Carnegie Mellon University

The Effect of On-screen Timer on Reading Comprehension

A/B Testing Term Project | Fall 2020

Team 2 | Ajay, Arun, Sanjay, Jianyu, Kaiqi, Prasun

December 18, 2020

Presentation Breakdown

- Introduction
- Experimental Design
- Data Collection
- Results Obtained
- Discussions & Limitations

Introduction

- The effect of UI design elements on student's performance on assessments has been of interest to experimental psychologists and cognitive scientists to improve the quality of assessments.
- We are interested in understanding the effect of a visible on-screen timer on reading compression in the context of timed assessments. Our experiment is designed to test if there is a causal relationship between the presence of a visible timer on the screen and student performance on reading comprehension.

Experimental Design

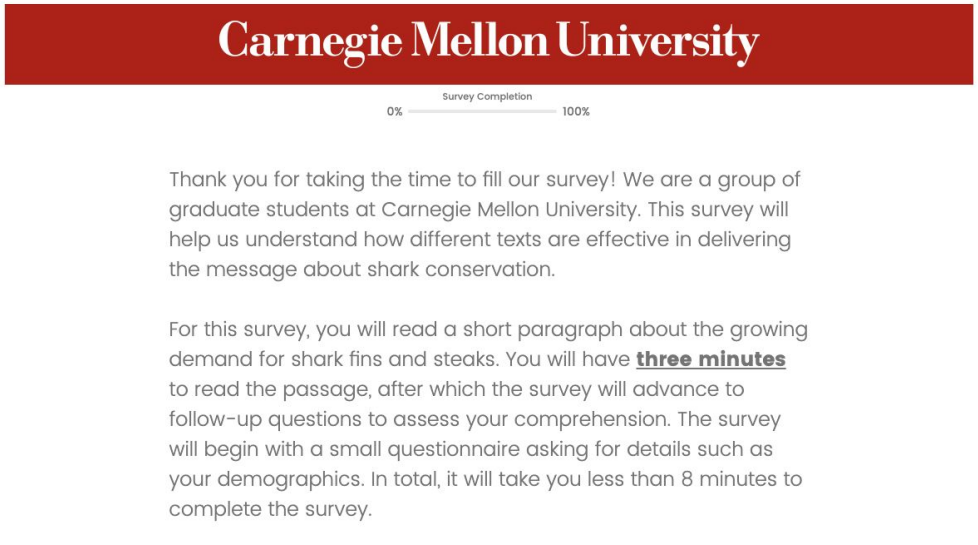
- We randomized subjects into **treatment** and **control** groups to eliminate any selection bias
- **Treatment:** presence of an on-screen timer.
- Unit of analysis: an individual

Survey Breakdown

- **Pretest:** To ensure that the experiment has been randomized correctly, we collected demographic data and information about the subject's familiarity with the topic.
- **Reading passage:** GRE-level 400 words essay with a maximum reading time of 3 minutes. The test group has a always-on-screen timer showing a countdown from 3 minutes to zero while the control group does not have this timer. Both the groups will be automatically taken to the next page after 3 minutes.
- **Assessment:** 8 questions (multiple choices and select all that apply) to check reading comprehension, with a maximum score of 12 and minimum of 0.

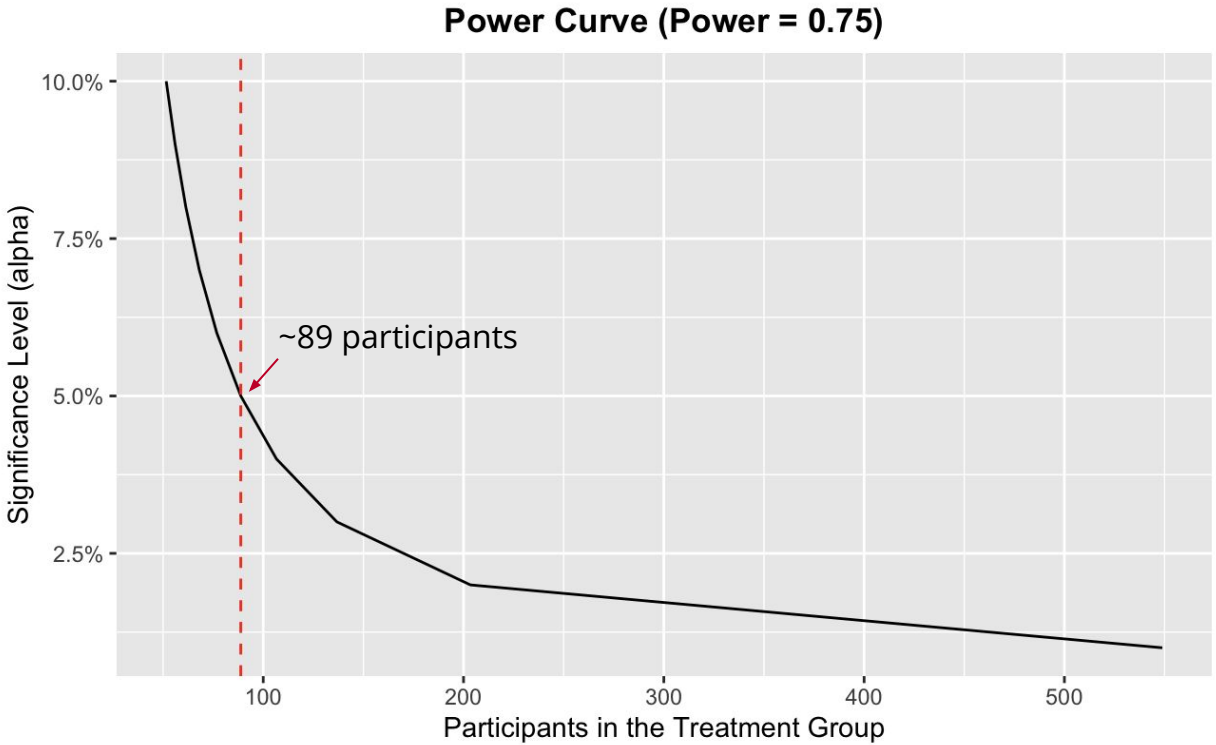
Experimental Design (contd..)

- **Channels**
  - Survey Modelling and Administration: Qualtrics
  - Data Cleaning: Excel/R
  - Data Analysis: R



Carnegie Mellon University

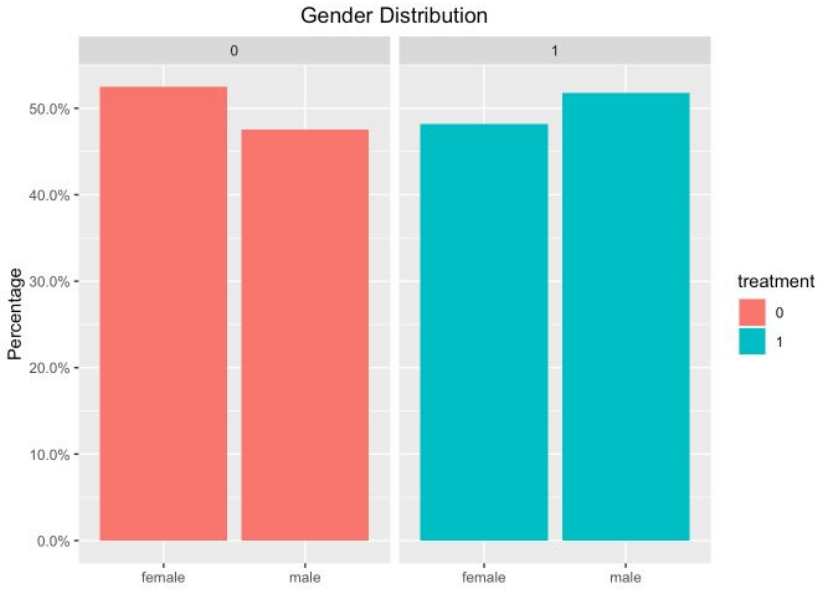
Data Collection



With the power of 0.75, significance level at 5% and estimated 60 control units, we would have needed 89 participants to get a desired lift + baseline probability of 0.1 from 0.01.

Carnegie Mellon University

Data Collection

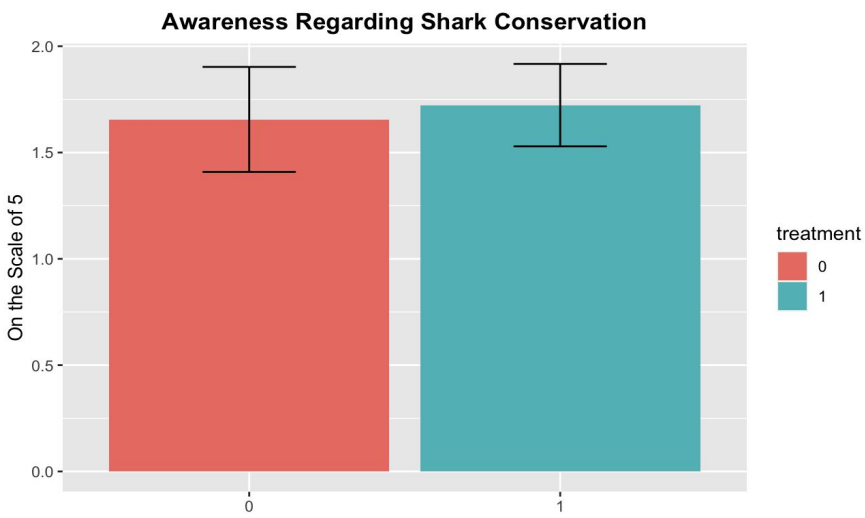


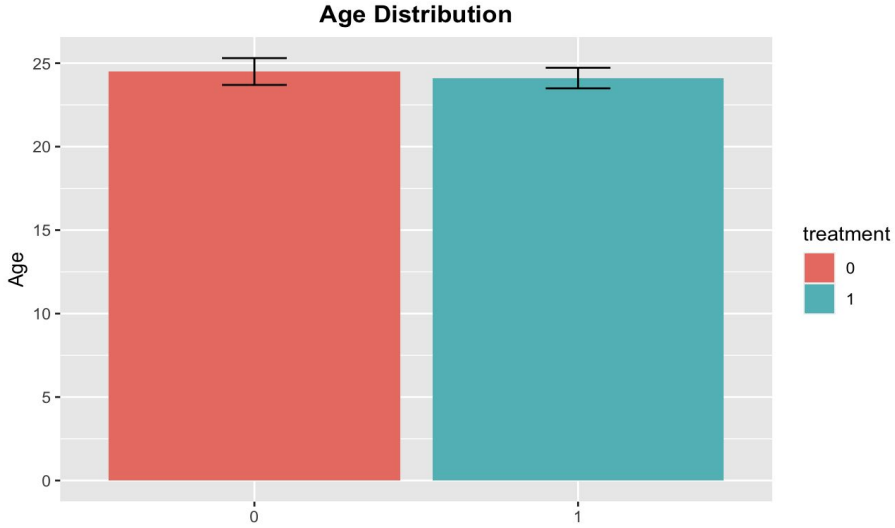

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  df_final$gender and df_final$treatment
## X-squared = 0.11377, df = 1, p-value = 0.7359
```

- The distribution of gender across treatment and the control group depicts that the gender is fairly distributed across both groups. This demonstrates that the treatment and the control group are similar "to begin with."
- p-value of 0.7359 on Chi-squared test

Carnegie Mellon University

Data Collection





The overlapping error bars in age distribution and subjects' prior awareness about shark conservation further affirm that the difference between the treatment and the control group is not statistically significant.

Carnegie Mellon University

Results

Intention to Treat (ITT) and Heterogeneous Causal Effect (Education as Modifier)

not statistically significant

	Dependent variable: final_score	
	(1)	(2)
treatment1	0.125	3.504
	(0.387)	(2.261)
educationassociate	7.799***	7.910***
	(2.846)	(2.830)
educationbachelors	6.427***	9.675***
	(2.069)	(3.037)
educationhigh sch	7.248***	7.285***
	(2.215)	(2.199)
educationmasters	6.289***	9.985***
	(2.120)	(3.107)
educationphd	5.865**	10.910***
	(2.477)	(3.555)
educationsome college	6.685***	7.927***
	(2.301)	(2.399)
treatment1:educationassociate		
treatment1:educationbachelors		-3.092
		(2.324)
treatment1:educationhigh sch		
treatment1:educationmasters		-3.828
		(2.344)
treatment1:educationphd		-6.544*
		(3.476)
treatment1:educationsome college		
Constant	-1.410	-4.966
	(3.261)	(3.939)
Interaction terms	No	Yes
Observations	108	108
R2	0.242	0.278
Adjusted R2	0.128	0.142
Residual Std. Error	1.860 (df = 93)	1.846 (df = 90)
F Statistic	2.124** (df = 14; 93)	2.040** (df = 17; 90)
Note:		*p<0.1; **p<0.05; ***p<0.01

Carnegie Mellon University

Results

Simple OLS Regression

	Dependent variable:	
	final_score	
	(1)	(2)
treatment1	0.125	-0.139
	(0.387)	(0.575)
gendermale	1.123***	0.834
	(0.389)	(0.607)
treatment1:gendermale		0.493
		(0.791)
Constant	-1.410	-1.236
	(3.261)	(3.284)
Gender Interaction Term	No	Yes
Observations	108	108
R2	0.242	0.245
Adjusted R2	0.128	0.122
Residual Std. Error	1.860 (df = 93)	1.866 (df = 92)
F Statistic	2.124** (df = 14; 93)	1.995** (df = 15; 92)
Note:		*p<0.1; **p<0.05; ***p<0.01

- Heterogeneous Causal Effect (Gender as Modifier here)
- Not statistically significant either
- Restricted the regression to observations with reading time more than 60 seconds, thus reduced observations

Carnegie Mellon University

Results: Completion Rate

```
##
## ================================================================
##                                Dependent variable:
##                    ---------------------------------------------
##                                    completion_rate
##                              (1)                   (2)
## ----------------------------------------------------------------
## treatment1                  2.690                3.911**
##                            (2.564)              (1.915)
##
## Constant                  95.498***            106.202***
##                            (23.667)             (16.465)
##
## ----------------------------------------------------------------
## Reading Time Filter          No                    Yes
## Observations                143                   115
## R2                         0.084                 0.095
## Adjusted R2                -0.016                -0.022
## Residual Std. Error  14.430 (df = 128)     9.707 (df = 101)
## F Statistic         0.837 (df = 14; 128) 0.815 (df = 13; 101)
## ================================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```

- The survey completion rate increased by 3.91 percentage points in the treatment group
- The difference is statistically significant at 5% level

Carnegie Mellon University

Results: Passage Reading Time

```
##
## ============================================================
##                                Dependent variable:
##                    ----------------------------------------
##                                      read_time
##                              (1)                  (2)
## ------------------------------------------------------------
## treatment1                  7.650               13.877*
##                            (10.466)             (7.425)
##
## Constant                   168.238*            194.207***
##                            (96.590)             (63.850)
##
## ------------------------------------------------------------
## Reading Time Filter           No                   Yes
## Observations                 143                  115
## R2                          0.080                0.129
## Adjusted R2                 -0.021               0.016
## Residual Std. Error  58.892 (df = 128)    37.643 (df = 101)
## F Statistic         0.795 (df = 14; 128) 1.147 (df = 13; 101)
## ============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```

- The treatment group spent 13.88 seconds more on average on reading the passage, and the difference is statistically significant at 10% level

Carnegie Mellon University

Results: Time on Answering Questions

	Dependent variable:	
	time_on_questions	
	(1)	(2)
treatment1	29.313	42.342**
	(20.562)	(20.958)
Constant	-92.659	23.189
	(178.842)	(178.556)
Reading Time Filter	No	Yes
Question Completion Time Filter	No	Yes
Observations	125	105
R2	0.112	0.088
Adjusted R2	-0.001	-0.031
Residual Std. Error	106.810 (df = 110)	101.160 (df = 92)
F Statistic	0.988 (df = 14; 110)	0.742 (df = 12; 92)
Note:		*p<0.1; **p<0.05; ***p<0.01

- The treated units spent 42.3 more seconds on average than did the control units
- The increase is statistically significant at 5% level

Carnegie Mellon University

Discussions and Limitations

- We measure four variables as our outcome variable of interest: (1) reading comprehension score, (2) survey completion, (3) reading time, and (4) question completion time.
- Our results are **underpowered,** meaning we do not have enough observations to attain our desirable lift at the 5% significance level. An iteration of the regression with increased sample size will yield much more robust results.
- The relative **difficulty** of the passage itself could be a factor that is not accounted for in our study. Different **durations of time** allowed to read the passage may impact the results. Future studies will benefit from varying levels of passage difficulty and time lengths.

Thank You!
Questions?