Information Bias and Selection of Female Professors

Stephanie El Khoury

Arizona State University

January 3rd, 2026

Ratings, Information Bias, and Professor Choice

- Ratings guide decisions
  - Healthcare: physician-rating sites (Bensnes and Huitfeldt, 2021; Chen, 2018)
  - Gig-Economy Platforms (Uber, DoorDash, Airbnb)
  - Higher education: teaching evaluations and professor review websites (PRWs)

    (Saygin and Zhang, 2025; Choshi et al., 2023; Yu et al., 2021)

Ratings, Information Bias, and Professor Choice

- Ratings guide decisions
  - Healthcare: physician-rating sites(Bensnes and Huitfeldt, 2021; Chen, 2018)
  - Gig-Economy Platforms (Uber, DoorDash, Airbnb)
  - Higher education: teaching evaluations and professor review websites (PRWs)

    (Saygin and Zhang, 2025; Choshi et al., 2023; Yu et al., 2021)
- Information bias: ratings can deviate from true quality
  - Manipulation, selective reviewing, and group-specific discrimination (Mayzlin et al., 2014; Hu et al., 2017; Askalidis et al., 2017; He et al., 2022; MacNell et al., 2015)

Ratings, Information Bias, and Professor Choice

- Ratings guide decisions
  - Healthcare: physician-rating sites(Bensnes and Huitfeldt, 2021; Chen, 2018)
  - Gig-Economy Platforms (Uber, DoorDash, Airbnb)
  - Higher education: teaching evaluations and professor review websites (PRWs) (Saygin and Zhang, 2025; Choshi et al., 2023; Yu et al., 2021)
- Information bias: ratings can deviate from true quality
  - Manipulation, selective reviewing, and group-specific discrimination (Mayzlin et al., 2014; Hu et al., 2017; Askalidis et al., 2017; He et al., 2022; MacNell et al., 2015)
- **Research questions:** How do information biases on PRW impact students' beliefs and their enrollment choices?
  - Gender penalties in academia (Antecol et al., 2018; Wu, 2018; Dupas et al., 2021; Sarsons et al., 2021)
  - Gender bias in teaching evaluations (Boring, 2017; MacNell et al., 2015; Binderkrantz and Bisgaard, 2023; Boring et al., 2016; Fernandes et al., 2023)

Research Question

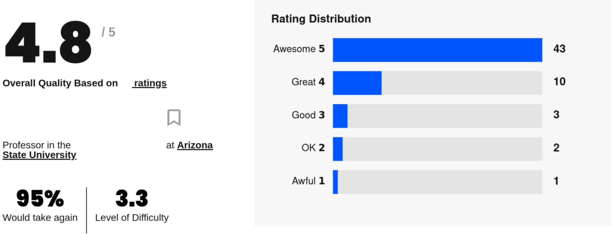
- Does an information bias against women exist in evaluation websites?
- Do students value reviews on evaluation websites and are they correcting for the existence of the bias?
- Can the information bias be corrected using an information intervention?

Literature and Contribution

Table of Contents

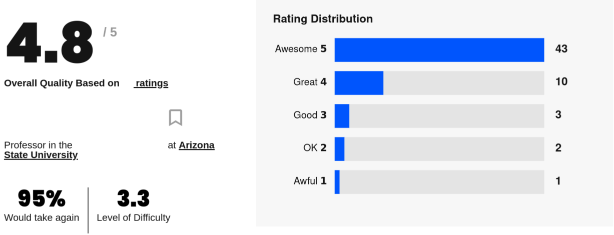
Table of Contents

Information: RateMyProfessor and the Gender Gap



Quality/Difficulty Definitions

Information: RateMyProfessor and the Gender Gap



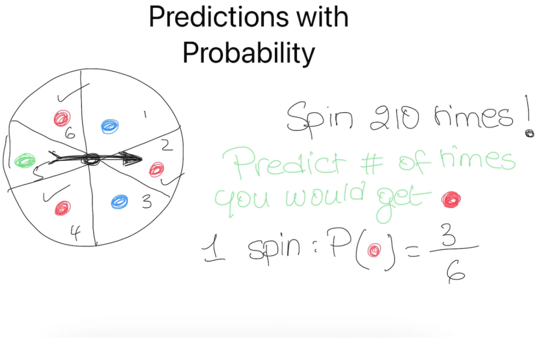
	Quality Rating	
	ASU Business School	Other Business Schools
Female Prof.	-0.128**	-0.168***
	(0.058)	(0.028)
Difficulty Rating	-0.660***	-0.736***
	(0.043)	(0.017)
Observations	1202	3781

Other Business schools include: CSU-Fullerton, FIU, MSU, GSU, WSU, UofAlabama, CSU. Uni. Level Regs

Quality/Difficulty Definitions

RCT #1: Do students rate female instructors lower?

- Online survey experiment; students randomly assigned a *male* or *female* voice-over teaching the same 2-minute probability lesson.
- No visual cues; identical script, age, pacing, and slides; professional voice actors with neutral North American accent.



RCT #1: Do students rate female instructors lower?

- Online survey experiment; students randomly assigned a *male* or *female* voice-over teaching the same 2-minute probability lesson.
- No visual cues; identical script, age, pacing, and slides; professional voice actors with neutral North American accent.

	Quality Rating
Female Prof $\times$ Has NOT Rated on PRWs	-0.061
	(0.062)
Female Prof $\times$ Has Rated on PRWs	-0.238***
	(0.054)
Observations	288

Table of Contents

Survey 1: Baseline Measurement

Timing:

- Administered at the start of the semester.

Content:

- **General Information:** demographics, preferences for professor characteristics, interactions with professor review websites and professors in the classroom, and typical costs of core courses and textbooks at ASU
- **Discrete Choice Experiment (BACE method)** ***[Drake et al., 2023]***:
  - 10 hypothetical professor profiles ⏩ Example
- **RCT #2:**
  - **Control:** 2-min neutral campus-activities video
  - **Treatment:** 2-min video on gender bias in PRW ratings ▸ Video

    → Females are sometimes rated harsher

    → Refers to literature (MacNell et al., 2015)

    → Shows gender gap on RMP in Economics Departments

Survey 2: Measuring Persistence of Treatment Effects

Timing:

- Administered **3 weeks after** Survey 1 and RCT #2

Purpose:

- Test whether students **internalized treatment information**

Table of Contents

Scenario Results: Willingness-to-Pay (USD)

	WTP
Female Professor	4
	(25)
Quality Rating	281***
	(20)
Difficulty Rating	531***
	(82)
(Difficulty Rating)^2	-123***
	(15)
Observations	1863

Scenario Results: Willingness-to-Pay (USD)

	WTP
Female Professor	4
	(25)
Quality Rating	281***
	(20)
Difficulty Rating	531***
	(82)
(Difficulty Rating)^2	-123***
	(15)
Observations	1863

Students value better rated professors

Scenario Results: Willingness-to-Pay (USD)

	WTP
Female Professor	4
	(25)
Quality Rating	281***
	(20)
Difficulty Rating	531***
	(82)
(Difficulty Rating)^2	-123***
	(15)
Observations	1863

Variable	Mean
Usefulness (Very, Extremely)	0.67
Often Usage (A lot, Always)	0.71
Percentage left a review	0.50
Use PRW in making decision	0.81
Ask friends for information	0.54

Scenario Results: Willingness-to-Pay (USD)

	WTP
Female Professor	4
	(25)
Quality Rating	281***
	(20)
Difficulty Rating	531***
	(82)
(Difficulty Rating)^2	-123***
	(15)
Observations	1863

Students are unaware of the existence of the bias

Scenario Results: Willingness-to-Pay (USD)

	WTP
Female Professor	4
	(25)
Quality Rating	281***
	(20)
Difficulty Rating	531***
	(82)
(Difficulty Rating)^2	-123***
	(15)
Observations	1863

Variable	Mean
Belief Size of the Bias	-0.017
	(0.017)

Table of Contents

Impact of Treatment on WTP, Beliefs, and Bias

	WTP Female$_{t+1}$	Size of Bias$_{t+1}$
Treatment	176.764**	-0.137***
	(68.755)	(0.027)
Observations	1419	1419

- **WTP:** Treatment increases willingness-to-pay for female instructors by $\sim$ $177 (10% of average course + textbook cost)

Impact of Treatment on WTP, Beliefs, and Bias

	WTP Female$_{t+1}$	Size of Bias$_{t+1}$
Treatment	176.764**	-0.137***
	(68.755)	(0.027)
Observations	1419	1419

- **WTP:** Treatment increases willingness-to-pay for female instructors by $\sim$ $177 (10% of average course + textbook cost)
- **Size of Bias:** Significant 0.14 point increase in perceived bias against female professors in PRW ratings

Table of Contents

Follow-Up Survey

Timing:

- Around **2-3 months** after the launch of the first core survey
- **After course enrollment** opens for the next semester

Purpose:

- Look at **long-run effects** of information intervention: rating behavior and course selection in subsequent semester

Tasks:

- Re-run RCT #1
- Ask questions about total courses taken and number taken with female professors

Impact of Treatment on Rating Behavior: Students Who Rated

	Quality Rating
Female Prof. $\times$ Not Treated	-0.402***
	(0.115)
Female Prof. $\times$ Treated	0.138
	(0.116)
Observations	174

Difference between the coefficients in Row 1 and Row 2 is statistically significant at the 1% level (t-stat = 3.96).

- Among untreated students, female professors receive lower ratings
- Among treated students, the penalty disappears
- The treatment closes the gender gap in ratings among students who rate

▸▸ All Results

Impact of Treatment on Courses Selection

	Courses Next Semester with Fem.
Treatment	0.323*
	(0.188)
Observations	134
IV (Receive Treatment)	X
Dependent Var. Mean (Control)	1.75

Estimating Eq.

Conclusion

Main Findings:

- Students value PRW ratings and use them to guide enrollment decisions
- However, they are largely unaware that these ratings are biased

Low-Cost Intervention:

- Increases student **WTP for female professors by 10%** of the course cost
- Students correct their beliefs about the existence of a bias against female instructors on PRWs
- Reduces the existence of a bias in student ratings among students that engage in rating on PRWs
- **Boosts enrollment with women** by 18 % for the subsequent semester for students that find PRWs useful

Gender Gap by Business Schools

	Female Coef.	Std. Error	Observations
Wayne State University	-0.081**	(0.040)	588
Florida International University	-0.139*	(0.080)	678
Georgia State University	-0.299***	(0.080)	591
Michigan State University	-0.307***	(0.112)	227
University of Alabama	-0.220***	(0.050)	761
California State University, Fullerton	-0.091*	(0.052)	936

Back

Literature and Contribution

- Literature on **statistical discrimination** (Aigner and Cain, 1977; Phelps, 1972; Bohren et al., 2025)
  - Statistical discrimination that stems from limited awareness of bias in signals
  - I bring this mechanism into the context of higher education by documenting that students

- Literature on **information frictions in educational choice** (Altonji, 1993; Arcidiacono, 2004; Stange, 2012; Stinebrickner and Stinebrickner, 2012, 2014; Hoxby and Turner, 2015; Thomas, 2019; Arcidiacono et al., 2020; Larroucau and Rios, 2022; Wiswall and Zafar, 2015)
  - I extend this insight to the context of instructor choice, showing that biased peer evaluations distort students' beliefs and sub-optimal decisions

Back

Literature and Contribution

- Literature on **online platforms and information design** (Luca, 2011; Jin and Sorensen, 2006; Chevalier and Mayzlin, 2006)
  - This paper brings these insights into the context of higher education
- Literature on **gender bias in evaluations of higher education** (Boring, 2017; Saygin and Zhang, 2025; MacNell et al., 2015; Binderkrantz and Bisgaard, 2023; Boring et al., 2016; Fernandes et al., 2023)
  - Bias is particularly pronounced among students who actively engage with professor review platforms
  - Biased signals distort students' beliefs and enrollment choices

Back

What is Quality?

Quality	
High	Low
Easy to understand	Hard to understand/follow
Cares about students	Tough grader
Accessible outside class	Waste of time
Gives great feedback	Does not respond to emails
Passionate	Boring
Helpful	Doesn’t care
Extra credit	I didn’t learn
Wants students to succeed	A lot of work

Back

What is Difficulty?

Difficulty	
High	Low
Hard to pass	Willing to help
Hard to understand material	Accessible outside class
Be prepared to work	Passionate
Makes you work for your grade	Easy to get good grade
A lot of homework	Wants students to succeed
Waste of time	Extra credit
Does not care	Puts in the effort

Back

BACE and Scenario Example

- BACE is a dynamic choice experiment method that elicits preferences efficiently
- Menu is chosen using the information provided by the subjects' previous choices

Professor A	**Professor B**
Textbooks in USD: \$128.2	Textbooks in USD: \$261.2
Gender: Female	Gender: Female
Ratings from Review Websites: Quality: 1.9/5 Difficulty: 3.4/5	**Ratings from Review Websites:** Quality: 1.4/5 Difficulty: 2.3/5

Back

Impact of Treatment on Rating

	Quality Rating
Has NOT Rated and Not Treated	
(R1) Female Prof.	0.020
	(0.491)
Has Rated and Not Treated	
(R2) Female Prof.	-0.402^{***}
	(0.115)
Has NOT Rated and Treated	
(R3) Female Prof.	0.617^{*}
	(0.244)
Has Rated and Treated	
(R4) Female Prof.	0.138
	(0.116)
Observations	174

Coefficient from (R2) is statistically different than the coefficient in (R4) at the 1% significance level with a t-statistic of 3.96.

Back

Estimating Equations (Course Selection)

First Stage:

$$T_i = \pi_0 + \pi_1 A_i + X_i' \pi_4 + v_{1i}, \tag{1}$$

Second Stage:

$$\begin{aligned} Courses\ w\ Fem_{i,t+1} &= \omega_0 + \omega_1 \widehat{T}_i + \omega_4 Courses\ w\ Fem_{i,t} \\ &+ \omega_5 Total\ Courses_{i,t} + \omega_6 Total\ Courses_{i,t+1} + X_i' \omega_7 + \varepsilon_i \end{aligned} \tag{2}$$

Back

Balance Test: RCT #1

	Diff (C-T)	SE	t-stat	p-value
Minority Status	0.008	0.059	0.15	0.895
Female Student	0.015	0.059	0.25	0.797
Year at ASU	0.142	0.130	1.10	0.279
Economics Major	0.031	0.035	0.90	0.371
Obs	288			

Balance Test: RCT #2

Variable	Diff. (C-T)	SE	*p*-value
WTP Female	73.166	57.550	0.204
WTP Quality	13.864	33.845	0.682
WTP Difficulty	-3.383	184.591	0.986
WTP Difficulty^2	24.16	50.008	0.629
Household Income	0.072	0.183	0.692
% Review on PRW	0.007	0.015	0.600
Minority	0.018	0.026	0.492
Female Student	0.009	0.026	0.736
Attention Check 1	-0.007	0.011	0.502
Attention Check 2	0.005	0.022	0.839
Age	0.111	0.144	0.442
Hours Study per Week	0.017	0.254	0.947
First Generation	0.034	0.024	0.146
Avg. Quality Rating Male ($\hat{\mu}_{RM}$)	-0.041	0.031	0.193
Avg. Quality Rating Female ($\hat{\mu}_{RF}$)	-0.009	0.035	0.806
Avg. Quality Male ($\hat{\mu}_{M}$)	-0.081	0.037	0.029
Avg. Quality Female ($\hat{\mu}_{F}$)	0.007	0.037	0.857
Useful PRW	-0.043	0.051	0.396
Total Observations		1,419	

Avg. Quality of Male has a $p - value < 0.05$. However, when running a normalized difference test following Imbens and Rubin (2015), the normalized difference between the treatment and control group (0.059) is less than 0.25, which provides evidence that this difference is not concerning to our analysis.

Frame Title

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Ilr Review*, 30(2):175–187.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, 11(1):48–83.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, 128(12):4475–4522.

Askalidis, G., Kim, S. J., and Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, 97:23–30.

Bensnes, S. S. and Huitfeldt, I. M. S. (2021). Rumor has it: How do patients respond to patient-generated physician ratings? *Journal of Health Economics*, 76:102415.

Binderkrantz, A. S. and Bisgaard, M. (2023). A gender affinity effect: The role of gender in teaching evaluations at a danish university. *Higher Education*.