```
title: "BA 830 - Effect of Physical Vs. Digital Book Reading on Grade" author: "Team members: Kaiyu Wang, Tianji Peng, Yapei Xiong, Zheming Xu, Zijing Wang" date: "12/05/2021"
```

## Introduction

During the coronavirus pandemic, students around the world switched from classrooms to remote online learning. Many people replace hardcopy textbooks and worksheets with websites and other digital resources. Even though digital books have been around for decades, but how well are we absorbing it all?

We, as college students, who experienced the whole change of remote courses and used the digital book and digital tests have to face these dramatic impacts during study life. When the epidemic came, we changed from offline to remote, which caused us to do all works on the Internet. However, since the change of environment, how efficiency we may learn from the course. By reading the course resources online, is it similarly understandable as the physical works, or it may be more efficient than traditional physical books. Is there really a gap in grades between reading physical books and digital books? With these questions, we decide to design an experiment to statistically conclude the effects of physical books and digital books on grades.

## Prior Research

Many studies and experts have announced that physical books are easier to comprehend than digital books. Print reading is kind of like meditation — focusing our attention on something still, And it's a whole different kind of immersion than responding to digital stimuli. (Kerry Benson,2020). There is article that also describes that print is easier to comprehend than digital text because it provides spatial and tactile cues to help readers process words on a page. Mindset may also be a factor. If people associate screen time with casual web-surfing they may rush through without fully absorbing the text (Benson 2020). So our hypothesis is that students will have better score if they read physical book.

To observe this, we created an experiment with 68 participants to test whether reading physical books or digital books impacted the understanding on some articles. These participants had to read an article and were given a Qualtrics survey tool about the contents of the article. All original volunteers were randomly assigned to the control group (digital book) or the treatment group (physical book). To find out if the hypothesis is right, we need to check the differences in grade between reading physical books and reading digital books.

```
#library
library(data.table)
library(tidyverse)
library(fixest)
library(fixest)
library(lubridate)
library(stargazer)
library(modelsummary)
library(kableExtra)
```

```
# read data
df<- fread('Final Book Reading Survey Results.csv')
# rename columns
names(df)[names(df) == 'A&S'] <- 'AorS'
names(df)[names(df) == 'prefer type'] <- 'prefer_type'
names(df)[names(df) == 'actual type'] <- 'actual_type'</pre>
```

```
# clean data
df$sex <- ifelse(df$Sex=="Male",1,0)</pre>
df$AorS[which(df$AorS == "SCIENCE.")] <- "SCIENCE"</pre>
df$major <- ifelse(df$AorS=="SCIENCE",1,0)</pre>
df$prefer <- ifelse(df$prefer_type=="physical books",1,0)</pre>
df$treatment <- ifelse(df$actual_type=="physical books",1,0)</pre>
df$note[which(df$note == "yes")] <- "Yes"</pre>
df$note_c <- ifelse(df$note=="Yes",1,0)</pre>
df$interested[which(df$interested == "yes")] <- "Yes"</pre>
df$interested_c <- ifelse(df$interested=="Yes",1,0)</pre>
df$finish_c <- ifelse(df$finish=="Yes",1,0)</pre>
df$grade<-df$Grade
df$grade<-gsub("%","",as.character(df$grade))</pre>
#create a new data set for future analysis
df_c < -subset(df, select = -c(1:9))
df_c$grade <- as.numeric(df_c$grade)</pre>
dim(df_c)
## [1] 68 8
#find mean points scored by treatment
df_c %>%
   group_by(treatment) %>%
   summarise(mean_grade = mean(grade))
## # A tibble: 2 x 2
##
    treatment mean grade
         <dbl>
##
                     <dbl>
## 1
                      57.1
             0
## 2
             1
                      66.4
##proportion test
#proportion test
prop.test(nrow(df_c[treatment == 1]), nrow(df_c), 0.5)
##
   1-sample proportions test without continuity correction
##
## data: nrow(df_c[treatment == 1]) out of nrow(df_c), null probability 0.5
## X-squared = 0, df = 1, p-value = 1
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
## 0.384 0.616
## sample estimates:
##
     р
## 0.5
```

The value of 0 on treat column determine the reading of physical book; and the value of 1 determines the reading of digital book. From second table, we can see that on average, the grade of paper book readers are slightly higher than e-books. Then, our group is using simple randomization to select our treatment group and control group. By using the proportion test, we can see the p value here is 1 which is larger than 0.05, so we fail to reject the null hypothesis and prove that our experiment of treatment and control group was well randomized.

# Method

## **Participants**

Since the main purpose of our experiment was to observe whether reading on a digital device would impact the quality and efficiency of study, we think students from all ages would be the best fit for this study. Therefore, our initial data collecting method was to create a questionary on the google form, and post the link to social media such as WeChat, Instagram, Facebook, etc. However, we have set our control group as reading a physical copy of the same material, so that we could have a more direct view about the impact by comparing two groups. To keep the consistency of the physical copies presented to participants, we had to limit our target for the control group to students who live in the Boston area. To imitate possible impact created by other elements, we kept our questionnaires for both control and treatment groups very similar. The only difference is that the questionaries for the treatment group included a link to the pdf version of material. We started data collecting by messaging people who live in the Boston area and asking them to participate. Meanwhile, we posted the questionnaire for the pdf version on our social media and asked friends who were not living in the Boston area to participate. Overall, we collected 68 observations, where 30 identify as male and 38 identify as female. Although this sampling method was very convenient, it also created a few biases. The most obvious one is that most of the control group participants were from the Boston area while partial of the treatment group were living outside of the Boston area such as the west coast. And most of the participants attended school right where they are living now. Secondly, since most of the participants were someone, we know from previous life, most of them have a background in business. This might limit the validity of our result.

# Randomization

When we try to decide which participants to put in a control or treatment group, we try to randomize by blocking on gender to have greater statistical power. We got 25 males who were willing to participate in the experiment, so we randomly assigned 13 of them to the treatment group and assigned the second half to the control group. For 34 female participants, we conducted the same method, so we assigned 17 to treatment and 17 to control. We chose to conduct blocking on gender because we want to control the number of male and female participants. An even number of participants may give more equal and accurate data on further analysis. We therefore did the randomization between the treatment and control group.

# Pre-Experiment Randomization Check

After we conduct blocking, we still need to run a randomization check on other variables. The results of these regression are present as below:

```
reg.sex <- lm(sex ~ treatment, data = df_c)
reg.sex%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.441	0.086	5.1	0
treatment	0.000	0.122	0.0	1

```
reg.major <- lm(major~ treatment, data = df_c)
reg.major%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
    "P-Value"), digits = c(0, 3, 3, 3, 3), align = c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.529	0.084	6.29	0.000
treatment	0.147	0.119	1.24	0.221

```
reg.prefer <- lm(prefer ~ treatment, data = df_c)
reg.prefer%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.412	0.086	4.81	0.00
treatment	0.176	0.121	1.46	0.15
	7 /			. 10

```
reg.note <- lm(note_c ~ treatment, data = df_c)
reg.note%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.353	0.077	4.58	0.000
treatment	-0.147	0.109	-1.35	0.182

```
reg.interested <- lm(interested_c ~ treatment, data = df_c)
reg.interested%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.441	0.087	5.09	0.000
treatment	0.088	0.123	0.72	0.474

```
reg.finish <- lm(finish_c ~ treatment, data = df_c)
reg.finish%>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.559	0.086	6.494	0.00
treatment	0.029	0.122	0.242	0.81

By running regressions for all the pre-experiment variables, we selected based on our survey, we found out that variables such as participants' sex, major, reading preference, note habit, and interest level were given us a p value bigger than 0.05. In this case, we are not able to reject the null hypothesis. It means that all those variables are well randomized between the treatment group and control group.

## **EDA**

#### Distribution of responses based on Gender

```
Nale Female Sex
```

#### Distribution of responses based on Prefer A or S

```
#Distribution of responses based on Prefer A or S

df %>% count(AorS)

## AorS n

## 1: ART 27

## 2: SCIENCE 41

ggplot(data = df, aes(x = AorS)) + geom_bar(color = '#2e2e2e' ,fill = '#04fbc4')

## AorS

ART SCIENCE

AorS
```

### Distribution of responses based on prefer type

### Distribution of responses based on finish

```
#Distribution of responses based on finish

df %>% count(finish)

## finish n

## 1: No 29

## 2: Yes 39

ggplot(data = df, aes(x = finish)) + geom_bar(color = '#2e2e2e', fill = '#04fbc4')

## 40

10

10

10

No Yes

finish
```

### Distribution of responses based on treatment group

### Distribution of responses based on major

# Data Analysis

### Regression

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	57.06	4.71	12.1	0.000
treatment	9.32	6.67	1.4	0.167

In the regression of model1 we looked at the grade as the outcome variable first. We get an intercept of 57.06, which represents the mean of the grade of our control group. The estimate for the treatment effect is 9.32 and has a standard error of 6.67. However, the p-value here is 0.167 which means that the difference is insignificant.

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	58.85	6.22	9.457	0.000
treatment   preferTRUE	4.07	7.41	0.549	0.585

For the regression above we added one fixed effects(prefer) to see if the if the prefer has an impact on the grade. It has an estimated treatment effect of 4.07 with an standard error of 7.41. However, it is also insignificant.

```
##
                                            reg.female
                            reg.male
## Dependent Var.:
                               grade
                                                  grade
                    56.27*** (6.939) 57.68*** (6.567)
## (Intercept)
## treatment
                       7.933 (9.814)
                                         10.42 (9.287)
## S.E. type
                                  IID
                                                    IID
## Observations
                                  30
                                                     38
## R2
                             0.02281
                                               0.03379
                            -0.01209
                                               0.00695
## Adj. R2
```

Because our general treatment effect is insignificant, we want to run a model about gender to see if the treatment effect will be influenced by gender or not. As we can see, both treatment effect for boys and girls are insignificant. So for both boys and girls, the treatment is insignificant. As we can see, girls have higher value for both intercept and ATE value. However, we can not say girls reading paper books will achieve better results than boys reading paper books because both p-value is higher than 0.05.

```
##
                        reg.physical
                                           reg.digital
## Dependent Var.:
                               grade
                                                 grade
##
                    54.50*** (6.760) 58.85*** (6.364)
## (Intercept)
                      20.15* (8.813)
## treatment
                                       -4.279 (9.918)
## S.E. type
                                 IID
                                                   IID
## Observations
                                  34
                                                    34
## R2
                             0.14041
                                               0.00578
## Adj. R2
                                              -0.02529
                             0.11355
```

Then, we want to check with the prefer. It is another model focus on prefer physical book or digital books. We could ninty percent confident that in average for the people who prefer the physical books, the treatment effect is 20. Which means that they will achieve 20 points higher if they take the physical test than if they take the digital test. And for the people who prefer the digital books, the treatment effect is insignificant.

```
##
                         reg.science
                                               reg.art
## Dependent Var.:
                                grade
                                                  grade
##
## (Intercept)
                    46.06*** (6.463) 69.44*** (5.930)
                     25.42** (8.630)
## treatment
                                        -13.71 (9.291)
## S.E. type
                                  IID
                                                    IID
                                                     27
## Observations
                                   41
## R2
                             0.18203
                                                0.08013
## Adj. R2
                              0.16105
                                               0.04333
```

Similar model about major. We are ninty five percent confident that in average, for people who are science major, the treatment effect is 25. Which means that for science major students who take treatment will have 25 points higher grade than science major students who take control. And for people who are art major, the treatment effect is insignificant.

```
##
                         reg.finish1
                                           reg.finish0
## Dependent Var.:
                                grade
                                                  grade
##
                    57.63*** (6.254) 56.33*** (7.156)
## (Intercept)
## treatment
                       14.52 (8.734)
                                         1.810 (10.30)
## S.E. type
                                  IID
                                                    IID
## Observations
                                   39
                                                     29
## R2
                             0.06950
                                                0.00114
## Adj. R2
                             0.04435
                                               -0.03585
```

Both average treatment effect is at 0 star confident level. The difference within finish group is insignificant.

```
## reg.note1 reg.note0
## Dependent Var.: grade grade
##
##
(Intercept) 49.67*** (9.559) 61.09*** (5.364)
## treatment 9.476 (15.75) 7.168 (7.227)
```

```
## S.E. type IID IID
## Observations 19 49
## R2 0.02085 0.02051
## Adj. R2 -0.03674 -0.00033
```

Both average treatment effect is at 0 star confident level. The difference within note group is insignificant.

```
reg.interest1
                                         reg.interest0
## Dependent Var.:
                               grade
                                                 grade
##
                    53.07*** (8.378) 60.21*** (5.157)
## (Intercept)
                                         8.227 (7.627)
## treatment
                       11.49 (11.34)
## S.E. type
                                                    IID
                                  IID
## Observations
                                  33
                                                     35
                             0.03203
                                               0.03405
                             0.00080
## Adj. R2
                                               0.00478
```

Both average treatment effect is at 0 star confident level. The difference within interest group is insignificant.

#### overall analysis for regression

As we can see, the average treatment effect for the overall(general) regression is insignificant. So, we want to check if the treatment is insignificant to every kinds of groups of people. And we notice that only major and prefer group is sensitive to the treatment. So we want to use heterogeneous treatment effects to check their conditional treatment effect.

# Heterogeneous Treatment Effects

Because only prefer and major group have the significant average treatment effect. We used the feols function to compute the heterogeneous treatment effects on grade if the person took the treatment survey and prefer reading physical books.

```
reg_het1 <- feols(grade ~ treatment*prefer, data = df_c, se = 'white')
etable(reg_het1)</pre>
```

```
##
                               reg_het1
## Dependent Var.:
                                  grade
## (Intercept)
                       58.85*** (5.349)
## treatment
                         -4.279 (10.52)
                         -4.350 (10.52)
## prefer
                         24.43. (14.39)
## treatment x prefer
## S.E. type
                       Heterosked.-rob.
## Observations
                                     68
                                0.09647
## R2
                                0.05412
## Adj. R2
```

#### Explanation of the Results

The CATE is 24 however it is insignificant, so we can not say that in average, the people who prefer physical books and take the physical test, will achieve higher grade than, people who prefer digital books and take the physical test.

## Heterogeneous treatment effects on grade - science major.

```
reg_het2 <- feols(grade ~ treatment*major, data = df_c, se = 'white')</pre>
etable(reg_het2)
##
                              reg_het2
## Dependent Var.:
                                  grade
##
                      69.44*** (5.917)
## (Intercept)
## treatment
                        -13.71 (9.190)
## major
                       -23.38* (8.810)
## treatment x major 39.13** (12.64)
## S.E. type
                      Heterosked.-rob.
## Observations
## R.2
                               0.15492
## Adj. R2
                               0.11531
```

#### Explanation of the Results

The CATE is 39.13. And we are 95 confident that in average, people who are science major will achieve 39 points higher grade than people who are art major, if both of them take the physical test.

# Limitations

One of most significant limitations to our regressions would be our sample size. Our sample only consisted of 68 survey results. This is not nearly enough observations to determine a trend on a larger scale so our statistical power is very low, so it also influenced the accuracy of the experiment. Our probability of rejecting the null when there is a true treatment effect of some size is substantially low. The other limitations are related to our environment. For physical book readers, we can only find people we may reach to. It is also one major reason that most our regression results are insignificant. By moral concerns, we did not record each participants' learning ability by the measurement of intelligence quotient, GPA, or reading speed, etc. Since each individuals have different learning speed or knowledge acceptance, this may somehow influence our analysis on effects. Meanwhile, except reading the article during the quiz, we did not restrict participants to use the any device or tool during the quiz time. Therefore, they may have more uncontrollable devices that may help them to get a higher grade.

About the participants, by the restriction of reading article physically, we have to primarily find people who are physically or geographically close to us. So, the majority of the physical book readers are come from our program or people who have some business background. It was also hard for us to control their knowledge base. As same of the issue of participants' background, we invite people who may not around us to do the digital version quiz for the control group. Since most of those participants are not live in Boston, we are hard to control their background also. Also, one limitation is also significant is the attitude of participants. If the participants were not likely follow the restrictions, like reading time or guess the answer, our analysis may be less accurate indeed.

# Conclusion

We designed an experiment to analyze the effects of physical book reading on the grade. We collected a certain variables to help us determine information of participates and how those variables affect their performance on grades, including not only the major, the reading preference, and the interest to the article of participants. In the process of the experiment, we designed to give participants 15 minutes to read the article whether the physical article or digital article. Then, they will take a 6 questions-quiz which were designed in 3 difficulty levels in an unlimited time. After the quiz, they need to fill out a survey about the participants' information or reading status. Collecting the final grade from the server, we use to analyze their performance in grades by assigning in treatment group or control group.

Consequently, after the analysis, we could not find a significant relationship between using physical books and the score of quiz. However, we observes that the participants of physical book have higher average quiz score during this experiment. We also conclude that in the 95% confidence the only one significant relationship is that the science major may have better grade than art major in the treatment group.

# **Biblography**

- Benson, Kerry. "Reading on Paper versus Screens: What's the Difference?" BrainFacts.org, July 22, 2020. https://www.brainfacts.org/neuroscience-in-society/tech-and-the-brain/2020/reading-on-paper-versus-screens-whats-the-difference-072820.
- Dillon, Andrew. "Reading from paper versus screens: A critical review of the empirical literature." Ergonomics 35, no. 10(1992):1297-1326. https://www.researchgate.net/publication/228707100\_Reading from paper versus screens A critical review of the empirical literature
- Sparks, Sarah D. "Reading on Screen vs. Print: New Analysis Thickens the Plot on Promoting Comprehension." Education Week. Education Week, August 24, 2021. https://www.edweek.org/teaching-learning/reading-on-screen-vs-print-new-analysis-thickens-the-plot-on-promoting-comprehension/2021/07.
- Wilde, Sari, et al. "Organizations Need a Dynamic Approach to Teaching People New Skills." https://hbr.org/2021/11/organizations-need-a-dynamic-approach-to-teaching-people-new-skills.

# **Appendix**

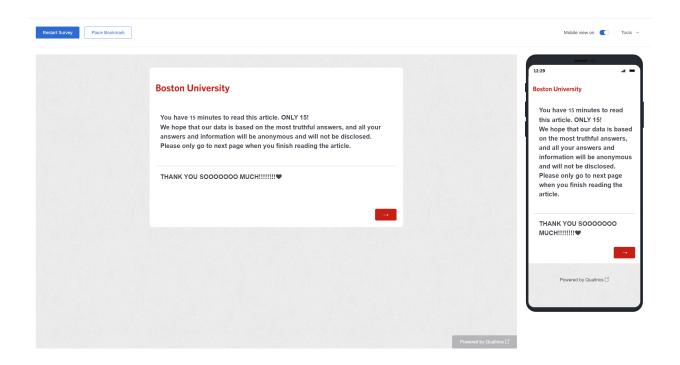


Figure 1: Survey - first page for Control Group

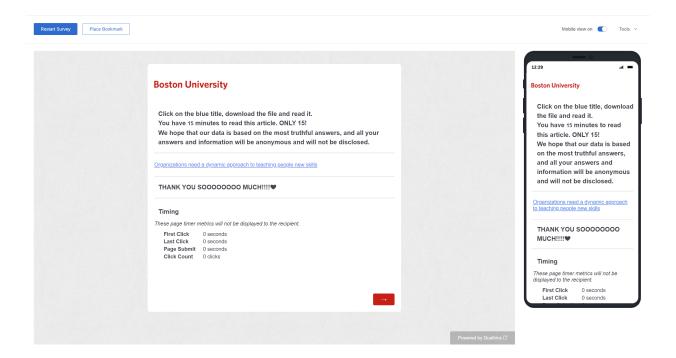


Figure 2: Survey - first page for Treatment Group



Figure 3: Survey - Quiz question 1-6

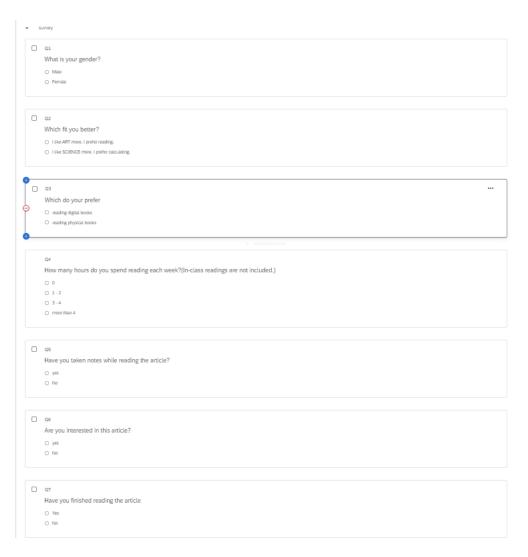


Figure 4: Survey - participant information

#### Harvard Business Review

### **Developing Employees**

# Organizations Need a Dynamic Approach to Teaching People New Skills

by Sari Wilde, Alison Smith, and Sara Clark

November 26, 2021



Jorg Greuel/Getty Images

**Summary.** As industries, organizations, customer needs, and work norms continue to shift and evolve, the need for rapid reskilling and upskilling will only intensify. These challenges require organizations to rethink the boundaries of current solutions to skills gaps. Rather than... **more** 

Figure 5: Article for reading