Effect of displaying reading time on the willingness to read an article

Team 3: geyues, sushmitd, nikitham, jiahuich, yufanyan, olasebik

1. Abstract

The goal of our experiment is to study the effect of showing the time taken to read an article on the reader. We want to test if the reading time affects the reader's motivation to either read the article or not. We want to test if readers alter their behavior based on the availability of the reading time.

The experiment was conducted via a randomized controlled survey created with Qualtrics that was distributed to the public through WhatsApp, Slack and WeChat. The data cleaning, exploratory data analysis and regression computation was carried out in R.

Through our analysis of the 186 survey responses, we find that while there is no significant difference in the willingness to read an article based on the treatment on the entire population, we do find interesting results for specific subpopulations such as the Black and African Americans and in the subgroup of people who prefer not to give away their race information.

2. Introduction

In today's fast-paced world, time is a commodity. We want to believe that our time is valuable and that we are using it wisely. Reading an article on a subject of interest is an excellent way to gain knowledge on the subject and we want to utilize the findings from studying the human psyche and prior research to see if we can increase the amount of time people spend reading articles.

We provide our respondents with the option to choose a domain they are most interested in and rule out selection bias by randomizing the readers who will be provided the reading time along with the title of the article. Prior research in psychology, like the paper on the 'Paradox of Choice' published by Messner and Wanke, suggests that *"The more we know about something — including precisely how much time it will consume — the greater the chance we will commit to it."*

3. Experiment Design

The A/B testing experiment addresses the question: whether having an estimated reading time label will impact users' interest/willingness in reading an article.

Unit of Analysis:

The unit of analysis is at the individual level. We distribute the survey to each individual and use the survey results to proceed with further analysis.

Independent Variable(s):

Interest/willingness level: we use numbers in the range of 1 to 7 to represent the user's interest level where 1 is the lowest level of interest and 7 is the highest level of interest.

Other Variables:

Variable name	Variable value
gender	male, female, other, prefer not to answer
age	0-17, 18-25, 26-40, 40+ (unit: years old)
ethnicity	white or Caucasian, Black or African American, Asian, Hispanic or Latino, multiracial or biracial, other, prefer not to answer
hours/day spend on computer/phone	< 1, 1-2, 2-4, 4-6, more than 6 (unit: hours)
Time spend reading articles on computer/phone	< 30mins, 30mins - 1hr, 1-2 hours, >2 hours
Article topic/category	entertainment, sports, fashion, technology, science, economy

Hypothesis:
We assume that a user's willingness to read an article tends to increase if the article is attached with reading time estimation

Survey Implementation:
We use Qualtrics to design and deliver the survey. We have distributed the survey to peers at Carnegie Mellon University and team members' personal networks through social media platforms such as Whatsapp, Slack, Wechat etc.

Step1. Demographic information & individual reading habit collection
The first section of the questionnaire asks for participant's demographic information as well as individual reading habit (how much time one spend on reading articles and the total time spent on computer/phone daily)

Step2. Randomization
Based on the article topic/category participants choose, we randomly assign treatment to each participant for each article category. The participant will either see an article title alone(control group) or will see an article title along with the estimated reading time(treatment group). Randomization is implemented to eliminate selection bias.

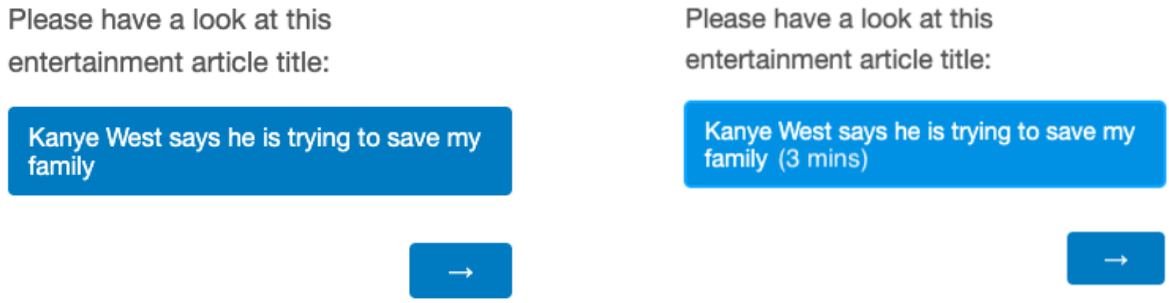

Figure 1-2 Example of the randomly assigned article title in the survey

Step3. Gather results

After participants filled out the surveys, we collected the questionnaire answers for further analysis. We filtered out empty/invalid answers (e.g some people didn't answer the question of willingness to read an article) and only kept valid surveys.

4. Exploratory Data Analysis

Before analyzing our modeling results, we did this exploratory data analysis to better understand our data.

4.1 User Demographic

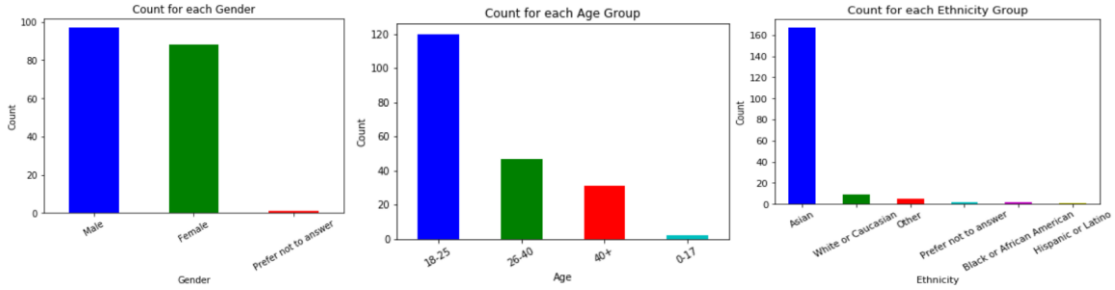

Figure 3-5 User demographic: gender, age and ethnicity distribution

- Gender Distribution: Male and Female are nearly 50%-50%.
- Age Distribution: Most of our participants are between 18 and 25 years old.
- Ethnicity Distribution: The majority of our participants are Asian.

4.2 Time spent on mobile phone and computers & Time spent on reading articles on mobile phone and computers

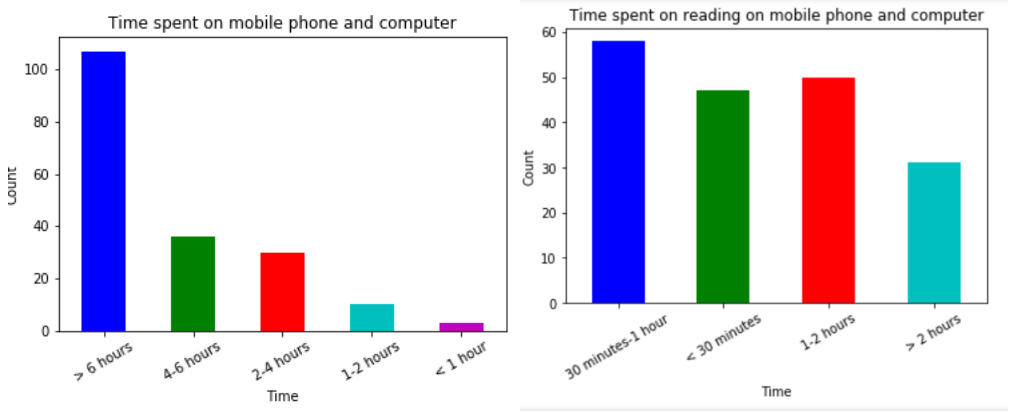

Figure 6-7 User time spent: on mobile phone and computer, on reading articles on mobile phone and computer

- Time spent on mobile phone and computer: Most of our participants spend more than 6 hours on their mobile phones or computers every day.
- Time spent on reading articles on computer and phone: Around 31% of our participants spend 30 minutes to 1 hour every day. Nearly 27% of them spend 1-2 hours, about 25% spend no more than 30 minutes, and others spend more than 2 hours.

4.3 Topic Preference

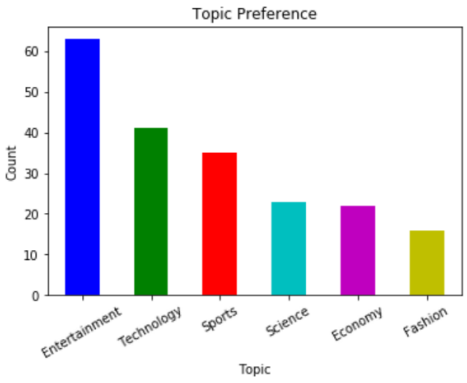

Figure 8: User topic preference distribution

● Topic interested in: The top 3 topics participants most interested in are Entertainment, Technology and Sports.

4.4 Willingness level distribution

In the survey, 50% of the participants are given the article title with time estimation tag, and others are given the article title without the tag under randomization.

4.4.1 Willingness level distribution without estimation time tag

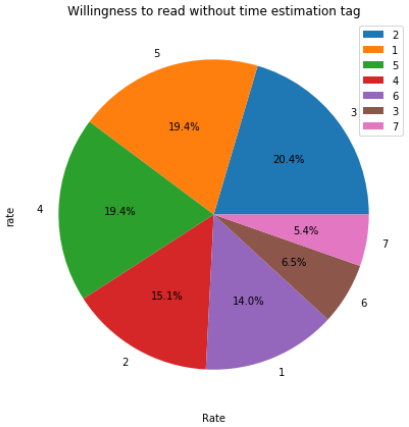

Figure 9: User willingness to read without time estimation tag

● The average level of willingness is 3.5591.
● The percentage of different levels of willingness is relatively evenly distributed except for level = 6 and level = 7.
● Most participants have medium willingness (level = 3) to read the article in their interested topic.
● The total percentage of level = 3, level = 4 and level = 5 is nearly 60%.

4.4.2 Willingness level distribution with estimation time tag

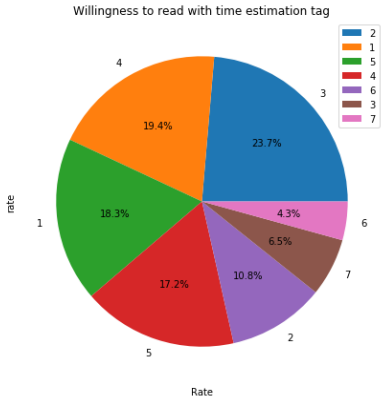

Figure 10: User willingness to read with time estimation tag

- The average level of willingness is 3.4516.
- The percentage distribution is relatively similar to distribution without a time tag.
- Compared with willingness level distribution without time tags, the number of participants who have very low willingness (level = 2) to read the article decreases, but the number of participants who have medium willingness (level = 3) and the number of participants who have relatively high (level = 5) increases.

5. Modeling Result

5.1 Analyze the willingness of control group and treatment group without or with control variable

We run a regression of the treatment on the willingness both with or without control variables which are time spent online per week, gender, age and ethnicity. On average, there is no statistically significant evidence that showing the reading time to readers would reduce their willingness to read the whole article.

Interpretation of the Results:

- First: The beta coefficient is -0.108 indicating that there is a negative causal effect of the treatment (i.e providing the reading time) on the willingness to read the article. However, there is no significant evidence that showing the reading time would reduce the willingness to read the whole article.
- Second: The beta coefficient is -0.068 indicating that there is a negative causal effect of the treatment (i.e providing the reading time) on the willingness to read the article. However, there is no significant evidence that showing the reading time would reduce the willingness to read the whole article.
- Third: The beta coefficient is -0.116 indicating that there is a negative causal effect of the treatment (i.e providing the reading time) on the willingness to read the article. However, since the results are not statistically significant, it is hard to come to a conclusion if there is a significant effect of the treatment on the outcome of interest.

Regression

	Dependent variable:		
	willingness		
	first	second	third
treatment	-0.108	-0.068	-0.116
	(0.250)	(0.249)	(0.257)
Time Spent: 1 - 2 hours		1.205	1.687
		(1.110)	(2.016)
Time Spent: 2 - 4 hours		0.947	1.436
		(1.023)	(1.933)
Time Spent: 4 - 6 hours		0.811	1.186
		(1.014)	(1.909)
Time Spent: More than 6 hours		0.247	0.648
		(0.989)	(1.911)
Gender: Female			-0.234
			(0.258)
Gender: Prefer not to answer			3.752
			(3.003)
Age: 18 - 25 years			2.164
			(3.081)
Age: 26 - 40 years			2.284
			(3.091)
Age: 40+ years			2.069
			(3.123)
Ethnicity: Black or African American			2.752
			(1.829)
Ethicity: Asian			1.319**
			(0.648)
Etnicity: Hispanic or Latino			2.169
			(1.815)
Ethnicity: Other			2.583**
			(1.043)
Ethnicity: Prefer not to answer			2.729**
			(1.359)
Constant	3.559***	3.023***	-0.752
	(0.177)	(0.977)	(2.487)
Observations	186	186	186
R2	0.001	0.042	0.098
Adjusted R2	-0.004	0.015	0.018
Residual Std. Error	1.704 (df = 184)	1.687 (df = 180)	1.684 (df = 170)
F Statistic	0.185 (df = 1; 184)	1.578 (df = 5; 180)	1.229 (df = 15; 170)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 11: Results of running regression of the treatment on the willingness to read the article with or without the reading time tag

Then we decided to run regressions of the treatment on different subgroups to explore if they are affected differently by the treatment. We assume different subgroups are affected differently by age, gender, ethnicity and reading time per day.

5.2 Analyze the willingness of different ethnicity

The result shows that there are differences among different ethnicities of the willingness to read the whole article. Asian people (what the constant represents) and Black or African American people are more willing to read the whole article when they are shown the estimated reading time. On average, Asian people have a 3.542 higher score of willingness to read the whole article when they are shown the reading time than people in the same ethnicity group who are not shown the reading time. On average, Black or African American people have a 3.090 higher score of willingness to read the whole article when they are shown the reading time than people in the same ethnicity group who are not shown the reading time. Both are statistically significant at a 99% significant level.

There are also interesting gaps when showing the reading time compared to not showing the reading time. For example, at first we thought that if the reading time was not shown, Black or African American people's willingness is only 2.000 (3.542-1.542). If the reading time was shown, their willingness becomes 6.632 (3.542+3.090). It is a significant change and we wonder if there might be cultural differences. But later we noticed that the number of Black or African American people who took the survey is only 2. This creates the huge gap.

	Dependent variable:
	willingness
	plain
treatment	-0.090
	(0.266)
ethnicityBlack or African American	-1.542***
	(0.184)
ethnicityHispanic or Latino	0.548***
	(0.193)
ethnicityOther	0.791
	(1.496)
ethnicityPrefer not to answer	3.458***
	(0.184)
ethnicityWhite or Caucasion	-0.542
	(0.537)
treatment:ethnicityBlack or African American	3.090***
	(0.266)
treatment:ethnicityHispanic or Latino	
treatment:ethnicityOther	1.256***
	(1.863)
treatment:ethnicityPrefer not to answer	-3.910***
	(0.266)
treatment:ethnicityWhite or Caucasion	-0.910
	(0.677)
Constant	3.542***
	(0.184)
Observations	186
R2	0.071
Adjusted R2	0.018
Residual Std. Error	1.685 (df = 175)
F Statistic	1.336 (df = 10; 175)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 12: Results of running regression of the treatment on the willingness of different ethnicity

5.3 Analyze the willingness to read the article of people with different online reading frequency

The results show on average, people who read articles for 30 min-1 hour per day have a lower willingness score of 0.753 when they are shown the estimated reading time compared with those in the same frequency group when they are not shown the estimated reading time. It is statistically significant at a 99% significance level. On average, people who read articles more than 2 hours per day have a lower willingness score of 1.464 when they are shown the estimated reading time compared with those in the same frequency group when they are not shown the estimated reading time. It is statistically significant at a 99% significance level.

It seems that when people read articles on their computer/phone for a longer time, they are less willing to read the whole article when they are shown the reading time. It is reasonable because when they read more time, they are more sensitive to how much more time they need while they

read less time, they might not care how much time is spent at the beginning. Or it might be because longer time reading makes people care more about time control.

OLS Regression: The Effect of Interaction of Adding a Reading Time and Reading Frequency on Willingness to Read the Whole Article

	Dependent variable:
	willingness
	plain
treatment	0.426
	(0.266)
frequency30 minutes to 1 hour	0.411**
frequencyLess than 30 minutes	0.184
frequencyMore than 2 hours	0.284
treatment:frequency30 minutes to 1 hour	-0.753***
treatment:frequencyLess than 30 minutes	-0.164
treatment:frequencyMore than 2 hours	-1.464***
Constant	3.316***
	(0.184)
Observations	186
R2	0.038
Adjusted R2	-0.0003
Residual Std. Error	1.700 (df = 178)
F Statistic	0.992 (df = 7; 178)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 13: Regression results of the treatment on the willingness of different online reading frequency

5.4 Analyze the willingness to read the article of different age groups

OLS Regression: The Effect of Interaction of Adding a Reading Time and age on Willingness to Read the Whole Article

	Dependent variable:
	willingness
	plain
treatment	-1.045
	(0.787)
age18-25 years old	0.786*
	(0.410)
age26-40 years old	1.567***
	(0.597)
age40+ years old	1.962***
	(0.629)
treatment:age18-25 years old	1.367
	(0.849)
treatment:age26-40 years old	0.178
	(0.957)
treatment:age40+ years old	
Constant	2.500***
	(0.360)
Observations	186
R2	0.047
Adjusted R2	0.015
Residual Std. Error	1.687 (df = 179)
F Statistic	1.470 (df = 6; 179)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 14: Regression results of the treatment on the willingness of different age groups

The result shows that on average, people who are 0- 17 years old have a higher willingness to read an article if they are shown the time needed to finish the article. However, the breakdown of the participants of the survey shows that they only make up about 1% of the respondents of the survey.

Consequently, it can be concluded that the willingness for that age group is being skewed by the small number of people that responded in that age group.

5.6 analyze the willingness to read the article of different genders

OLS Regression: The Effect of Interaction of Adding a Reading Time and gender on Willingness to Read the Whole Article

	Dependent variable:
	willingness
	plain
treatment	0.041
	(0.787)
genderMale	0.367
genderPrefer not to answer	-0.378
treatment:genderMale	-0.306
treatment:genderPrefer not to answer	
Constant	3.378***
	(0.360)
Observations	186
R2	0.008
Adjusted R2	-0.014
Residual Std. Error	1.712 (df = 181)
F Statistic	0.344 (df = 4; 181)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 15: Regression results of the treatment on the willingness of different gender groups

On average, people who identify as female are more likely to read the article if they are shown the reading time. This result is significant at the 99% confidence interval.

6. Conclusion

In conclusion, there is no statistically significant evidence that showing the reading time to readers would reduce their willingness to read the whole article. There are differences among different subgroups, i.e. ethnicity, age, gender and reading time online. We cannot give definite conclusions on ethnicity subgroups or age sub groups as some of the groups have too small a number of people. But we can conclude that if people spend more time on online reading, they would be less likely to read the whole article when the reading time is shown. This might be because they are more sensitive to the reading time after reading a long time.

There are limitations in our experiment. First of all, there is an obvious imbalance in the percentage of each ethnicity group, which might be due to the fact that most group members come from Asia. Asian people have the highest percentage of 89.78%. This might cause the inaccuracy of results. Second, the total number of respondents is 209. But after data cleaning, the valid number of respondents is only 186. Small samples might also cause bias.

7. References

[1] Messner, C. and Wänke, M., 2011. Unconscious information processing reduces information overload and increases product satisfaction. *Journal of Consumer Psychology*, 21(1), pp.9-13.