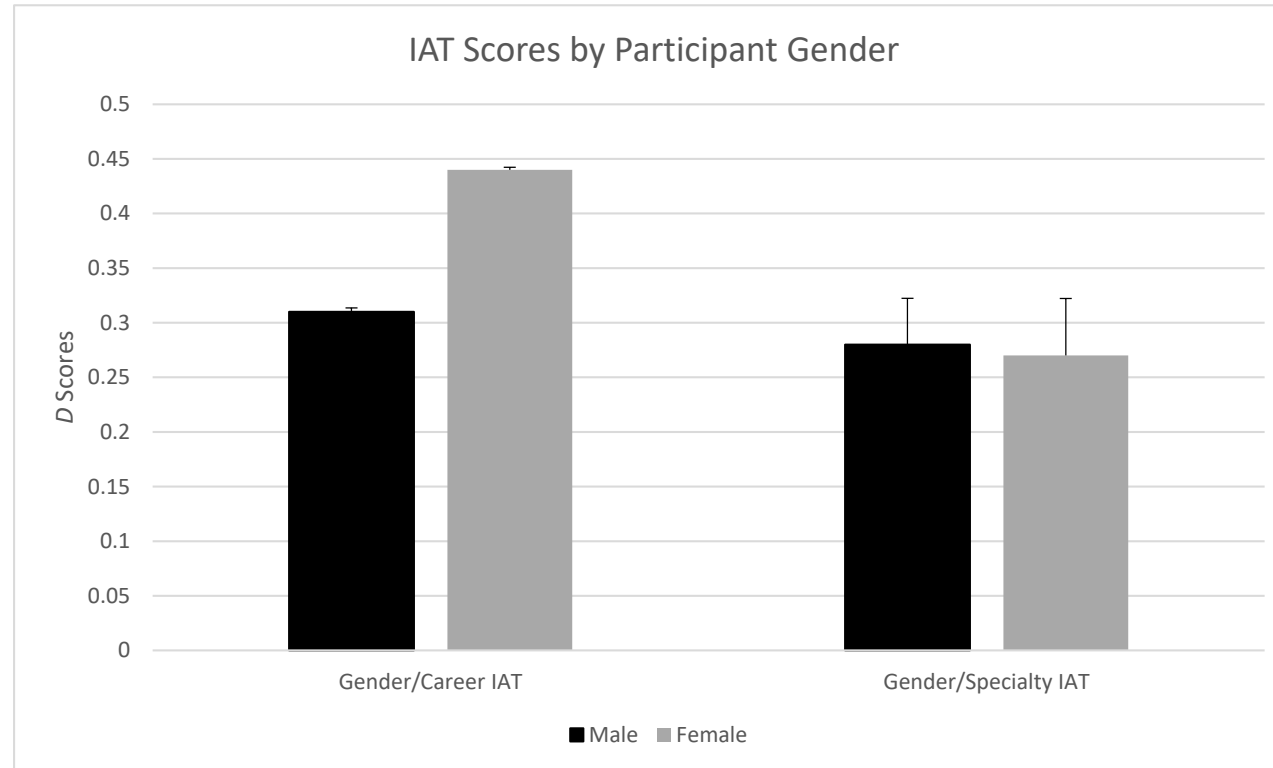
# Where Do The Women Go?



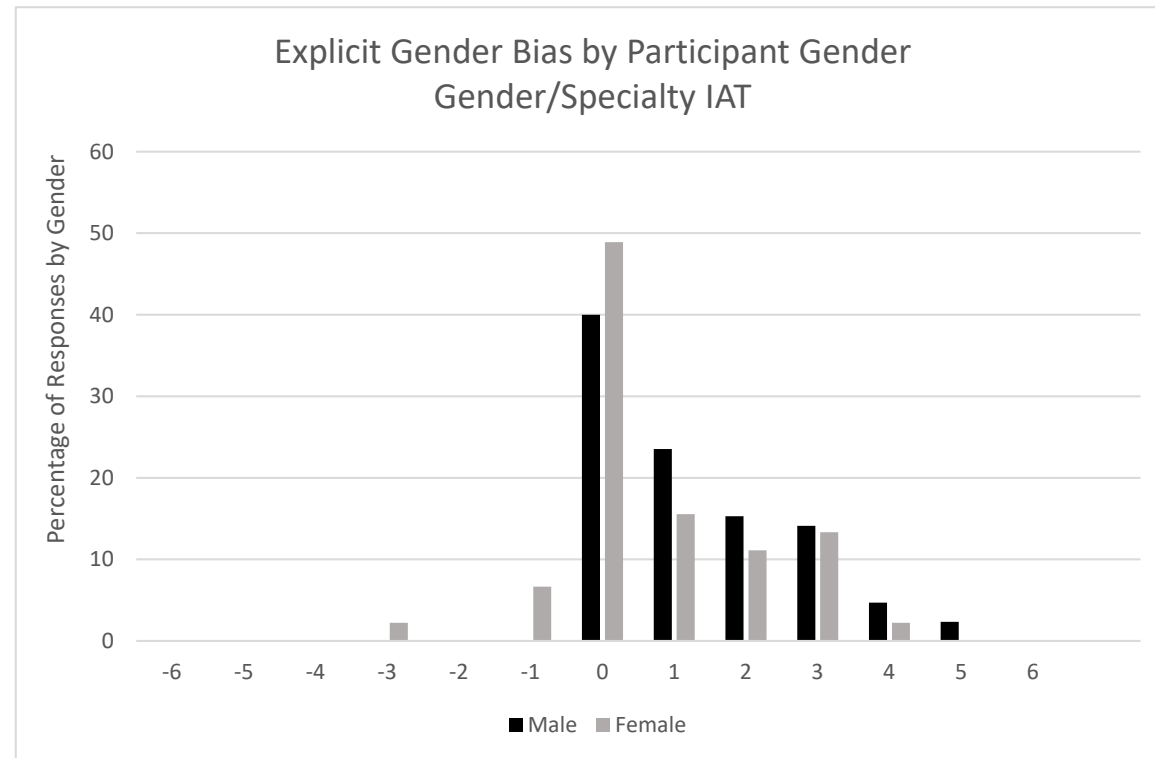
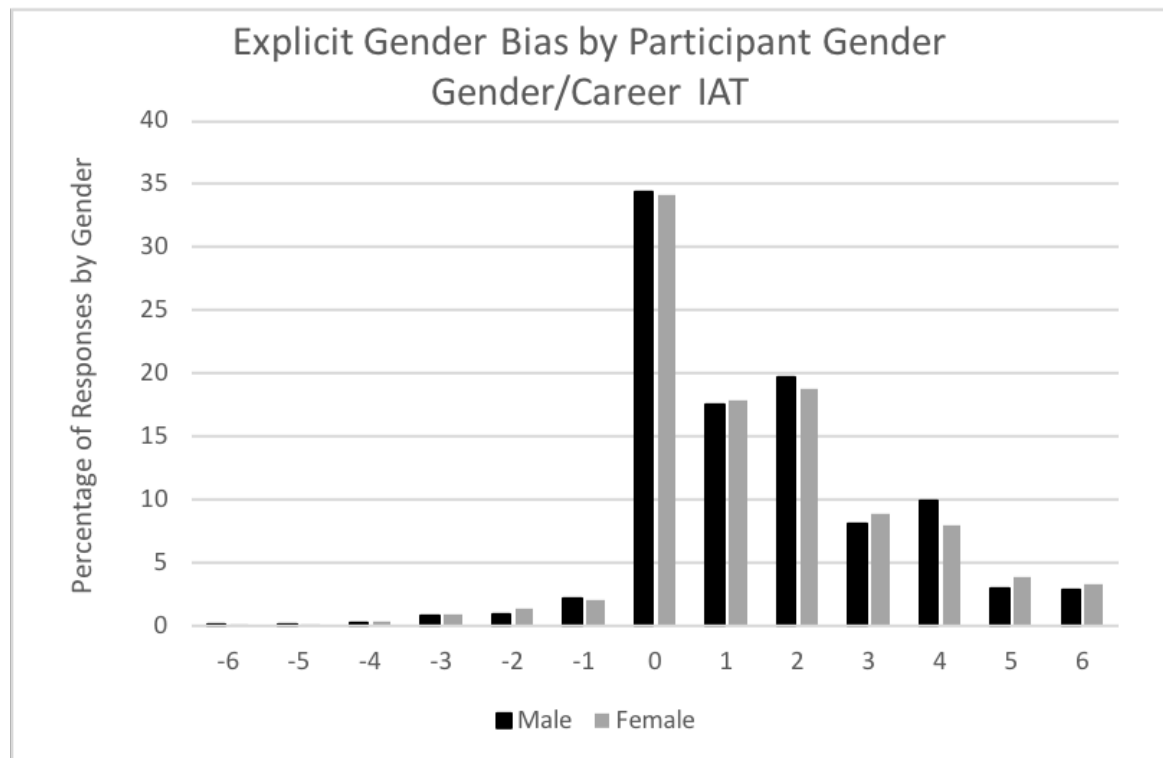
# How do we measure bias?

- Implicit Association Tests
  - [implicit.harvard.edu](http://implicit.harvard.edu)
  - Measure reaction times to assess bias
  - Age, gender, obesity, race, etc.

# Gender/Career, Gender/Specialty



# Gender/Career, Gender/Specialty



# Previous Work

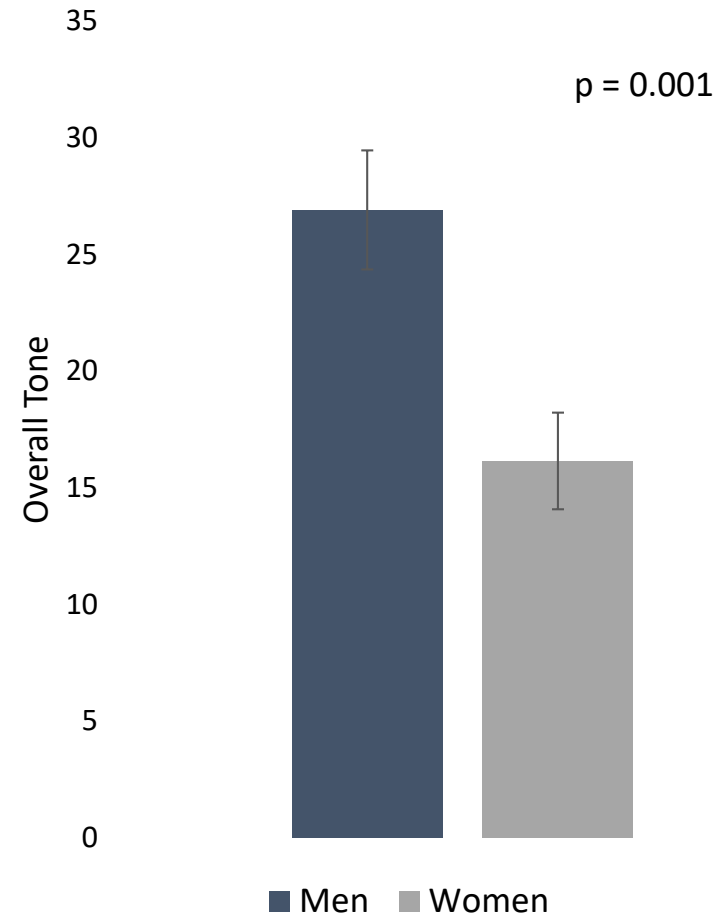
- De-identified, gendered language removed
- Two independent evaluators coded a subset of comments to thematic saturation to determine codes
- Unit of analysis – single idea

# Representative Comments

Theme	Male	Female
<b>Professional Competency</b>	"He is thoughtful and conducts himself professionally...A real gentleman."	"She is very professional and <i>pleasant to deal with.</i> "
<b>Reference to Future</b>	"I predict a <i>great future</i> for him as an academic neurosurgeon/ scientist and <i>leader nationally.</i> "	"She is a great resident. She has <i>potential to succeed</i> as an academic neurosurgeon."
<b>Disposition</b>	" <i>Calm and determined</i> demeanor in difficult situations, <i>pleasant to work with</i> "	"Interacts with everyone in a pleasant demeanor and <i>never seems to get upset or angry</i> "
<b>Job Domains</b>	"Below average for his level of training. Highly motivated to do extra reading on his own time, a quality that should be emulated by other residents."	"Her orthopaedic <i>knowledge base is excellent</i> and she is a <i>motivated and enthusiastic learner.</i> "
<b>Overall Performance</b>	" <i>Terrific</i> resident. Great job! X is <i>a star</i> ... He is performing quite well this year for his level of training."	"X is a <i>good resident</i> . She has a good knowledge base, works hard and thinks about what she is doing."



# Overall Tone

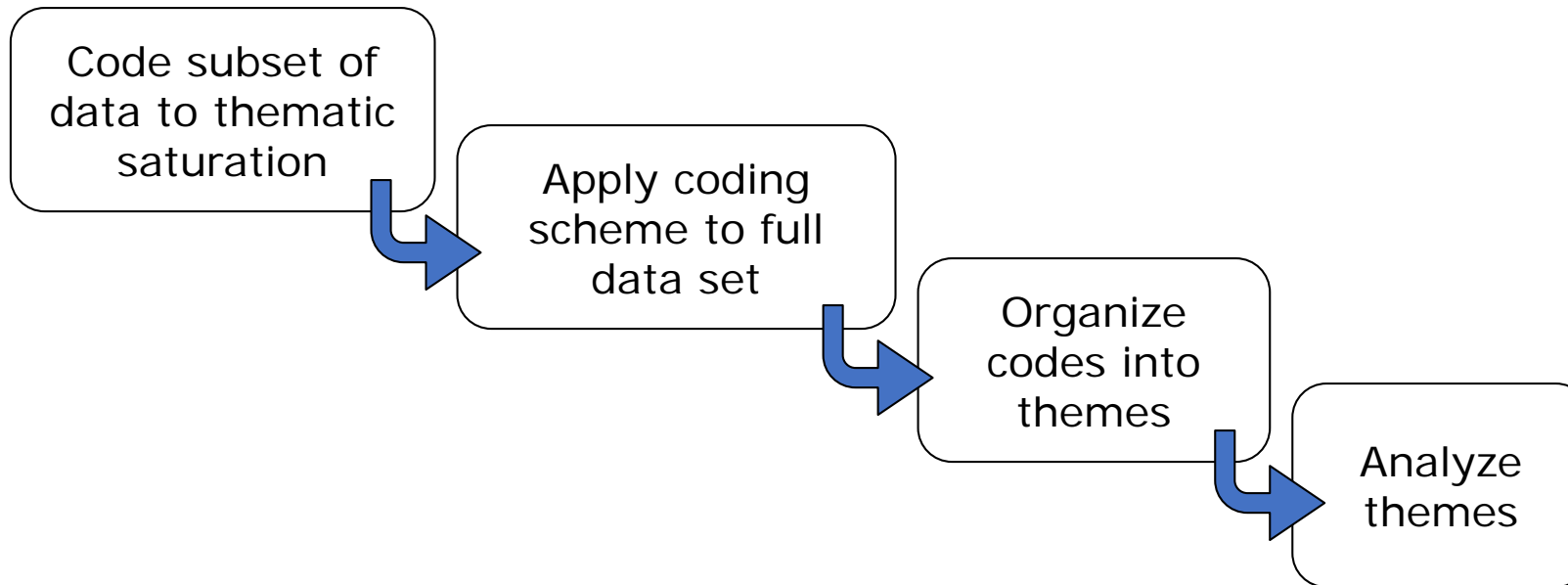


# Current Study

- 590 evaluations
- Of 101 medical students from 2016-2017
- Pediatrics, Obstetrics & Gynecology, Surgery, Medicine

# Methods – Qualitative Analysis

- Inductive thematic analysis to assess written comments



# Methods – Content Analysis

Communal <sup>1</sup>	Grindstone <sup>2</sup>	Standout <sup>2</sup>		Ability <sup>2</sup>
Caring	Conscientious	Exceptional	Dramatically	Intelligent
Kind	Diligent	Best	Extraordinary	Bright
Empathy	Meticulous	Outstanding	Very	Talent
Compassionate	Disciplined	Superb	Tremendous	Expert
Communicate, Communication	Organize, Organization	Excellent	Incredible	Competent
Rapport	Solid	Phenomenal	Stellar	Smart
Cheerful	Hardworking	Star, Superstar	Extremely	Gifted
Pleasant	Dependable	Terrific	Impeccable	Adept
Cooperative	Thorough	Remarkable	Notable, Notably	Skilled
Helpful	Dedicated	Superior	Unique	Analytical
Warm	Careful	Leader	Unparalleled	Proficient
Sympathetic	Reliable	Scholar	Most	Instinct
Teamwork	Responsible	Exemplary	Amazing	Adroit
Team player		Fantastic	Highly	Natural
Thoughtful		Wonderful		Insight
Upbeat				Confident
Joy				
Interpersonal				

<sup>1</sup> Madera et. al, 2009, J Applied Psychology

<sup>2</sup> Trix and Psenka, 2003, Discourse and Society

# Methods – Content Analysis

- Relative Use
  - $R > 1 \rightarrow$  higher use in men's evaluations
  - $R < 1 \rightarrow$  higher use in women's evaluations

$$R = \frac{\left( \frac{\text{Number of Category Words for Men}}{\text{Number of Words for Men}} \right)}{\left( \frac{\text{Number of Category Words for Women}}{\text{Number of Words for Women}} \right)}$$

# Results – Themes

## Entrustability

- Attention to Detail
- Ownership of Patients
- Professionalism
- Responsibility
- Work Ethic

## Growth

- Improvement
- Seeking/Responding to Feedback
- Room for Improvement

## Generic Feedback

- General Performance
- Future Potential

## Medical Skills

- Clinical Skills
- History
- Physical Exam
- Differential Diagnosis
- Plan
- Written Note
- Presenting Patients
- Technical Skills

## Interpersonal Skills

- Communication Skills
- Enthusiasm/ Engagement
- Empathy
- Likeability
- Rapport
- Team Membership
- Thoughtfulness

## Medical Knowledge

- Intelligence
- Fund of Knowledge
- Use of Evidence-Based Medicine

# Results—Representative Comments

Theme	Male	Female
<b>Growth</b>	<p>“Continue to read and be proactive.”</p> <p>“Routinely <i>sought</i> feedback on his performance.”</p>	<p>“Needs to work on <i>confidence/assertiveness</i> moving forward.”</p> <p>“She is definitely <i>receptive</i> to criticism and is willing to listen to feedback.”</p>
<b>Interpersonal Skills</b>	<p>“Compassionate member of our team.”</p> <p>“He was a <i>pleasure</i> to work with.”</p>	<p>“She is also very compassionate and caring for her patients.”</p> <p>“She was a <i>delightful</i> addition to the L&amp;D team.”</p>
<b>Medical Skills</b>	<p>“Good <i>suturing</i> skills in the OR.”</p>	<p>“She was helpful in the OR without <i>getting in the way</i>.”</p> <p>“Very <i>comfortable</i> in the OR.”</p>
<b>Medical Intelligence</b>	<p>“He is very bright and works very well with patients and in the team setting.”</p>	<p>“Bright intellectually motivated and a hard working student.”</p>
<b>Entrustability</b>	<p>“He was respectful, reliable, and worked well with the entire care team.”</p>	<p>“She was diligent, motivated, and eager to learn.”</p>

# Results – Content Analysis

Word Category	Unadjusted Word Use		Adjusted Relative Use+ <i>(proportion for men)</i> <i>(proportion for women)</i>	p-value
	<i>Men</i>	<i>Women</i>		
Ability	225	234	1.06	0.53
Standout	1306	1503	0.96	0.25
Grindstone	319	394	0.89	0.13
<b>Communal</b>	<b>408</b>	<b>524</b>	<b>0.86</b>	<b>0.02</b>



# Results—Grades and Likert Scores

## Grades

<b>Clerkship</b>	<b>Men</b>	<b>Women</b>	<b>p</b>
Medicine	3.63	3.51	0.30
Ob/Gyn	3.51	3.63	0.33
Pediatrics	3.38	3.52	0.16
Surgery	3.43	3.39	0.81

## Likert Scores

<b>Competency</b>	<b>Men</b>	<b>Women</b>	<b>p</b>
Medical Knowledge	4.37	4.43	0.38
Patient Care	4.26	4.30	0.49
Interpersonal Communication	4.31	4.34	0.17
Professionalism	4.45	4.51	0.04
Problem-based learning	4.53	4.62	0.47

# Conclusions

- Qualitative differences:
  - Thematic differences in growth, interpersonal skills, medical skills
  - Communal words more likely to be used to describe women
- No quantitative differences

# Solutions

- Education of faculty, residents, and fellows
- Standardization of language and format
- Promotion of holistic review
- Frame performance as promising

# Questions?

[marenloe@wustl.edu](mailto:marenloe@wustl.edu)

[arghavan@Stanford.edu](mailto:arghavan@Stanford.edu)

# Content Analysis: Standout Words

Word	Unadjusted Word Use		Adjusted Relative Use <sup>+</sup> <i>(proportion for men)</i> <i>(proportion for women)</i>	p-value
	Men	Women		
Excellent	169	72	1.29	0.07
Outstanding	69	25	1.52	0.08
<b>Leader</b>	<b>28</b>	<b>5</b>	<b>3.09</b>	<b>0.02</b>
Best	15	11	0.75	0.54
Exceptional	13	6	1.19	0.81
Superb	13	4	1.79	0.45
Star/Superstar	7	2	1.93	0.51
Terrific	8	1	4.41	0.17
Superior	7	0	*	*
Scholar/Scholarly	5	0	*	*
Phenomenal	1	1	0.55	1.00
Fabulous	0	2	*	*
<b>Total</b>	<b>335</b>	<b>129</b>	<b>1.43</b>	<b>&lt;0.01</b>

# Results – Content Analysis

Word Category	Unadjusted Word Use		Adjusted Relative Use+ <i>(proportion for men)</i> <i>(proportion for women)</i>	p-value
	Men	Women		
Ability	54	21	1.42	0.19
Grindstone	71	34	1.15	0.54
Communal	68	46	0.81	0.28