Assessing Gender Bias in Educational Videos

Elizabeth Khan, Estrella Ndrianasy, Chandni Shah, Michelle Shen, Catherine Tsai

W241 Fall 2021

Berkeley
UNIVERSITY OF CALIFORNIA

Agenda

1. **Research Question**
   - Background
   - Hypothesis
2. **Data**
   - Recruitment & reasoning behind recruitment
3. **Experimental Design / Methods**
   - Overview
   - Outcome Measure
4. **Results**
   - Randomization / Covariate Balance
   - Analysis
5. **Limitations**
6. **Discussion / Conclusion**

Research Question

- **Background:**
  - Studies have concluded female instructors face challenges with gender stereotypes

- **Research Questions:**
  - **Primary:** Does an instructor's perceived gender influence ***perceived quality of instruction***?
  - **Secondary:** Does an instructor's perceived gender influence ***retention of content***?

- **Hypothesis:**
  - Treatment of changing perceived gender of instructor will impact measured instructor ratings and content retention
  - Expect primary and secondary outcomes measures to ***decrease*** with treatment of female-perceived instructor

Data

- Requirements for subject recruitment company:
  - Administer Qualtrics survey to subjects
  - Record subject responses
  - Ensure sufficient sample size
  - Enable blocked design
  - Include inclusion & exclusion criteria
  - Disqualify those who didn't pass attention check
  - Deliver responses to us in a timely manner

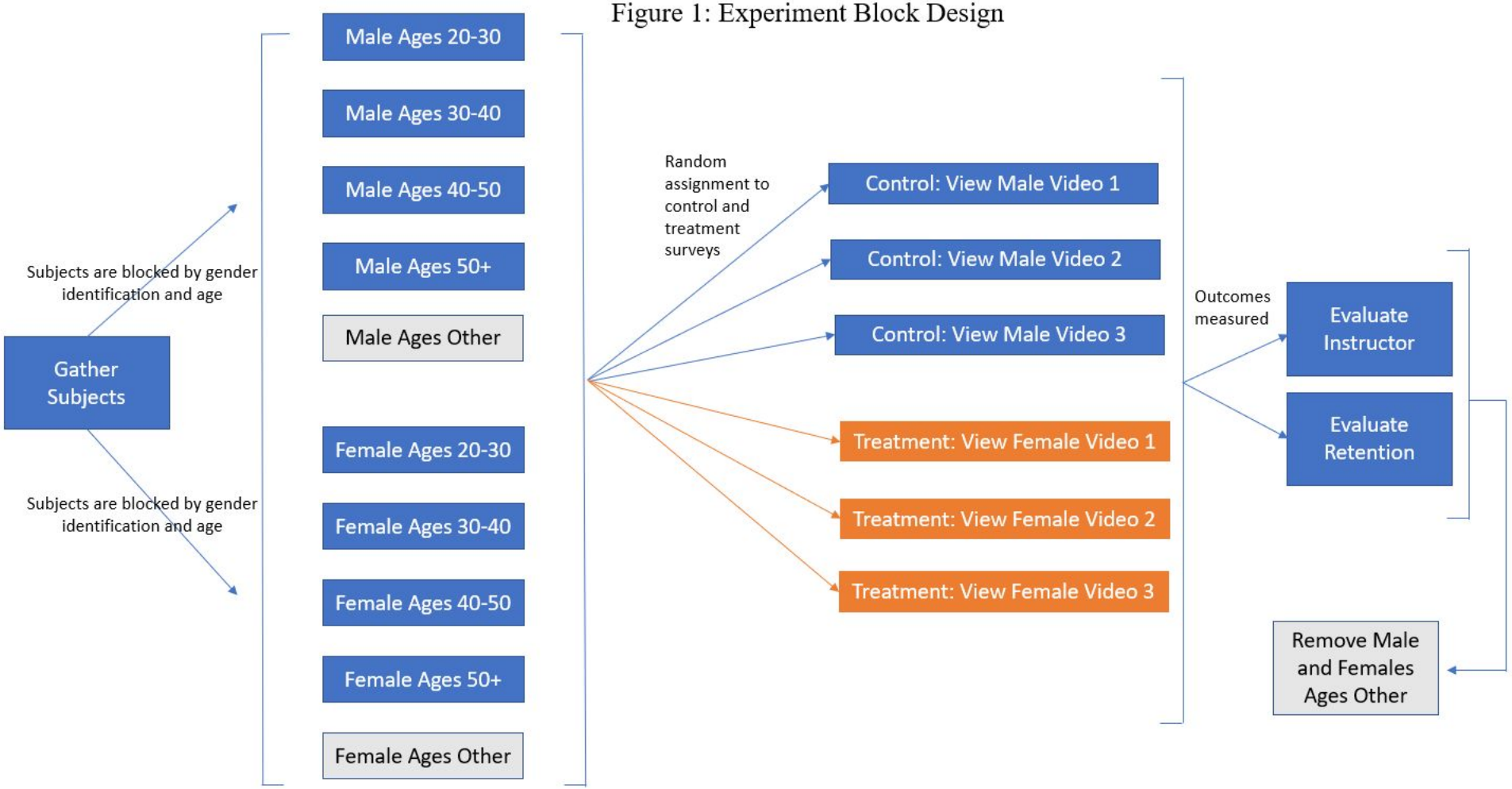
- Social science recruitment company, SurveySwap, was the best fit

Experimental Design

- 2x4 **blocked design** on subjects' gender and age groups
- Power test led to **sample size of n=112**
- **Inclusion criteria:** located in United States and native English speakers
- **Exclusion criteria:** Subjects younger than 20 years old or whose age was not discernable
- **Control group:** received survey with male voice over video followed by outcome-related questions
- **Treatment group:** received survey with female voice over video followed by outcome-related questions
- **Primary outcome:** subject perception of instructor
- **Secondary outcome:** information retention

Experimental Design

Figure 1: Experiment Block Design



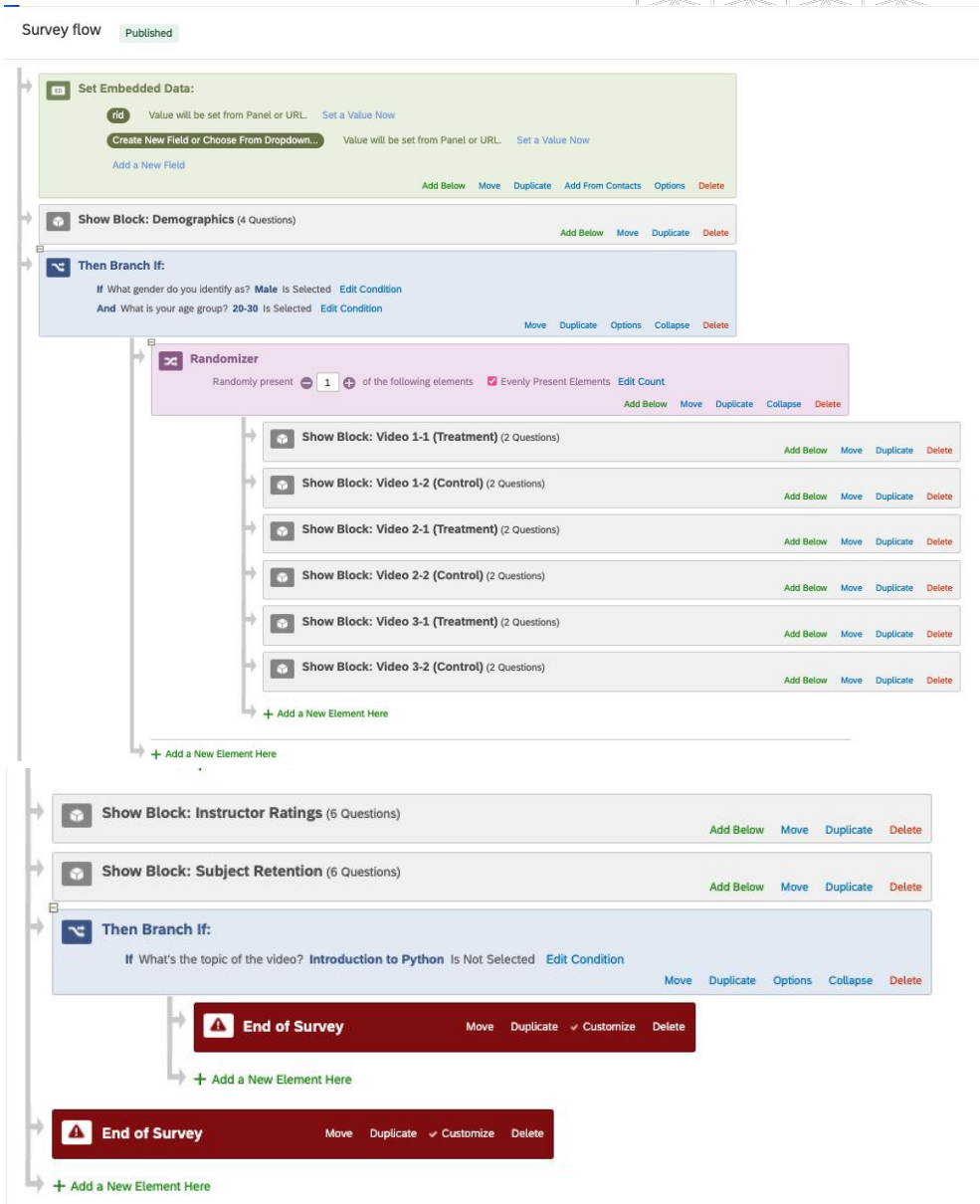
Methods

- Created a 3 min 10 second long introductory Python video
- 3 men and 3 women recorded voice overs using a script with timestamps
- Made 1 questionnaire with 13 questions for demographics, outcomes of interest, and attention check
- Combined into 6 separate surveys on Qualtrics; only difference was the video
  - Added logic
- Survey Swap administered Qualtrics survey to users and sent us results

Survey Creation

How would you rate the instructor's enthusiasm?

1 - not enthusiastic
2 - slightly enthusiastic
3 - moderately enthusiastic
4 - very enthusiastic
5 - extremely enthusiastic



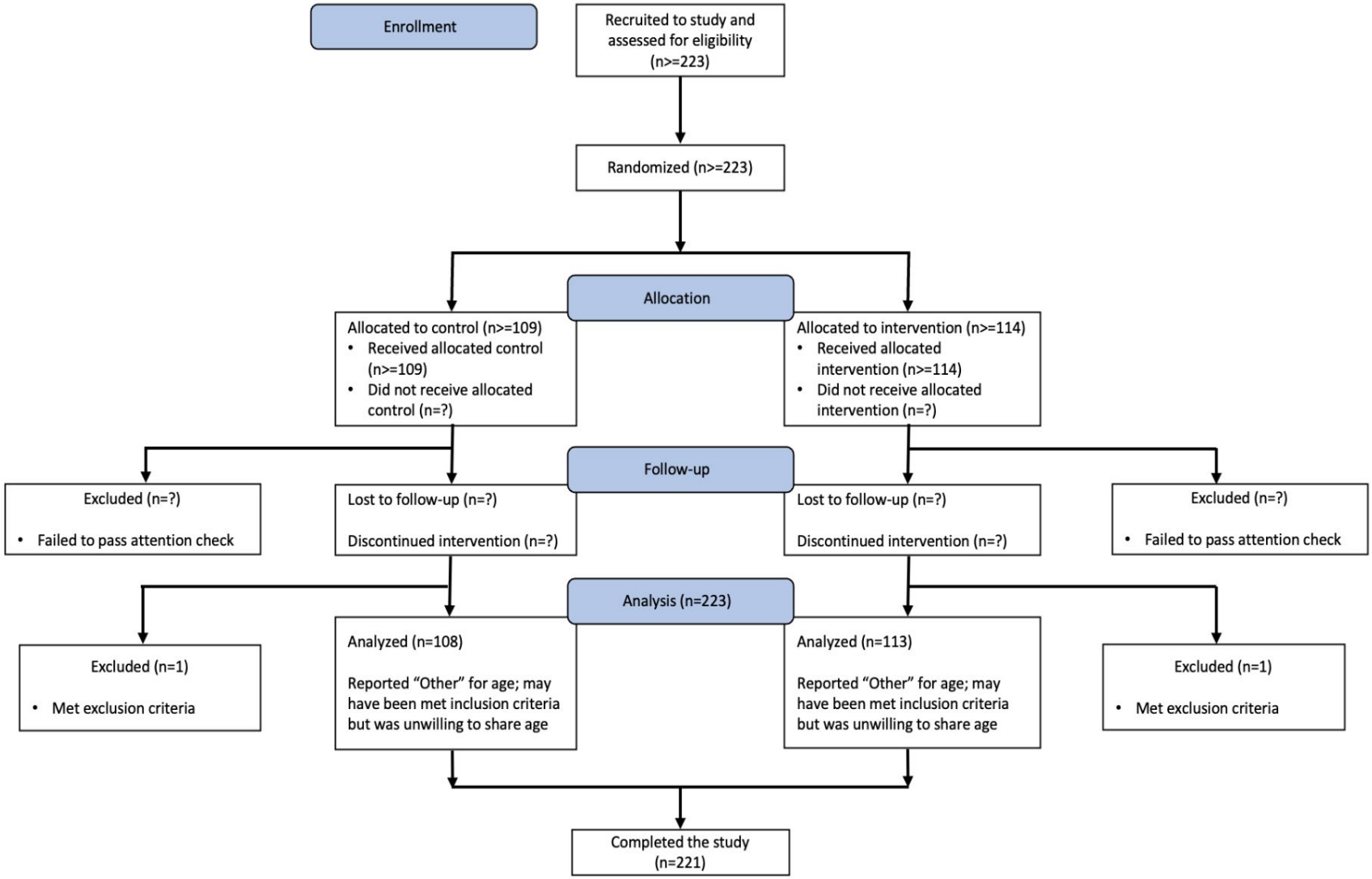
Berkeley
UNIVERSITY OF CALIFORNIA

Introduction to Python Video

Control - Male Instructor Voice (3 versions)

Treatment - Female Instructor Voice (3 versions)

Figure 2: Consort



Results

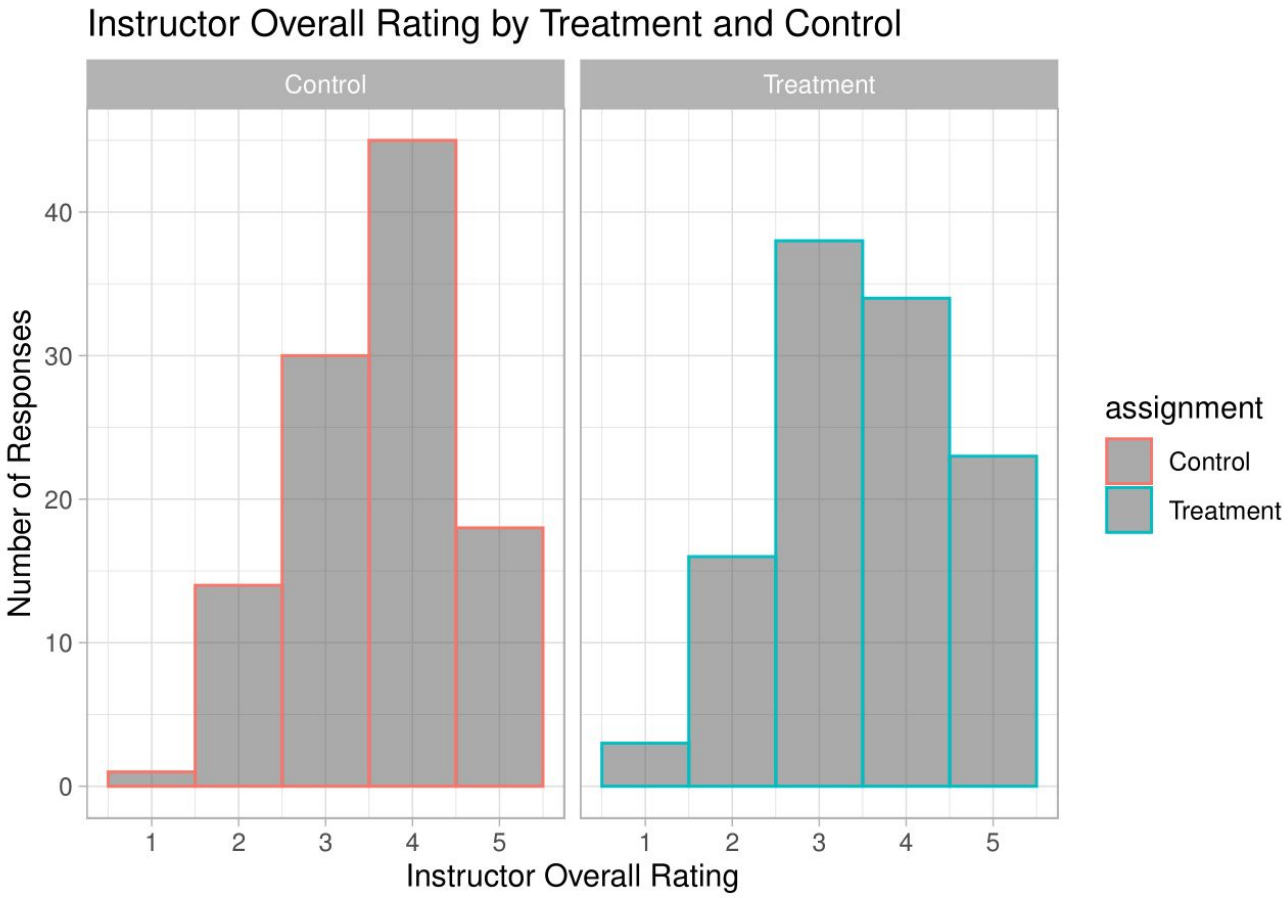
- Overall Respondents
  - 221 Total: Female - 142, Male - 79

Table 1: Respondents by Age Group and Gender

Characteristic	**Overall**, N = 221	**Female**, N = 142	**Male**, N = 79
assignment			
Control	108 (49%)	68 (48%)	40 (51%)
Treatment	113 (51%)	74 (52%)	39 (49%)
Age			
20-30	44 (20%)	25 (18%)	19 (24%)
30-40	82 (37%)	56 (39%)	26 (33%)
40-50	48 (22%)	32 (23%)	16 (20%)
50+	47 (21%)	29 (20%)	18 (23%)

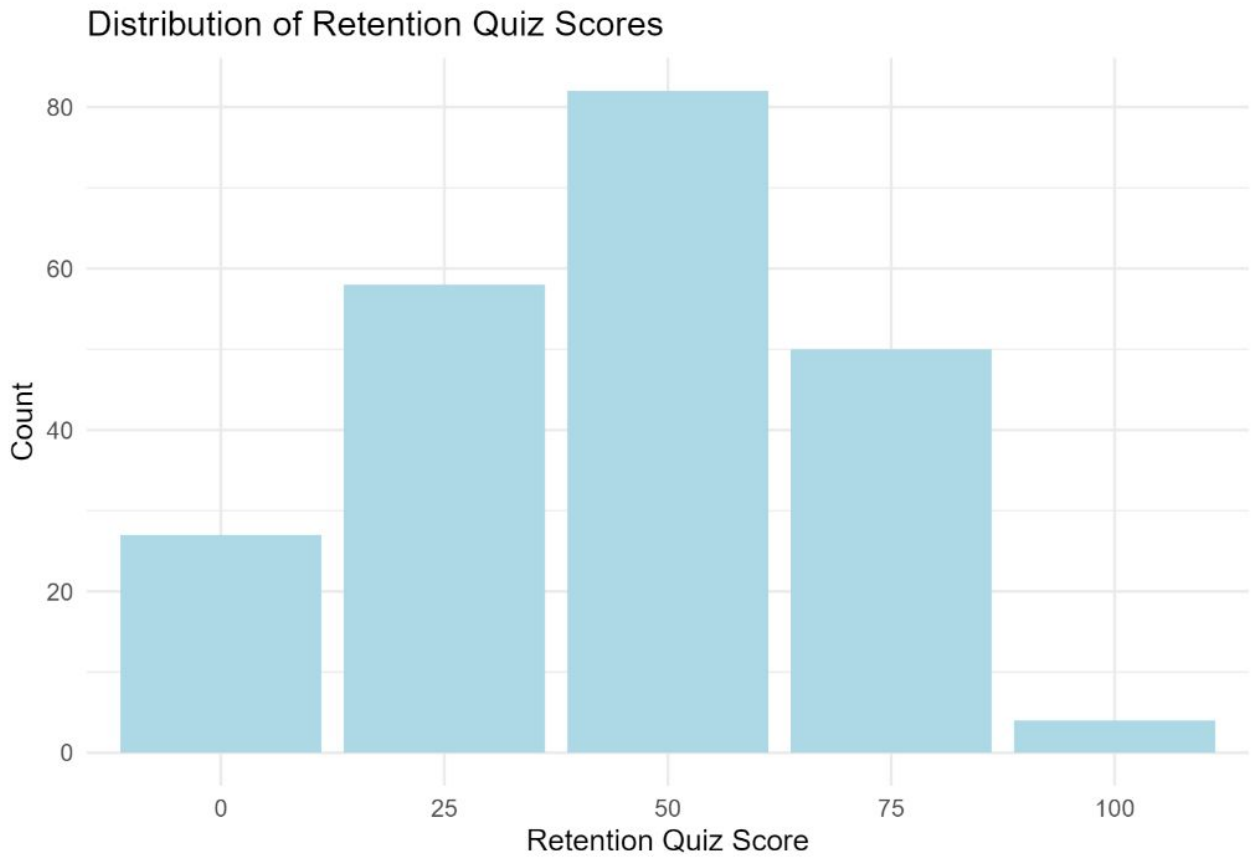
Results

- Distribution of the Instructor Ratings
  - Distributions are similar - mean treatment 3.60 vs. control 3.50
    - Mode: Control 4, Treatment: 3



Results

- Despite the attention check (1st question in the “Subject Retention” section)
  - Median score is 50 (2 out of 4 questions answered correctly)



Berkeley
UNIVERSITY OF CALIFORNIA

Covariate Balance

Ran a covariate balance check on Gender, Age, and Education

- P-values indicate no significant difference between control and treatment groups
- Confirmed randomization - treatment assignment was successful

Table 3: Covariate Balance Test

	Control (N = 108)	Treatment (N = 113)	Mean - Control	Mean - Treatment	t-test (p-value)
Gender					
Male	40 (37.04%)	39 (34.51%)	0.37	0.35	0.697
Female	68 (62.96%)	74 (65.49%)	0.63	0.65	0.697
Age					
20-30	23 (21.30%)	21 (18.58%)	0.21	0.19	0.616
30-40	39 (36.11%)	43 (38.05%)	0.36	0.38	0.766
40-50	26 (24.07%)	22 (19.47%)	0.24	0.19	0.41
50+	20 (18.52%)	27 (23.89%)	0.19	0.24	0.33
Education					
Less than High school	1 (0.93%)	6 (5.31%)	0.01	0.05	NA
High school diploma	33 (30.56%)	33 (29.20%)	0.31	0.29	0.827
Some College No degree	32 (29.63%)	35 (30.97%)	0.3	0.31	0.829
Associates degree	16 (14.81%)	11 (9.73%)	0.15	0.1	0.253
Bachelors degree	21 (19.44%)	22 (19.47%)	0.19	0.19	0.996
Masters degree	5 (4.63%)	6 (5.31%)	0.05	0.05	0.817

Linear Regression Approach

- Model 1: Simple Model

$$Outcome = \beta_0 + \beta_1 MaleInstructorVideo$$

- Model 2: Blocks on Age Group and Gender

$$Outcome = \beta_0 + \beta_1 MaleInstructorVideo \ + \beta_2 Age:30-40 \ + \beta_3 Age:40-50 \ + \beta_4 Age:50 + \beta_5 Male$$

- Model 3: Interaction Terms between Gender and Treatment

$$Outcome = \beta_0 + \beta_1 MaleInstructorVideo \ + \beta_2 Age:30-40 \ + \beta_3 Age:40-50 \ + \beta_4 Age:50 + \beta_5 Male - \beta_6(Male * MaleInstructorVideo)$$

Linear Regression Results

Primary Outcome: Overall Instructor Effectiveness

	Dependent variable:		
	Overall Instructor Effectiveness Rating		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	−0.097	−0.085	−0.097
	(0.135)	(0.138)	(0.178)
	p = 0.472	p = 0.537	p = 0.585
Age: 30-40		−0.109	−0.108
		(0.199)	(0.200)
		p = 0.587	p = 0.590
Age: 40-50		0.097	0.100
		(0.214)	(0.217)
		p = 0.650	p = 0.645
Age: 50+		0.003	0.004
		(0.211)	(0.212)
		p = 0.988	p = 0.985
Male		0.237	0.219
		(0.138)*	(0.188)
		p = 0.086	p = 0.244
Male:Male Instructor Video			0.035
			(0.279)
			p = 0.901
Baseline	3.600	3.530	3.540
	(0.092)***	(0.185)***	(0.191)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.002	0.022	0.022
Adjusted R^2	−0.002	−0.001	−0.005
Residual Std. Error	1.000 (df = 219)	1.000 (df = 215)	1.000 (df = 214)
F Statistic	0.521 (df = 1; 219)	0.962 (df = 5; 215)	0.800 (df = 6; 214)

Primary Outcome: Instructor Professional Rating

	Dependent variable:		
	Instructor Professional Rating		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	0.045	0.062	0.035
	(0.124)	(0.125)	(0.162)
	p = 0.716	p = 0.623	p = 0.829
Age: 30-40		−0.308	−0.307
		(0.165)*	(0.165)*
		p = 0.062	p = 0.063
Age: 40-50		−0.169	−0.164
		(0.183)	(0.185)
		p = 0.355	p = 0.375
Age: 50+		−0.349	−0.347
		(0.178)*	(0.179)*
		p = 0.051	p = 0.053
Male		−0.019	−0.057
		(0.127)	(0.192)
		p = 0.879	p = 0.768
Male:Male Instructor Video			0.075
			(0.260)
			p = 0.774
Baseline	3.780	4.000	4.010
	(0.094)***	(0.143)***	(0.155)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.001	0.020	0.021
Adjusted R^2	−0.004	−0.002	−0.007
Residual Std. Error	0.914 (df = 219)	0.914 (df = 215)	0.916 (df = 214)
F Statistic	0.135 (df = 1; 219)	0.899 (df = 5; 215)	0.760 (df = 6; 214)

Note: *p<0.1; **p<0.05; ***p<0.01
Note: Uses Robust Standard Error

Berkeley
UNIVERSITY OF CALIFORNIA

Linear Regression Results

Primary Outcome: Instructor Knowledge Rating

	Dependent variable:		
	Instructor Knowledge Rating		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	0.068	0.069	0.022
	(0.123)	(0.124)	(0.160)
	p = 0.579	p = 0.577	p = 0.889
Age: 30-40		−0.212	−0.210
		(0.184)	(0.185)
		p = 0.250	p = 0.257
Age: 40-50		−0.049	−0.040
		(0.188)	(0.190)
		p = 0.797	p = 0.834
Age: 50+		0.064	0.067
		(0.179)	(0.179)
		p = 0.722	p = 0.709
Male		0.108	0.042
		(0.125)	(0.172)
		p = 0.389	p = 0.808
Male:Male Instructor Video			0.132
			(0.253)
			p = 0.603
Baseline	3.860	3.900	3.920
	(0.086)***	(0.172)***	(0.176)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.001	0.021	0.022
Adjusted R^2	−0.003	−0.002	−0.005
Residual Std. Error	0.907 (df = 219)	0.907 (df = 215)	0.909 (df = 214)
F Statistic	0.311 (df = 1; 219)	0.913 (df = 5; 215)	0.802 (df = 6; 214)

Primary Outcome: Instructor Enthusiasm Rating

	Dependent variable:		
	Enthusiasm Rating		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	0.364	0.369	0.409
	(0.153)**	(0.154)**	(0.193)**
	p = 0.018	p = 0.017	p = 0.035
Age: 30-40		0.043	0.041
		(0.218)	(0.219)
		p = 0.846	p = 0.853
Age: 40-50		−0.306	−0.313
		(0.245)	(0.248)
		p = 0.212	p = 0.208
Age: 50+		−0.144	−0.147
		(0.240)	(0.241)
		p = 0.547	p = 0.541
Male		0.479	0.536
		(0.157)***	(0.235)**
		p = 0.003	p = 0.023
Male:Male Instructor Video			−0.113
			(0.319)
			p = 0.723
Baseline	2.780	2.680	2.670
	(0.115)***	(0.205)***	(0.212)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.025	0.081	0.081
Adjusted R^2	0.021	0.059	0.055
Residual Std. Error	1.130 (df = 219)	1.110 (df = 215)	1.110 (df = 214)
F Statistic	5.710** (df = 1; 219)	3.770*** (df = 5; 215)	3.150*** (df = 6; 214)

Berkeley
UNIVERSITY OF CALIFORNIA

Linear Regression Results

Primary Outcome: Instructor Clarity Rating

	Dependent variable:		
	Instructor Clarity Rating		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	0.176	0.185	0.045
	(0.137)	(0.140)	(0.183)
	p = 0.201	p = 0.187	p = 0.805
Age: 30-40		−0.098	−0.092
		(0.202)	(0.203)
		p = 0.629	p = 0.652
Age: 40-50		0.002	0.029
		(0.223)	(0.223)
		p = 0.992	p = 0.898
Age: 50+		−0.035	−0.026
		(0.219)	(0.218)
		p = 0.873	p = 0.907
Male		0.220	0.023
		(0.138)	(0.198)
		p = 0.113	p = 0.910
Male:Male Instructor Video			0.393
			(0.280)
			p = 0.162
Baseline	3.630	3.590	3.650
	(0.095)***	(0.189)***	(0.194)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.007	0.021	0.029
Adjusted R^2	0.003	−0.002	0.002
Residual Std. Error	1.020 (df = 219)	1.020 (df = 215)	1.020 (df = 214)
F Statistic	1.640 (df = 1; 219)	0.902 (df = 5; 215)	1.070 (df = 6; 214)

Secondary Outcome: Quiz Score

	Dependent variable:		
	Quiz Score		
	Simple	Blocks Included	Gender Interaction Terms
	(1)	(2)	(3)
Male Instructor Video	−0.020	−0.024	−0.019
	(0.034)	(0.033)	(0.042)
	p = 0.556	p = 0.469	p = 0.641
Age: 30-40		−0.032	−0.032
		(0.047)	(0.048)
		p = 0.499	p = 0.499
Age: 40-50		0.032	0.031
		(0.050)	(0.051)
		p = 0.521	p = 0.540
Age: 50+		0.095	0.095
		(0.053)*	(0.053)*
		p = 0.071	p = 0.075
Male		−0.052	−0.045
		(0.035)	(0.050)
		p = 0.141	p = 0.369
Male:Male Instructor Video			−0.013
			(0.072)
			p = 0.854
Baseline	0.449	0.454	0.452
	(0.023)***	(0.042)***	(0.043)***
	p = 0.000	p = 0.000	p = 0.000
Gender Fixed Effects	No	Yes	Yes
Age Group Fixed Effects	No	Yes	Yes
Observations	221	221	221
R^2	0.002	0.047	0.048
Adjusted R^2	−0.003	0.025	0.021
Residual Std. Error	0.250 (df = 219)	0.246 (df = 215)	0.247 (df = 214)
F Statistic	0.350 (df = 1; 219)	2.140* (df = 5; 215)	1.780 (df = 6; 214)

Berkeley
UNIVERSITY OF CALIFORNIA

Limitations

- Potential violations of the exclusion restriction
  - Pitch and accent
  - Assumptions about video instructor

- Company for recruiting participants provided limited information on participants who may have dropped out of study (noncompliers and attrition)
  - Population that failed the attention check could be inherently different than actual respondents

Discussion / Conclusion

Our experiment set out to understand if there is generalized gender bias in educational videos

Results

- Only Instructor Enthusiasm rating exhibited a statistically significant treatment effect (Control Group = 2.78 out of 5, ATE = .364)
- Male subjects rated Instructor Enthusiasm statistically significantly higher than female subjects

Implications on Social Science and Future Work

- We did not find strong evidence of gender bias in our outcome measures
- We urge caution when interpreting these results as they may not generalize to broader populations
- Does gender bias exist in a non-academic setting?

Thank you!