

Analysis of COVID-19 Impacts to Education Attainment

Introduction In many aspects of life, the pandemic has brought on long-term negative affects. From an educational perspective, the pandemic has widened existing achievement and opportunity gaps between students on a global scale. These effects have been immediate, but the scale will continue to shift as more research is conducted.

As an analyst, supporting learning and development, the impact of COVID-19 to students has been an especially important topic in recent years. One of the major employee benefits is education assistance, and we have noticed an unprecedented change in student behavior in light of the pandemic. In many ways, student behavioral changes are parallel in many activities, with major factors being qualitatively owed to change in caregiver responsibilities, financial hardships, and work changes due to workforce reductions.

As the workforce continues to fluctuate and socioeconomic insecurity remains high, educational attainment and student sentiment at all stages of the educational journey are more important to understand now than ever.

The Problem Statement You Addressed By diving into the qualitative data collected (and artificially created) by researchers, the aim was to better understand the barriers that students are facing in their educational journey in light of the pandemic. Quantifying student behavior through a qualitative lens helps drive policy and programs that can better nurture education or student well-being.

How You Addressed This Problem Statement The problem was addressed through an exploratory data analysis, which should serve as a conversation and project plan for further research and development. The data was publicly collected by research teams across the world with student subjects varying in educational attainment and socioeconomic background.

More will be addressed in the limitations section, but due to the nature of public data, there was not enough datapoints to successfully and confidently build a model. Ideally, a predictive model would be built using historical student data pre-pandemic and to date (understanding the pandemic is not at a close) in order to predict out future student behavior within these new constraints and dependent variables (considering fields such as caregiving responsibilities, access to technology, access to educational funds, etc).

Given the constraints of the data, explanatory modeling was built to look at the current state of student activity and sentiment. Analysis highlights provided below.

Analysis

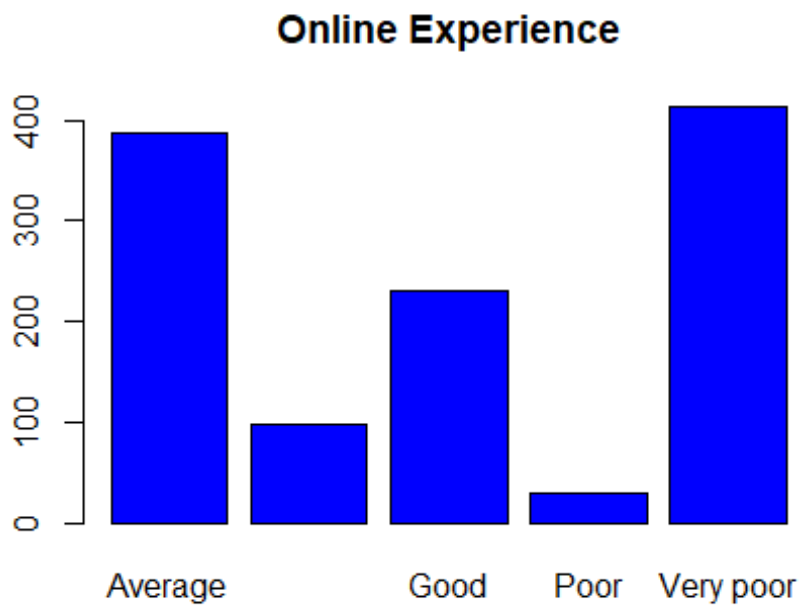
```
## set working directory to dataset
setwd("C:/Users/alexi/OneDrive/Documents/1 Data Science/DSC520")

## read DF
DF1 <- read.csv("COVID-19 Survey Student Responses.csv")
```

```
count <- table(DF1$Rating.of.Online.Class.experience)
count

##
##      Average Excellent      Good      Poor Very poor
##      387          98      230        30      413

barplot(count, col = "blue", main = "Online Experience")
```



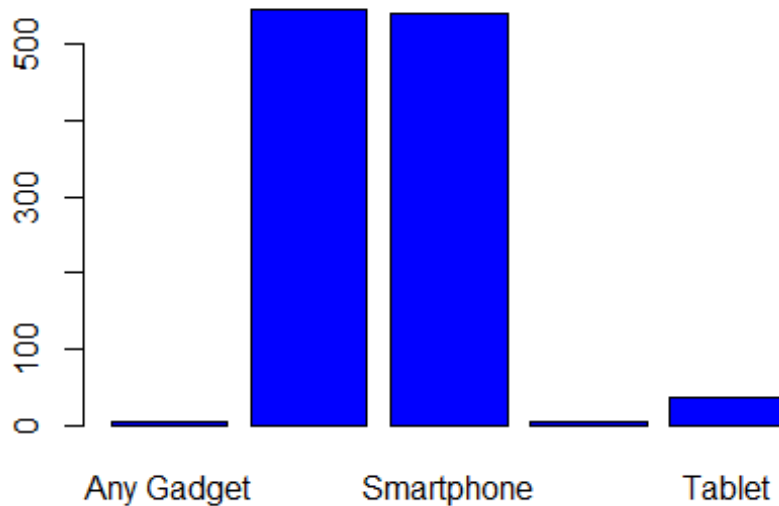
As seen in the above visual, the majority of responses (36%) were very poor. Average, being the most neutral of responses was reported second at 34%. 28% of students reported having a positive experience with online learning.

```
count <- table(DF1$Medium.for.online.class)
count

##
##      Any Gadget      Laptop/Desktop
##      5          545
##      Smartphone Smartphone or Laptop/Desktop
##      539          5
##      Tablet
##      37

barplot(count, col = "blue", main = "Technology Used in Online Learning")
```

Technology Used in Online Learning



In near equal responses, students were relying on a smartphone as often as they were a more traditional laptop/desktop. Presumably, even at the secondary level, a smartphone would not be able to as comfortably support a student's online learning platform the same as a laptop or desktop. This brings into question the spread of technology access across learners and if a disparity in access affects a students' ability to learn at the same level as their peers if access is limited to a smartphone.

```
cor(DF1$Time.spent.on.self.study, DF1$Time.spent.on.Online.Class)
## [1] 0.1193179
```

Despite reporting a positive relationship, the result is small and may not indicate a strong relationship between the amount of time spent in class and the amount of time a student spent studying. Without strong indicators of time, it can be difficult to further quantify relationships between time and ratings provided by students, as the numeric indicators are not significant enough within this population of data. Larger responses from a wider variety of students may be more beneficial, as this population supports just short of 1,200 responses.

This can be further seen when diving into qualitative indicators such as online rating of class. There were no strong causal relationships within responses that would indicate that those who spent more or less time studying or in online class were more or less likely to rate their online learning experience any which way.

Overall, this dataset provided a solid jumping point to look into further student responses during the pandemic, but it is important to note the nuance in responses that less than 20% of students reported any health issues during the pandemic.

Diving into the final dataset, this dataset is a panel dataset that compares three semesters before COVID-19 and “after” COVID-19. This is a constructed dataset but was meant to mimic Portland, Oregon.

```
## set working directory to dataset
setwd("C:/Users/alexi/OneDrive/Documents/1 Data Science/DSC520")

## read DF
DF2 <- read.csv("DF2.csv")

table(DF2$covidpos)

##
##    0    1
## 936 464
```

Similar to the last dataset, approximately 1/3 of respondents reported a positive COVID results. Below, high-level distribution of responses for all fields were evaluated in order to discern any early indicators of interest.

Narrowing into the fields, access to technology is parallel with the former dataset. The average number of computers in this respondent group was 2 with ~12% reporting 0 computers in the household. Since this dataset does not report on alternative technology like the last dataset, it is difficult to definitively say if one population has higher access to technology or not.

Once the baseline is established, the performance over time can be evaluated under the panel dataset, which will be labeled as DF3.

```
DF3 <- read.csv("DF3.csv")

counts <- table(DF3$readingscore)
counts

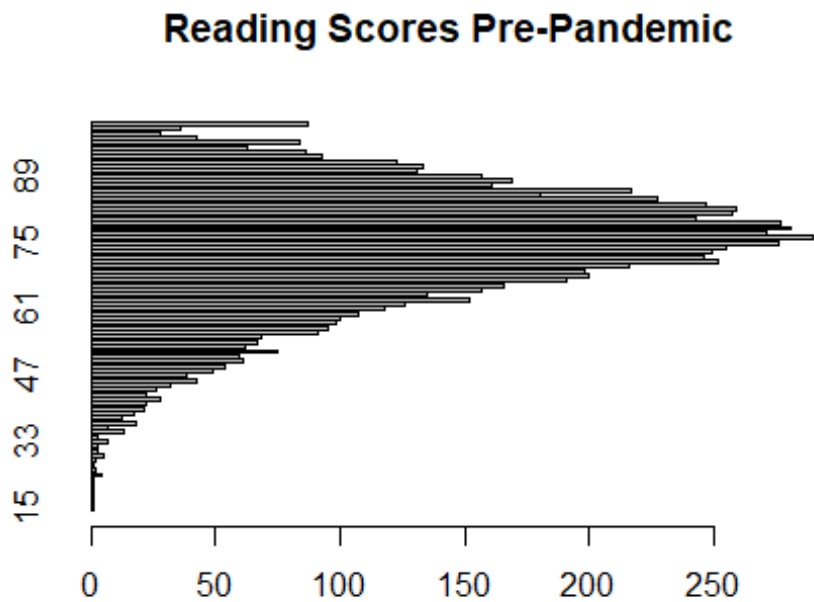
##
## 15 17 20 21 22 23 24 26 27 28 29 30 31 32 33 34 35 36
37 38
## 1 1 1 1 1 1 1 4 2 1 2 5 3 3 7 3 13 7
18 12
## 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
57 58
## 17 21 22 28 22 26 32 42 38 49 54 61 59 75 62 67 68 91
95 98
## 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76
77 78
## 100 107 118 126 152 135 157 166 191 200 198 216 252 246 249 255 276 290
271 281
## 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96
97 98
## 277 243 257 259 247 227 180 217 161 169 157 131 133 123 93 86 63 84
42 28
```

```
## 99 100
## 36 87

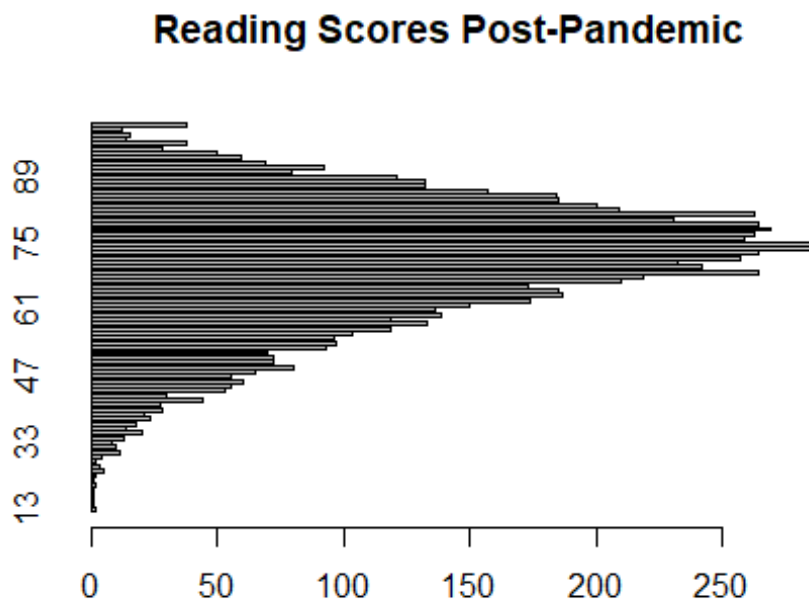
counts2 <- table(DF3$readingscoreSL)
counts2

##
## 13 15 19 21 23 24 25 26 27 28 29 30 31 32 33 34 35 36
37 38
## 2 1 1 1 1 2 1 2 5 3 2 4 11 10 8 13 20 14
18 23
## 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
57 58
## 21 28 27 44 30 53 55 60 55 65 80 72 72 70 93 97 96 103
119 133
## 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76
77 78
## 119 139 136 150 174 187 185 173 210 219 264 242 232 257 264 286 285 259
263 269
## 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96
97 98
## 264 231 263 209 200 185 184 157 132 132 121 79 92 69 59 50 28 38
14 15
## 99 100
## 12 38

barplot(counts, main = "Reading Scores Pre-Pandemic", horiz = TRUE)
```



```
barplot(counts2, main = "Reading Scores Post-Pandemic", horiz = TRUE)
```



Pre-pandemic, the average reading score was 74 with a minimum score of 15, Post-pandemic, the average reading score was 71 with a minimum score of 13. In 3 semesters, average performance diminished by 3 points.

Implications While it's important for students to understand what factors influence their educational journey, it can be assumed that students are aware of their limitations and barriers to success. The targeted audience for this exercise would be leaders and shareholders within the broader learning and development center of excellence, in my case. For other research purposes, educators or education facilitation companies may find better understanding the performance changes within student populations beneficial in shifting marketing messages, individual engagements, and so forth.

In all datasets presented within the milestones of the broader project, it can be seen that student performance and overall attitudes worsened from the start of the pandemic to date. Reasonably, this was to be expected given the general attitude of the collective public over the past several years, but the lasting effects to educational attainment and broader workforce needs to be looked into in order to support the widening education gap between students.

For future research, these results could influence further studies done on varying levels of study, as potential barriers listed in this project may not be the same to students who are adults in higher education.

Limitations To build onto this research, it is important to widen the scope of not only participants but time accounted for. Historical data comparisons to current date are

important to test the assumptions of researchers and what amount of noise or outliers would be anticipated within a population.

Demographic data or more granular geographic data would be beneficial in more completely fleshing out the population of students and how their socioeconomic responses compared to Census data or other privately collected data of either student body populations or workforce demographics.

Lastly, the pandemic is a unique experience to every individual, and it is difficult to make sweeping generalizations without some degree of caveat or exception.

Concluding Remarks Overall, this was an interesting look into the state of students and their educational experience. It will be important to continue looking into any behavioral or performance shifts in the coming years as online learning continues to evolve or students return to traditional classroom experiences.