Carnegie Mellon University

A/B Testing

Final Project

Team 18

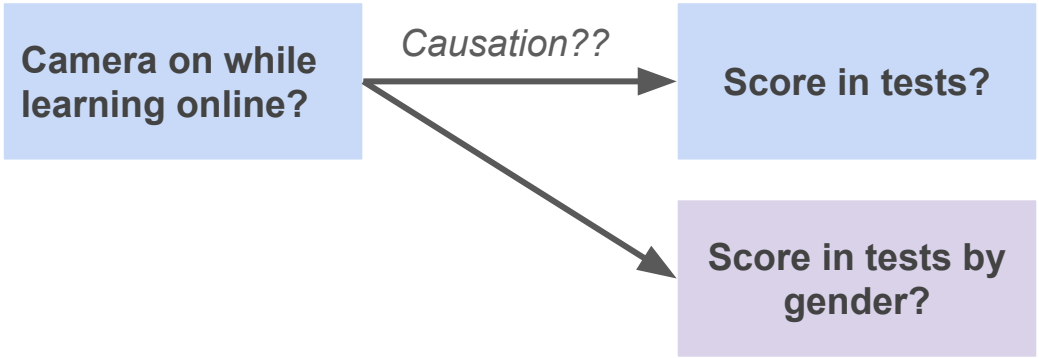
Carnegie Mellon University

Our Team

1. Fatima Salebhai
2. Pawanjeet Singh
3. Pallavraj Sahoo
4. Shashank Singh
5. Seemal Muzzafar
6. Yashvi Thakkar

Causal Question

1. **Does the audience pay more (or less) attention to the content of a video if they know they are being recorded?**
2. **Does gender affect the learning when being recorded?**



- We try to answer these questions by conducting an A/B test.

Background

- **Video conferencing** has become an increasingly **popular** way for students to interact with their classrooms.
- Video conferencing can foster a **sense of trust** and **connection** among students.
- Understand the relationship between keeping the **camera switched on** and **educational outcomes.**
- **Hypothesis: Being recorded lead to poorer educational outcomes**

Literature Review

Impact of self-consciousness and “being observed” on learning:

- **Self-consciousness** can lead to a **decrease in performance**, particularly when the task requires **creativity or problem-solving skills** [1]
- Research has suggested that **being observed** can lead to a **decrease in performance** when the observer is **perceived as critical or evaluative** [2]
- Videoconferencing in synchronous lectures have show to increase students' attention due to a **sense of community** but the self consciousness from turning the video on is still unexplored [5,6]
- Presence of a camera or the knowledge that one is being recorded can lead to self-consciousness and distraction, which can negatively impact performance on a task [7]

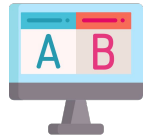
Effect of Gender:

- Women feel **more fatigue** when they have video **camera turned on** due to **self-consciousness** [8]
- Men tend to perform better when being observed by an actual audience, while women tend to perform worse when they are being observed. This may be due to gender differences in socialization and self-consciousness [9]

Data

- Conducted a **survey** from 15th Nov 2022 to 12th Dec 2022 (a month) to know collect current knowledge on **A/B testing**, showed a video and retested them on the concepts.
- Randomly asked **viewers to turn on video** and then viewers could choose to turn the camera on or not?
- Fetched data related to **facial features** of participants who kept their cameras on
- Fetched **demographic data** like age, gender, employment, race etc

A/B Test Design



- Randomly divided survey takers into **Treatment and Control group**
- **Treatment: Randomized recording a participant/ keeping cameras on.**

Observation about data collected:

1. **Control Group** - never offered to record: **141 observations.**
   Assumption: There is simply no possibility that the control group would have gotten a chance to put their video on and choose to record themselves. Therefore, all 141 observations are **compliers.**
2. **Treatment Group** - asked to record themselves: 113 observations.

Analysis

Averages of each group:

Group	Avg Pre-Quiz %	Avg Post-Quiz %
Control	62.76%	59.57%
Treatment	60.17%	60.17%
Compliers (Treatment)	60.20%	60.71%

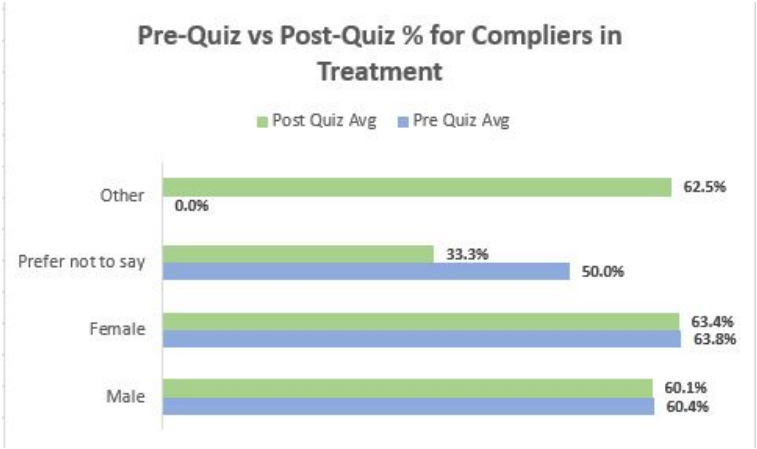
Interestingly, for the control group, the **average pre-quiz to post-quiz percentages** go **down from 62.76% to 59.57%**.

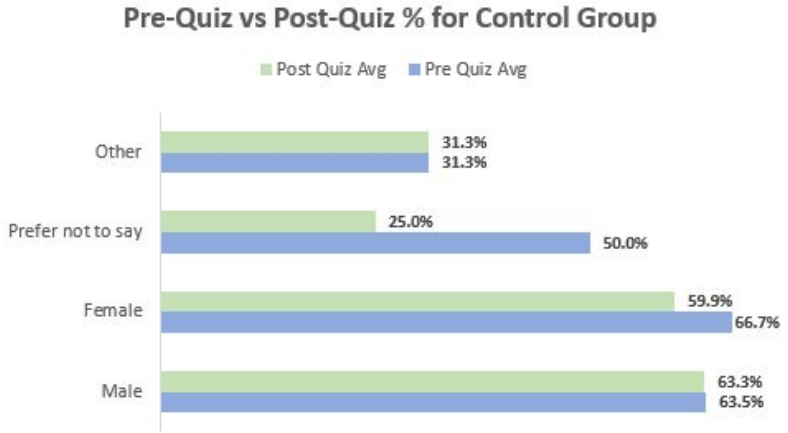
While for the compliers in the treatment group, there is a **very marginal increase from 60.20% to 60.71%**.

This could mean that not being asked to record leads to negative learning outcomes and quiz performance. While knowing that you are being recorded marginally enhances quiz performance.

Analysis

Control vs Treatment Group - Pre vs Post-Quiz Performance breakdown by Gender





The sizable result from above is the difference in the change in performance of females (compared to men) from pre-quiz to post-quiz, depending on whether or not they are given treatment.

This is further exacerbated by carrying out ITT with heterogeneous effects.

Regressions

1. Intention to Treat: % correct_post_quiz (outcome) ~ recording_eligible (Randomized treatment)

This is the Intent to Treat (ITT) effect, which is 0.0241, but this is statistically insignificant. This does not differentiate between the effect for compliers vs non-compliers in the treatment groups.

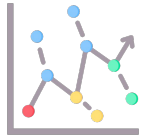
2. Local Average Treatment Effect: % correct_post_quiz(outcome) ~camera_on |recording_eligible

```
[441] LATE = ITT/(len(data[(data["compliers"]==1)])/len(data[(data["recording_elegible"]==1)]))

[442] LATE

      0.027789839339991664
```

LATE in this scenario is ITT/(% of Compliers in the dataset)

LATE is approximately +0.027

Regressions



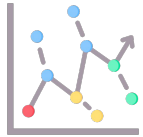
3. Intention to Treat with gender heterogeneous effects:

```
linear_model_hg = ols('correct_post_quiz ~ recording_elegible+Male_Female+
                       recording_elegible*Male_Female',
                      data=data).fit()
```

We find that the coefficient of the interaction term (recording_eligible*male_female i.e. treatment*heterogeneous_dummy_var) is insignificant, which tells us that there is no heterogeneous causal effect, that is, the effect is not different for men vs women.

But coeff. of male_female is significant. Interpretation: On average, holding all else constant, being a man increases the post-quiz performance by 0.31 points. This means that recording has no significant impact on the post-quiz percentage but gender plays a significant role.

Regressions



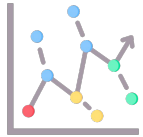
4. (LATE with gender heterogeneity) First and Second Stage using IV2SLS:

```
formula = ("correct post quiz ~ 1+recording elegible+survey_duration
+correct_pre_quiz+azure_features+video_error+[cam_allowed ~
Male_Female]")
mod = IV2SLS.from_formula(formula,data)
iv_res = mod.fit(cov_type="robust")
```

We try to identify the local average treatment effect of recording eligibility on post quiz score while moderating for gender.

But we do not find anything significant. This could be happening because the data we are working with is limited and does not truly capture the underlying relationships between these variables.

Conclusion



- Descriptive statistics gave results that the option of being recorded while watching a video/taking a quiz marginally improves performance (59.57% for control group vs 60.7% for compliers in the treated group).

- When this effect was studied for gender, recording deteriorated performance/scores for men (63.3% for control vs 60.1 for treatment%) while enhancing it for women (63.4% for treated vs 59.9% for women).

- Subsequent statistical Whitney-Mann U Tests showed how none of the differences were statistically significant. And the A/B tests also did not give any significant results for ITT and LATE (with or without heterogeneous effects on gender).

Thank you!