COVID School Closures and Their Impact on Students*

Cristina Burca, Yan, Sakura

February 12, 2024

Abstract

COVID-19 was an outbreak of virus that forced many instituations to shut down for 2-3 years. Schools were no different and this paper aims to look at the effects of the said closures in school and how it affected the population. This paper finds that with more inperson schooling provided the less the enrollment rates drop(more in depth analysis in later sections).

Introduction

This paper will discuss several ways that the Covid-19 pandemic may have affected schools students and education as a whole. We will be looking at data procuded from all around the world , although we will be focusing on the data collected from the United States of America. First, we will be looking at school closures in the five major continents and some dductions we can make from them. Next, we will look at test scores for

*Code and some data from this paper are available at: github repo the USA in Mathematics as well as ELA(English Learning Arts), and how they changed from year to year, before, during, and after the pandemic. Lastly, the outcomes of this paper are that there is an over all increase in the percent of enrollment as more in person schools are opened in the educational districts of United States. The concept fo educational districts will be described later in the data section for the user's aid but in this context it is not the most important.

We will also discuss the differences in our findings from the original paper as well as addition analasis from the datat provided from their replication package.

Data

Data Source and Collection

We use R Core Team (2022) to make this paper as well as the graphs and topic were taken inspiration from Jack and Oster (2023). Various helpful packages were used in order to clean, sort and graph this paper in a way such that the reader will not have difficulty undertsnading neither the topic nor the data sets of this paper. The packages are, Wickham (2016), Wickham et al. (2019), Wickham et al. (2023), Wickham, Hester, and Bryan (2023), Xie (2014), Firke (2023), Zhu (2021), Wickham, Vaughan, and Girlich (2024), Wickham and Miller (2021), Hyndman and O'Hara-Wild (2021).

Data Cleaning / Methodology

What data set did we clean and why. Explain the variable here too ig The data provided originally was called nces Some data sets cleaned were

The paper's third set of graphs represnt percent change in enrollment by share of in-person offering by district The third set of graphs shown in this paper is the enrollment rates in the different educational districts of the US.

Measurment

The second set of this paper's graphs focused on test scores of students in the United States of America, specifically their score changes in ELA(English Language Arts) and Mathematics. The data this was compared to was the percentage of in person attendance as well as the geographic locale of the students.

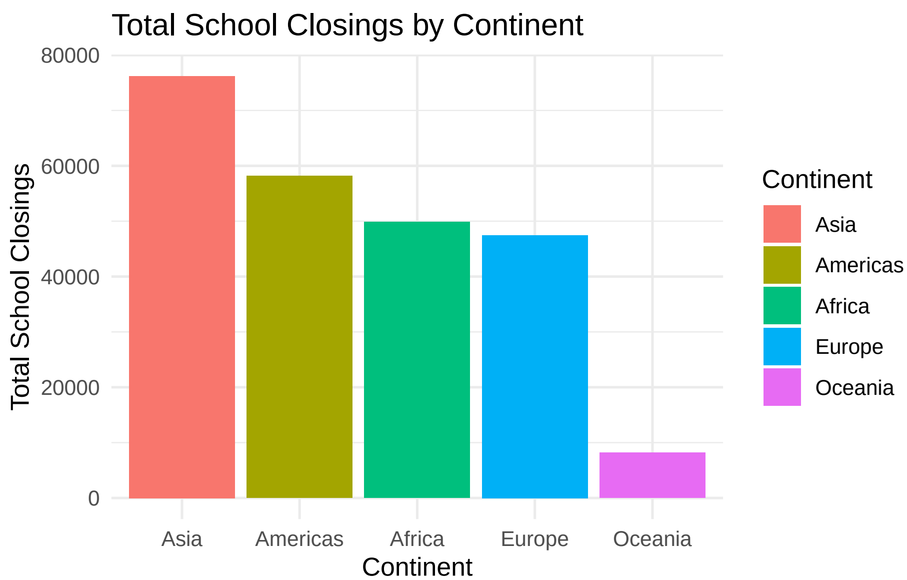
As mentioned before the third claim that this paper looks at is the enrollment rates with respect to the share of in-person learning or virtual learning. For this category there are 2 measurements. The data originally come with the number of enrollments in each leaid for each year however, the data is cleaned up so the reader may see the rates related to it. The x-axis show the share of in-person or virtual schools in the district. The share of in-person was only changed in terms of how this paper binned the data; the share value was from 0 to 100% and it was cut into 11 equal pieces. The measurement for the x-axis is the same as it is on the raw datasheet, which is percentage representing share of districts that were either in-person or virtual, values from 0 to 100%. The measurement for the y axis is the enrollment rate. This has been changed from the raw data set as originally it recorded the number of students enrolled for each grade by leaid for a certain year. As mentioned in the methodology, the original value was changed into a rate then a percentage to better show the reader what the actual changes were. If this step was not done then comparing the declines and the inclines of enrollment rate would have been hard due to the large numbers. Anyhow, the measurement for the y-axis is the enrollment rate in percentage and it doesn't have a limit.

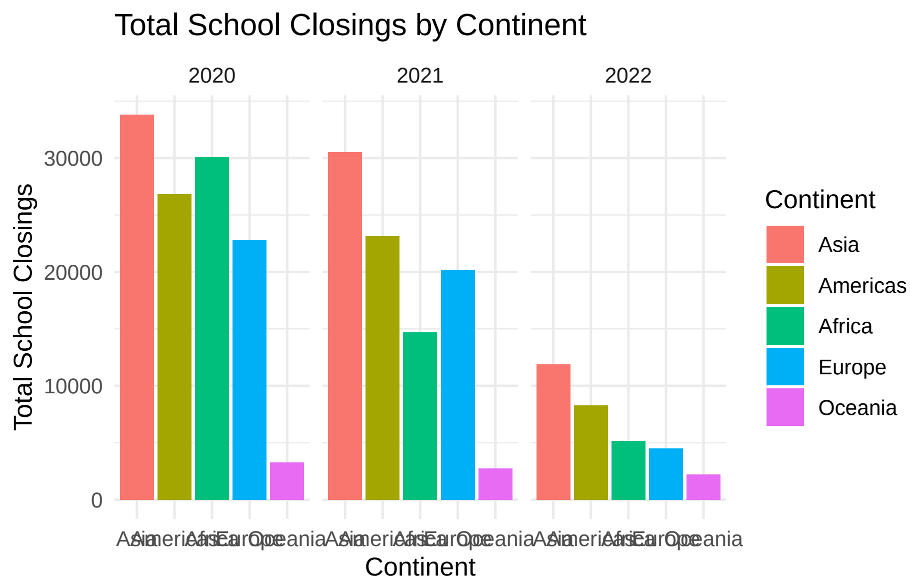
Results

First Graph

As the original paper has divided the countries by Region, this paper has divided the countries by continent, to compare and contrast the findings based on the separation of the countries.

The original paper states that in their findings, the United States had longer terms of school closures, while given by the graph I have produced, it seems





Second Graph

Here we will look at the change in percentage of test score of students between the years of 2017 and 2022. Covid-19 had a major impact on student test scores, whether it be because of the quality of remote learning or the resources the students may have or not have had while learning remotely. the first of the two figures is a near identical replication of the the on in the original paper we are replicating while the second figure is gives us some additional information to analyze that wasn't examined as thoroughly in the orginal paper.

```
#| echo: false
#| warning: false
#| label: fig-InPersonAttendence
```

```r
#| fig-cap: Average Grade Change by In Person Attendence

score_data = read_csv("../../inputs/data/scores_lm_demographics.csv")
```

```
New names:
Rows: 9823 Columns: 63
-- Column specification
-------------------------------------------------------- Delimiter: "," chr
(9): state, DistrictName, subject, lea_name, fips, zip_location, urban... dbl
(51): ...1, county_code, covid_level, year, NCESDistrictID, lunch, miss... lgl
(2): spec_ed_students, english_language_learners date (1): ReportingDate
i Use `spec()` to retrieve the full column specification for this data. i
Specify the column types or set `show_col_types = FALSE` to quiet this message.
* `` -> `...1`
```

```r
clean_score_data_inperson <- score_data |>
  select(subject, change_2017_2018, change_2018_2019, change_2019_2021, change_202
  rename_with(~ sub("^change_(\\d{4})_(\\d{4})$", "Spring_\\2", .), starts_with("c
  mutate(share_inperson_grouped = cut(share_inperson * 100, breaks = seq(0, 100, b

# Pivot the data to a long format
score_data_long_inperson <- clean_score_data_inperson |>
  pivot_longer(cols = starts_with("Spring"), names_to = "time_period", values_to =

# Group by 'subject', 'share_inperson_grouped', and 'time_period', then summarize
score_data_summary_inperson <- score_data_long_inperson |>
  group_by(subject, share_inperson_grouped, time_period) |>
  summarise(
    mean_change = mean(change_score, na.rm = TRUE),
    .groups = 'drop'
  )
```

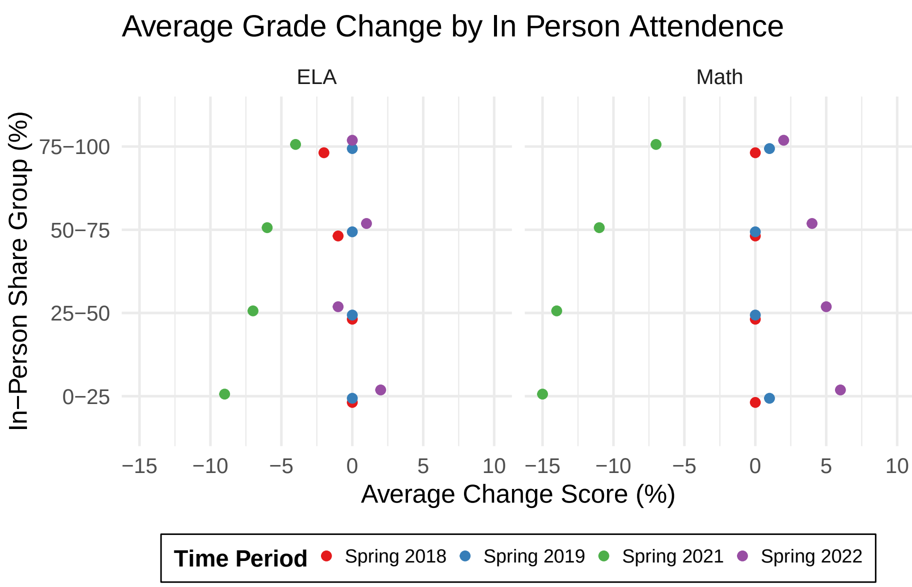
```r
# Now prepare data for the 'urban_centric_locale' grouping
score_data_long_locale <- clean_score_data_inperson |>
  pivot_longer(cols = starts_with("Spring"), names_to = "time_period", values_to =

# Group by 'subject', 'urban_centric_locale', and 'time_period', then summarize
score_data_summary_locale <- score_data_long_locale |>
  group_by(subject, urban_centric_locale, time_period) |>
  summarise(
    mean_change = mean(change_score, na.rm = TRUE),
    .groups = 'drop'
  )

ggplot(score_data_summary_inperson, aes(y = share_inperson_grouped, x = round(mean
  geom_point(position = position_dodge(width = 0.2)) +
  scale_x_continuous(limits = c(-15, 10), breaks = seq(-15, 10, by = 5)) +
  labs(
    title = "Average Grade Change by In Person Attendence",
    y = "In-Person Share Group (%)",
    x = "Average Change Score (%)",
    color = "Time Period"
  ) +
  scale_color_brewer(palette = "Set1", labels = c("Spring 2018", "Spring 2019", "S
  theme_minimal() +
  theme(
    legend.position = "bottom",
    legend.background = element_rect(fill = "white", size = 0.3, linetype = "solic
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 10, face = "bold"),
    legend.key.size = unit(0.2, "cm")
  ) +
  facet_wrap(~subject)
```

```
Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
```

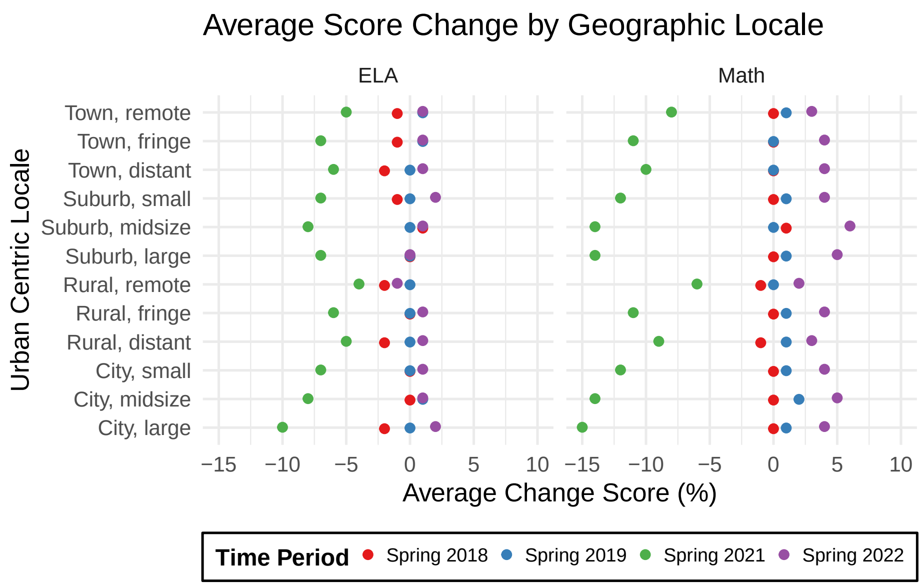
```
i Please use the `linewidth` argument instead.
```



The first observation that can be made is the percent change in test score between Spring 2019 and Spring 2021 where the students who who attended class less than 25% of the time had an average of 15% less in their Math tests and approximately 9% in their ELA tests. Although it can be seen that the more the students attended in person the less their test scores would decrease, never the less the students still saw a decrease in their test scores between the two years even if they had gone to school in person. One thing the original paper failed to include although it was in their data set, was the change in test scores of Spring 2022. While the test scores do increase in Math and ELA, ELA less so than Math, they do not increase as much as they fell in the previous year.

```r
#| echo: false
#| warning: false
#| label: fig-GeoLocale
#| fig-cap: Average Grade Change by In Person Attendence


# Plot for 'urban_centric_locale'
ggplot(score_data_summary_locale, aes(y = urban_centric_locale, x = round(mean_cha
  geom_point(position = position_dodge(width = 0.1)) +
  scale_x_continuous(limits = c(-15, 10), breaks = seq(-15, 10, by = 5)) +
  labs(
    title = "Average Score Change by Geographic Locale",
    y = "Urban Centric Locale",
    x = "Average Change Score (%)",
    color = "Time Period"
  ) +
  scale_color_brewer(palette = "Set1", labels = c("Spring 2018", "Spring 2019", "S
  theme_minimal() +
  theme(
    legend.position = "bottom",
    legend.background = element_rect(fill = "white", size = 0.5, linetype = "solic
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 10, face = "bold"),
    legend.key.size = unit(0.2, "cm")
  ) +
  facet_wrap(~subject)
```



Here we will look at the Geographic Locales of students, and how their grade have changed based on what type of municipality they live in. We can observe that Cities and suburbs of all sizes both in Math and ELA Scores. This may be due to a variety of reasons among them the Covid-19 in each of the specific localities. As cities are densely populated they are more likely to have had stricter lock down rules during the outbreak of Covid-19. Furthermore, suburbs are usually near cities and would likely have similar lock down rules to the cities they are near.

Enrollment rate in the US

It is no secret that COVID-19 affected the school enrollment rates, when the students were forced to learn in a virtual environment. In this section the paper will uncover any trends in enrollment rates with relation to the number of in-person learning offered in each district. For all of the graphs below the y axis will be from -0.12 to 0.05 to ensure reader us able to comapre the trends right beside each other.

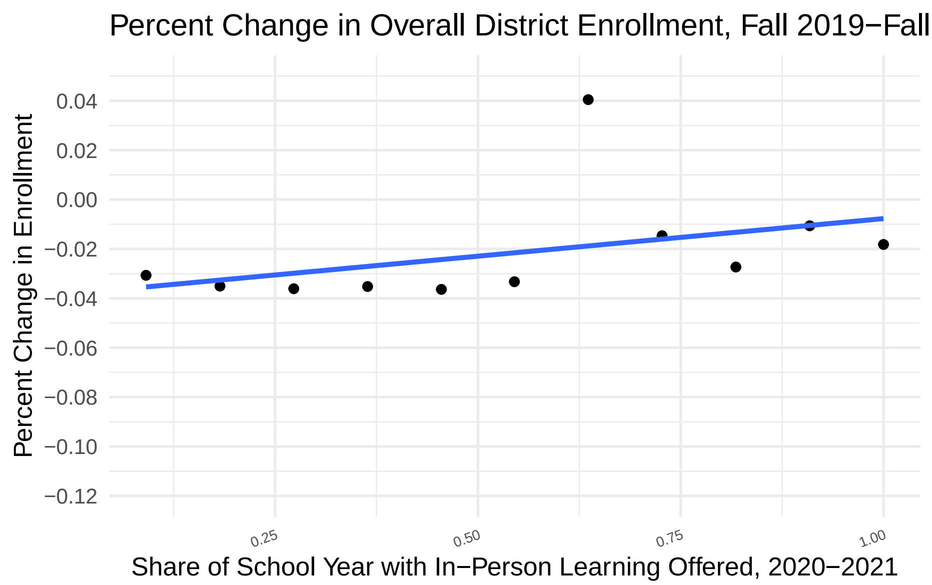


Figure 1: The weighted mean of the percent of overall district enrollment was calulated and graphed. The weights were decided on the bins which were cut in eleven sections based on share of in person school overall.

First, we take a look at the overall enrollment through out the districts. We can see that most values are in the negatives with one value in the positives in the [0.545, 0.6360] bin. We however see an upward trend signifying that we have a a positive change in the percent of rate of change. This means that the enrollment rate increases with more shares of the district going in person. Diving in deeper to look at the values (provided in the Appendix).The graph shows that for 0% to 9% of in-person schooling, we have a decrease of 3% in the enrollment rate. For 9% to 55% share or in person schooling, the rate of enrollment is around -35% meaning there is decrease of 35% and the rate goes to 4% increase in enrollment rate for 55% to 64% of in person schooling in the district. Then the value drops to 1% decrease in enrollment rate when the share of in person offered is 73% to 64% approximately. Then it jumps to a 2% decrease in enrollment then it levels out to 1% decrease in enrollment rate for the last 73% to 100% share in in-person learning. The overall trend is increasing but there are brief moments of decreasing enrollment rates.

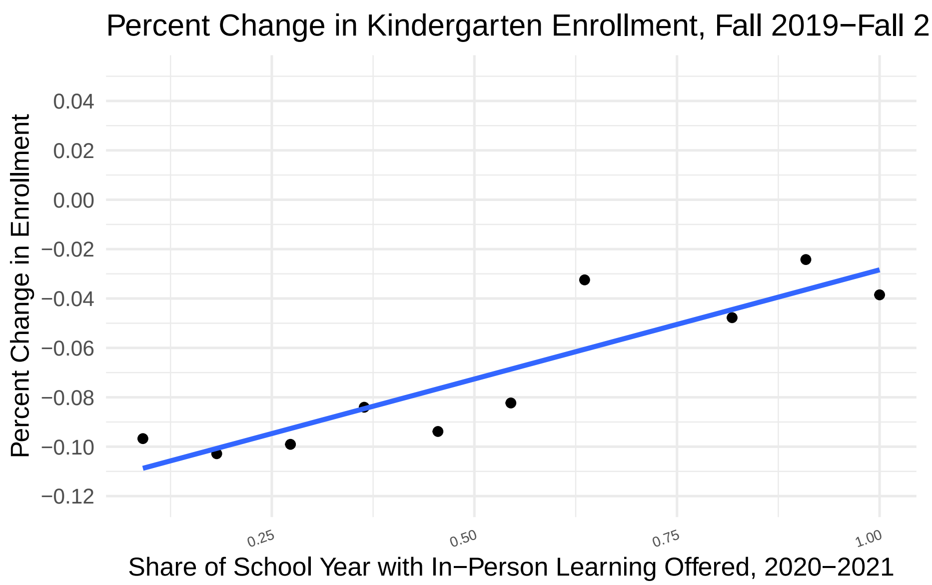


Figure 2: The weighted mean of the percent of kindergarden enrollment in all districts was calulated and graphed. The weights were decided on the bins which were cut in eleven sections based on share of in person school overall.

In comparison we look at the kindergarten enrollment in the same year. Looking at the values individually, the graph shows that for the initial shares 9%, 18% and 27%, we get an enrollment rate of kindergartners in the US is -9%, -10% and 10% respectively. They all show decreasing enrollment rate. The for the 36% of in-person schooling we get a 8.4% enrollment rate and there is a decrease of 9.4% and 8.2% in enrollment rate when there is 45% and 55% share of in person school respectively, for kindergartens. Then the enrollment rates increase to -3.2% and 7% for the proceeding 64% and 73% of in person school share. Afterwards there is a decrease 4.8% in enrollment rates when 82% of schools are in person. Then there is another decrease to -2.4% in enrollment rate when there is 91% share of in person rates. And lastly for 91% to 100% of the in person shares the enrollment rates are -3.9%. This means that although the enrollment rates are negative for the most part there is an increasing trend. This means that the enrollment rates were increasing as more kindergartners were attending in person. Also compared to Figure 1, the slope of the line of best fit is steeper meaning that the rate of increas in enrollment rates is higher. More on the reason to be discussed in later sections.

The purpose of this paper is to compare and analyse the outcomes of the school closures during the pandemic to better understand the impact of it. To see the comparison, this paper also analyses the enrollment rate in 2019. The graph shows that the initial 9% share of in-person school has a decrease of 1% in enrollment.These values stay relatively the same - hovering around the -0.10% to -0.019% enrollment rate- until the 55% to 64% share of in-person school where the enrollment rate jumps to 0.02. Then in the 73th percentile, we see a decrease to -1% in enrollment rate but the procceeding values even out to approximatly 0$ (or more specifically 0.001% and 0.004%). Although the values are negative for the large portion of the graph, there is still an upward trend we saw in previous graphs where the more proprtion of in-person schooling, the larger the enrollment rate.

Lastly, in order to properly make a conclusion about enrollment rates and in-person schooling the reader must also see the opposite correlation, hence the paper looks at the enrollment rates for share of virtual learning as well, above. The graphs shows that for the first 9% of schools doing virtual learning in the districts have a enrollment rate of -2%, which increase to

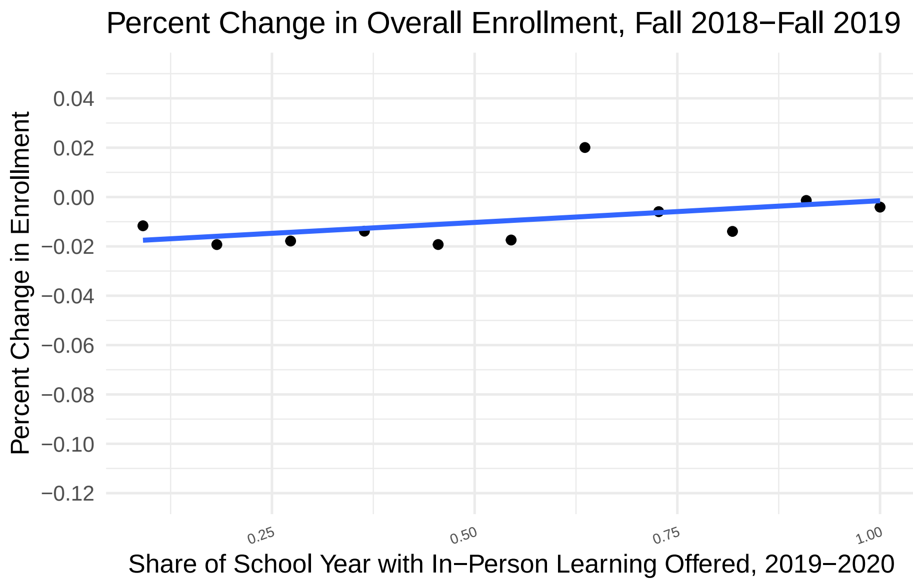


Figure 3: The weighted mean of the percent of total enrollment in 2019 accross all districts was calulated and graphed. The weights were decided on the bins which were cut in eleven sections based on share of in person school overall.

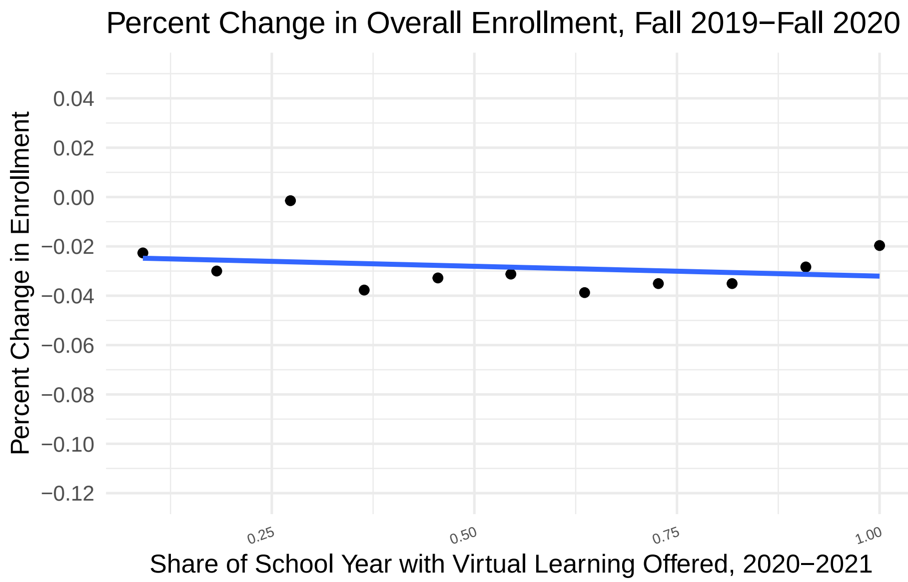


Figure 4: The weighted mean of the percent of total enrollment in 2020 accross all districts was calulated and graphed. The weights were decided on the bins which were cut in eleven sections based on share of virtual school overall.

approximately -3% in the 18% share of virtual school. Then the value jumps up to the maximum value in the entire graph at 0.001% decrease in enrollment rate when 18% to 27% of schools are virtual. Then for the shares from 27% to 72% we see that the enrollment rate fluctuates between 3.1 to 3.7% decrease in enrollment rate where the change is so small the difference is not changing the graph at all. Finally we see a slight increase in enrollment rate with -2.8% and then proceeding with an even higher enrollment rate of -1.9% for 90% and 100% share of virtual school respectively. Looking at the graph overall, the line of best fit seems to have a negative slop compared to the line of best fit computed for the other graphs.

Discussion

Interesting point 1

Change in Test Scores of 2022

In the near identical Replication of the graph in the original paper(**?@fig-InPersonAttendence**), we have included the year 2022 in the change in test scores, where the original paper failed to do so. This could could be due to a variety of reasons, among them possibly the fact that the original authors may have thought it to be unimportant, or they may have been trying to make the change in test scores to seem more permanent, as if the the virus had more long term effect. While the test scores did not increase as much as they fell, they increased never the less, although we do not have data for spring 2023, it is possible we could have made further conclusions with this additional information.

Change in enrollment rate in kindergarden vs all

The third set of graphs that were shown in the results section was graphs that showed the change in enrollment with the shares of in-person school that started (Figure 1, Figure 2). As discussed in the section it seemed like the overall trend of the graph was that enrollment rates were higher the more in-person school there was. This was seen in both cases but more predominantly in the kindergarten graph (Figure 2). This result was interesting as this shows with younger ages the enrollment rates being higher means the children enjoy going to school in person. This is important as kindergarten is when most children start developing the fundamental social skills required to interact with other people.

Ethics and Bias could talk about mental health maybe but it might apply to other "interesting point"

weakness and limitations

Do our own part (how to solve it!)

Furthur research

As stated previously we do not have data for the test scores in 2023, which could give us additional information about how the pandemic affected students grades, as well as the younger students who were in elementary and middle schools when the pandemic began and how their educational foundations may have been affected.

how to solve the limitations

Furthermore, the best way to see the how test scores would change would be to wait several years and analyze the data that would be collected after a few ages complete these tests. While in the short term the test scores of Spring 2023 would be benefitial information to have before making any more conclusions than we have.

Furthur questions?

Appendix

Table 1: Enrollment rates throughout all districts in the U.S. 2020 catagorized by share of inperson learning

Share of inperson	average total enrollment rate
0.0909	-0.0306700
0.1820	-0.0350468
0.2730	-0.0360982
0.3640	-0.0352125
0.4550	-0.0363727
0.5450	-0.0332659
0.6360	0.0404678
0.7270	-0.0146317
0.8180	-0.0273175
0.9090	-0.0106241
1.0000	-0.0181826

Table 2: Enrollment rates for kindergartners throughout all districts in the U.S. 2020 catagorized by share of inperson learning

Share of inperson	average kindergarten enrollment rate
0.0909	-0.0967289
0.1820	-0.1028427
0.2730	-0.0990449
0.3640	-0.0840324
0.4550	-0.0938456
0.5450	-0.0822999
0.6360	-0.0324613
0.7270	0.0748270
0.8180	-0.0477577
0.9090	-0.0242239
1.0000	-0.0385117

Table 3: Enrollment rates throughout all districts in the U.S. 2019 catagorized by share of inperson learning

Share of inperson	average enrollment rate
0.0909	-0.0116537
0.1820	-0.0192675
0.2730	-0.0178225
0.3640	-0.0138305
0.4550	-0.0192475
0.5450	-0.0174195
0.6360	0.0200701
0.7270	-0.0059215
0.8180	-0.0139261
0.9090	-0.0013853
1.0000	-0.0040847

Table 4: Enrollment rates throughout all districts in the U.S. 2020 catagorized by share of virtual learning

Share of virtual	average enrollment rate
0.0909	-0.0226540
0.1820	-0.0299878
0.2730	-0.0014796
0.3640	-0.0377017
0.4550	-0.0327944
0.5450	-0.0312298
0.6360	-0.0387331
0.7270	-0.0350839
0.8180	-0.0350736
0.9090	-0.0283413
1.0000	-0.0196509

Reference

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Hyndman, Rob, and Mitchell O'Hara-Wild. 2021. *Tsibble: Tidy Temporal Data Frames and Tools*. https://CRAN.R-project.org/package=tsibble.

Jack, Rebecca, and Emily Oster. 2023. "COVID-19, School Closures, and Outcomes." *Journal of Economic Perspectives* 37 (4): 51–70. https://doi.org/10.1257/jep.37.4.51.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy

D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.