

# Assessing Factors Affecting Student's Performance during COVID-19 Pandemic\*

Data from USA

Vanshika Vanshika

Shivank Goel

Navya Hooda

February 15, 2024

We analyze data from 11 U.S. states to assess the correlation between factors that may be indicative of student performance during COVID-19 learning disruptions. This paper replicates data analysis upon pandemic schooling and its affect on student test scores, and further attempts to understand other mechanisms that led to learning losses during this period. Several characteristics were related to student performance: mode of learning, loss of COVID-19 case rates, employment and student-to-staff ratio. The learning losses were found to be around 13–14 percentage points in math and approximately 8 percentage points in ELA between 2019 and 2021 (Jack and Oster 2023). In particular we discover that all factors except student-to-staff ratio were relevant predictors and contributed to noticeable differences when compared to 2019 (pre-covid), these findings suggest that the factors that normally contribute to student performance had exacerbated effects during COVID-19 as they led to a larger magnitude of learning losses.

## Table of contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Introduction</b>                                   | <b>2</b> |
| <b>2</b> | <b>Data</b>   | <b>4</b> |
| 2.1      | Data Source and Collection . . . . .                  | 4        |
| 2.2      | Data Cleaning and Processing . . . . .                | 5        |
| 2.2.1    | Student-to-Staff Ratio Across the 11 States . . . . . | 5        |
| 2.3      | Data Modifications . . . . .                          | 6        |

---

\*Code and data are available at: <https://github.com/shivankgoel003/Student-Performance-COVID-19-US-States-Factors>; Minimal and Full Replication on Social Science Reproduction platform available at <https://www.socialsciencereproduction.org/reproductions/1470/published/index>

|          |   |           |
|----------|---|-----------|
| <b>3</b> | <b>Results</b>  | <b>6</b>  |
| 3.1      | Impact of Learning Model . . . . .                            | 7         |
| 3.2      | Impact of COVID-19 Case Rates . . . . .                       | 10        |
| 3.3      | Impact of Staff Availability . . . . .                        | 13        |
| 3.4      | Impact of Loss of Employment Income on Passing Rate . . . . . | 14        |
| <b>4</b> | <b>Discussion</b>   | <b>14</b> |
| 4.1      | Findings . . . . .  | 14        |
| 4.2      | Accounting for Bias . . . . .                                 | 16        |
| 4.3      | Limitations . . . . .   | 17        |
| 4.4      | Future Research . . . . .                                     | 18        |
|          | <b>References</b>   | <b>19</b> |

# 1 Introduction

The COVID-19 pandemic has been one of the most challenging times globally in recent decades. The measures taken in place to reduce the transmission across the world were unprecedented but necessary in curbing the spread of disease. However, the cost of these measures posed disruption in every part of life whether it be businesses, healthcare, or education. For young people in particular, COVID-19 posed disruptions and threatened their sense of normalcy. The transition to COVID-19 learning manifested an environment of uncertainty and stress at a time of critical psychological development (Imogen H. Bell and Thompsona 2023).

The COVID-19 pandemic has created the largest disruption of education systems in history, affecting nearly 1.6 billion learners in more than 190 countries and all continents (Card 2023). COVID-19 led to rapid adaptations in learning models across schools and districts in the United States. From 2020 to 2021, educational institutions faced unprecedented challenges, transitioning between in-person, hybrid, and virtual learning models in response to evolving public health guidelines and infection rates. This shift in learning models has had severe impacts on student performance unlike anything we’ve seen in recent years. As reported by NAEP, from 2019 to 2022, the average performance on the National Assessment of Educational Progress (Nations 2020) declined by 8 points in 8-th grade math, representing approximately 75% of a grade level, and by 3 points in eighth-grade reading, equivalent to roughly 25% of a grade level. While the pandemic was bound to impact learning outcomes of students, its likely that various factors played a role in the overall results, and not just the pandemic alone.

In normal times, student performances have been studied and analyzed to highlight strong indicators of student performance. Among these important factors affecting students’ academic achievement are their economic background and conditions, personal health conditions, and teacher involvement. In particular, it is suggested that the adequacy of the physical conditions that make up the school, such as the information technology facilities, location, library and media facilities, and schoolyard are attributes important role in the student’s academic

achievement. (Ozcan (2021)) In the context of COVID-19 schooling, all these factors were impacted in some way. The physicality of a school environment was shifted to a virtual hybrid space, students' health whether their own or of those around them was impacted due to the transmission of COVID-19, the pandemic brought in employment losses, and teachers were expected to rapidly provide learning in new formats. This changing dynamic raises the question of how these factors impacted student performance during COVID-19.

In this paper, we embark on a comprehensive analysis to explore the estimand whether factors like loss of employment, COVID case rates, student-to-staff ratio in virtual learning, and the mode of learning had a significant impact on student performance, measured by the pass rates of students in grades 3-8 across the 11 U.S. states namely Ohio (OH), Colorado (CO), Connecticut (CT), Minnesota (MN), Massachusetts (MA), Mississippi (MS), Rhode Island (RI), Virginia (VA), Wyoming (WY), Wisconsin (WI), and West Virginia (WV). Through examination, we show the transformative impact of the pandemic on the educational landscape across these states. Our analytical framework explores whether the patterns of the factors stated above contributed to the decline in test scores between 2019 and 2021, shedding light on the nuanced associations between the two.

We explore the origin of this decline in detail by using the extensive dataset from the Household Pulse Survey (Bureau 2020), searching for household dynamics that may have caused students' lower academic performance and states' diminished passing rates. Our findings reveal that the learning losses found during the pandemic were both substantial and highly heterogeneous across communities. We claim that these losses cannot be attributed to one cause but rather stem from a variety of diverse factors as we stated previously. Notably, the loss of household income during the pandemic exhibited a direct correlation with learning loss, with mathematics suffering a more significant decline compared to ELA courses. Additionally, we also explored the correlation between passing rates in each state and the type of learning mode used popularly, which allowed us to demonstrate the varying impacts of in-person, hybrid, or virtual models on student pass rates. Furthermore, we highlight the significant effect of varying COVID-19 case rates on learning outcomes, in general, we notice that states experiencing higher infection rates manifested greater learning losses compared to their counterparts with lower case counts.

Recognizing the intricate effects of the pandemic on education is vital for crafting effective strategies moving forward. By grasping the various factors affecting student learning amidst these challenges, we can provide schools and districts with the tools they need to overcome obstacles and ensure every student receives the necessary support to succeed academically and beyond. We have the opportunity to strengthen the education system to counteract learning setbacks and pave the way for a brighter future for all students.

This paper is structured as follows. In the Data Section, we denote the sources of the datasets utilized in the paper we are replicating, the specific techniques we used to process them, and highlight their key factors. In the Results Section we dive into the findings we discovered following replication and workings of the original paper, we assessed how they were relevant to our study of the factors contributing to the decline in student performance. In the Discussion

Section, we address biases and weaknesses in the data that contribute to our findings, and how we approached analyzing those areas despite the limitations.

## 2 Data

### 2.1 Data Source and Collection

The data utilized in this study is obtained from the compilation of district-level schooling models and state standardized assessment data, originally featured in the paper “Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts” (Jack and Oster 2023) published in the American Economic Review: Insights in June 2023. (AEA 2022). The primary dataset encompasses schooling mode data from the 2020–2021 academic year across 11 U.S. states, integrating various educational approaches during the pandemic, including in-person, hybrid, and virtual learning environments. These datasets were sourced from COVID-19 School Data Hub (USA 2022), district-level state standardized assessment data from spring 2016–2019 and 2021, and additional data demographic statistics from the National Center for Education Statistics (Education Statistics (2023)) to establish a comprehensive analytical framework. We also used additional year 2020 household data by Household Pulse Survey (Bureau 2020) from U.S. Census Bureau spanning weeks 1 to 12.

Our analyses incorporate three categories of data: district-level schooling modes, state standardized assessment results from the academic years 2016–2019 and 2021, and auxiliary that encompass district demographics and county-level variables. The following subsections detail the sources, collection methodologies, and data cleaning procedures undertaken to ensure the accuracy and reliability of the datasets for our analysis.

**State Score Data:** This dataset encompasses state-level academic performance statistics from 2016 to 2020. Variables include the proportion of in-person, virtual, and hybrid learning, participation rates, standardized test pass rates, COVID-19 case rates per 100,000 in school zip codes, peak monthly cases, total enrollment figures, and political voting shares by district. These variables are measured by the standardized test scores of each state. Each state records their scores differently and on different websites, so the methodology of measurements is unknown.

**Learning Model Data:** This dataset details the learning models adopted by school districts in Colorado, Connecticut, Ohio, Virginia, West Virginia, Wyoming, Mississippi, Rhode Island, Minnesota, Massachusetts, and Wisconsin. Each state’s data outlines the specific learning model (in-person, virtual, hybrid) employed over time, enrollment figures, and staffing counts. These variables are measured by NCES Demographic Data. Both district- and school-level demographic data from the National Center for Education Statistics Common Core of Data (<https://nces.ed.gov/ccd/>) were accessed through Urban Institute’s Education Data Portal using Stata 16 (<https://educationdata.urban.org/documentation/>). Data documentation can be found at the website above. COVID-19 Community Case Rate Data: We

included COVID-19 Community Case Rate Data by District from the COVID-19 School Data Hub website accessed through the “Download Community Case Rate Data” in the Community Case Rate Data by District box on the ‘For Researchers’ page of the CSDH website ([https://www.covidschooldatahub.com/for\\_researchers](https://www.covidschooldatahub.com/for_researchers)).

## 2.2 Data Cleaning and Processing

We used R (R Core Team 2022) for data cleaning and processing, utilizing packages like tidyverse (Wickham et al. 2019) for data manipulation and janitor (Firke 2023) for cleaning column names. Other packages used includes ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), readr (Wickham, Hester, and Bryan 2024), tibble (Müller and Wickham 2023), janitor (Firke 2023), reshape2 (Wickham 2007), knitr (Xie 2023), ggbeeswarm (Clarke, Sherrill-Mix, and Dawson 2023), ggrepel (Slowikowski 2024), kableExtra (Zhu 2024) and here (Müller 2020). The cleaning process involved standardizing learning model categories, selecting columns of interest, and simplifying data for analysis. We integrated data from different states, and made sure that variations in reporting and measurement were addressed while combining them into one. For instance, some states categorized their schooling mode as “In-person,” while others used “In-Person.” To ensure they are equivalent, we standardized these terms. We also conducted a check for missing data values, and imputed or excluded them as appropriate based on the context. The variables were also renamed for clarity and consistency across different states’ datasets. A sample of cleaned state score data can be seen in Table 1.

Table 1: Sample of Cleaned State Score Data

| State | Year | Share In-person | Share Virtual | Share Hybrid | Participation | Pass Rate | Covid cases rate per 100k | Total Enrollment |
|-------|------|-----------------|---------------|--------------|---------------|-----------|---------------------------|------------------|
| CO    | 2016 | 0.8586572       | 0.0600707     | 0.0812721    | 86.28333      | 0.3296667 | 29.04677                  | 1132             |
| CO    | 2016 | 0.9399293       | 0.0600707     | 0.0000000    | 91.75000      | 0.2460000 | 31.04686                  | 8723             |
| CO    | 2016 | 0.0000000       | 0.2756184     | 0.7243816    | 95.86667      | 0.2526667 | 31.53898                  | 18657            |
| CO    | 2016 | 0.0000000       | 0.5759717     | 0.4240283    | 98.58334      | 0.0950000 | 31.54341                  | 5725             |
| CO    | 2016 | 0.2155477       | 0.0600707     | 0.7243816    | 96.35000      | 0.3298333 | 29.81067                  | 1001             |
| CO    | 2016 | 0.2155477       | 0.4664311     | 0.3180212    | 97.26667      | 0.3730000 | 31.54496                  | 35817            |

### 2.2.1 Student-to-Staff Ratio Across the 11 States

This variable was not directly provided in our original paper, to retrieve this relationship we required 2 variables, average pass rates across each state, and the average student-to-staff ratio for each state. To find these we had to use 2 different types of datasets to finally combine them to reach our variable of interest. We first used “cleaned\_state\_scoredata.dta” in .csv format, this dataset contains pass rates for various districts across the 11 U.S. states. Since we require an average across the states we grouped all the state pass rates in 2021 for each state by filtering and calculated it by dividing the pass rate sum by the number of total pass rates reported in one specific state, this yields the average pass rate for each state as needed. We then proceed to find the student-to-staff ratio. This information is specific to each state and found in the replication data. We use each state’s xlsx file, which contains the variables

Table 2: Sample of Cleaned Student-to-Staff Ratio

| State | Pass Rate | Avg. Students per Teacher |
|-------|-----------|---------------------------|
| CO    | 32.31432  | 6.548523                  |
| CT    | 50.27606  | 9.755664                  |
| MA    | 43.86268  | 6.736223                  |
| MN    | 46.03220  | 7.324085                  |
| MS    | 59.95366  | NA                        |
| OH    | 58.88783  | 4.858887                  |
| RI    | 34.79347  | NA                        |
| VA    | 54.15240  | 11.150020                 |
| WI    | 38.65995  | 6.557867                  |
| WV    | 32.07227  | 6.190121                  |
| WY    | 51.90012  | 3.579474                  |

‘enrollment\_total’ and ‘staff\_count’. The statistics are reported biweekly in this data, so we choose to aggregate them by finding all unique values for enrollment and teacher numbers, with that we find the ratio by dividing the total number of students by staff count. We do this calculation for each state by looping through each file and yielding a simple average for each state are newly created and cleaned results seen in Table 2.

## 2.3 Data Modifications

In this study, we developed a unique dataset by carefully selecting and combining data from the Household Pulse Survey (Bureau 2020), available from the U.S. Census Bureau for weeks 1 to 12. This data was measured by using the survey to get both qualitative and quantitative data. This process involved choosing key variables that were crucial for our research. We aimed to analyze the connection between economic changes, particularly the loss of employment income, and their impact on educational outcomes, focusing specifically on school passing rates. This helped us to understand the relationship between these economic factors and educational performance. To ensure consistency, we standardized terminologies and measurements, and aligned variables such as timeframes and demographic categories across the datasets. For example, we combined the state score data with the weekly household data, aligned variables relevant to our study. The merged data file was thereafter saved as a CSV file to study the trends later on.

## 3 Results

Here are two comprehensive overviews detailing passing rate trends. Furthermore, we delve into the factors contributing to these trends.

1. Passing Rate Trends Over Different Years and Subjects: Figure 2 shows the decline in passing rates in English Language Arts (ELA) and Mathematics from 2016 to 2021, with

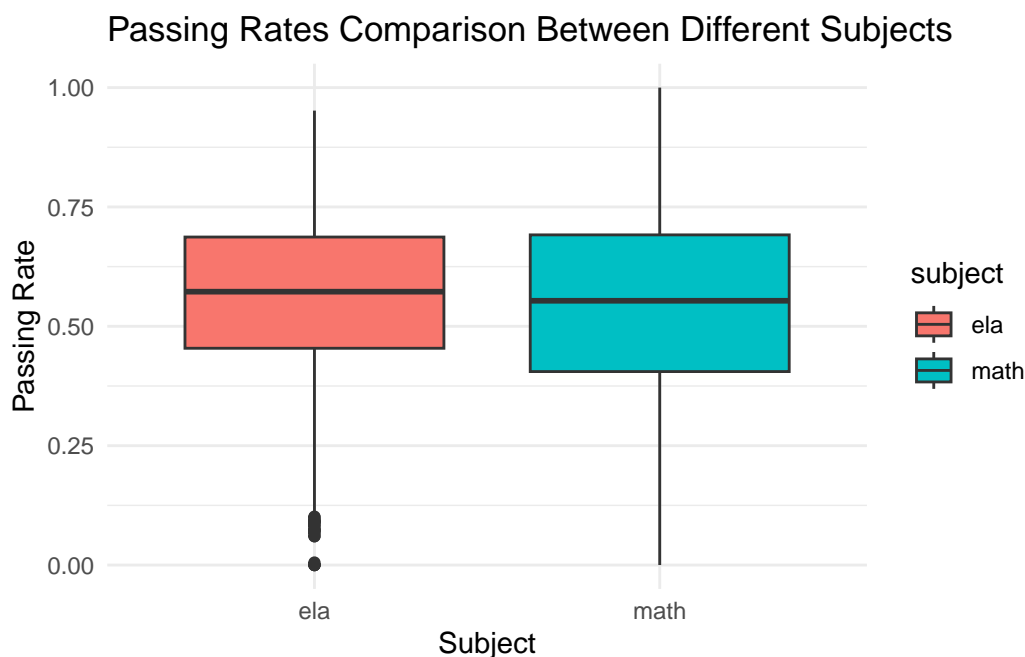


Figure 1: Comparison between Maths and ELA

larger decline in maths than ELA Figure 1, depicting a consistent downward trend across both subjects especially during the pandemic period from 2019 to 2021.

2. Variations in Passing Rates Between Different States: The box plot comparison Figure 3 across states showcases the wide range of passing rates, with some states showing a widespread, which indicates the variability in performance within those states.

### 3.1 Impact of Learning Model

As the COVID-19 pandemic changed the educational model, schools across the United States adopted various learning methods to continue educating students such as virtual, hybrid, and in-person learning. This segment of our study focuses on how these learning models were implemented across different states and their impact on the academic performance of students.

We visualized the distribution of learning models across states using box plots. The box plot for in-person learning ,Figure 4, reveals a wide range of engagement across states, with Virginia showing a median of 0 percent in-person learning from 2019 to 2021, whereas Wyoming had a median of approximately 80 percent in-person learning.

The virtual learning model box plot Figure 5 showed a significant number of outliers, suggesting that certain states had districts with a notably high share of virtual learning. The states

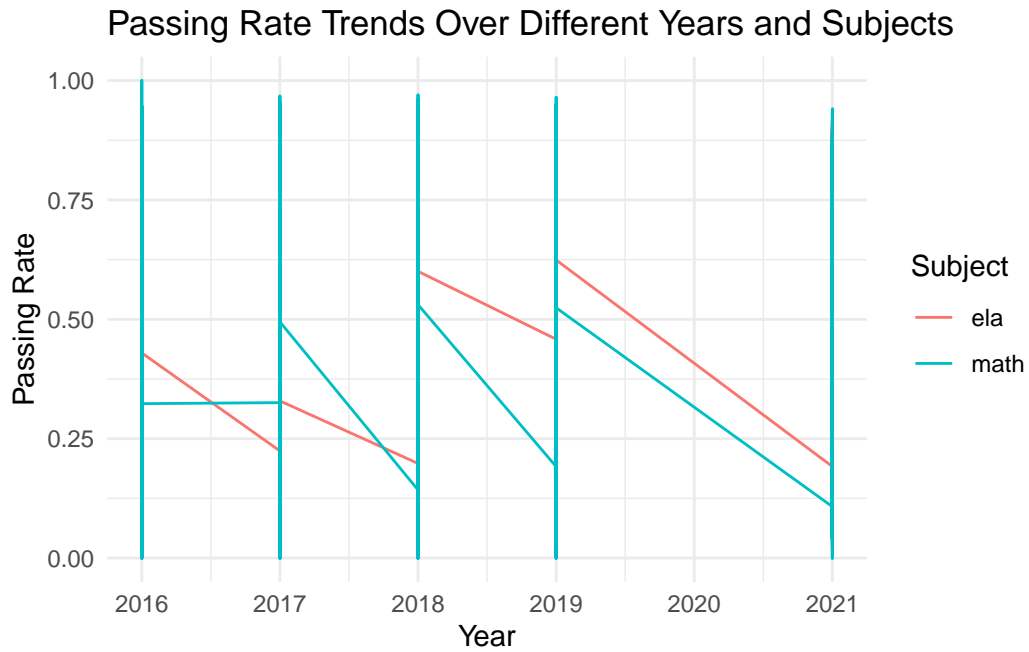


Figure 2: Passing Rate Trend

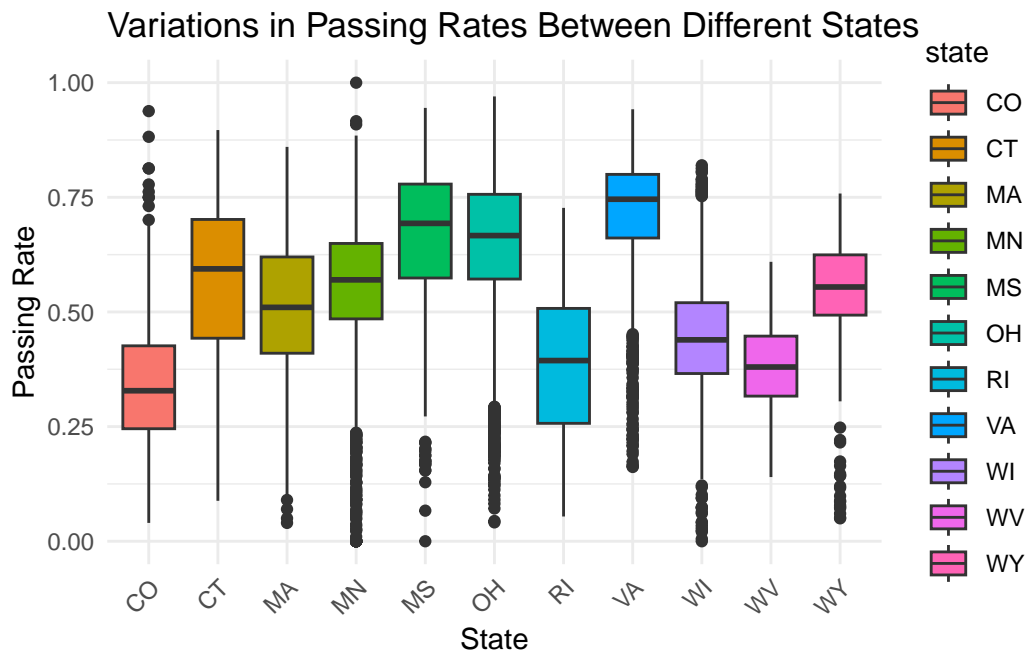


Figure 3: Variations in Passing Rates Between Different States



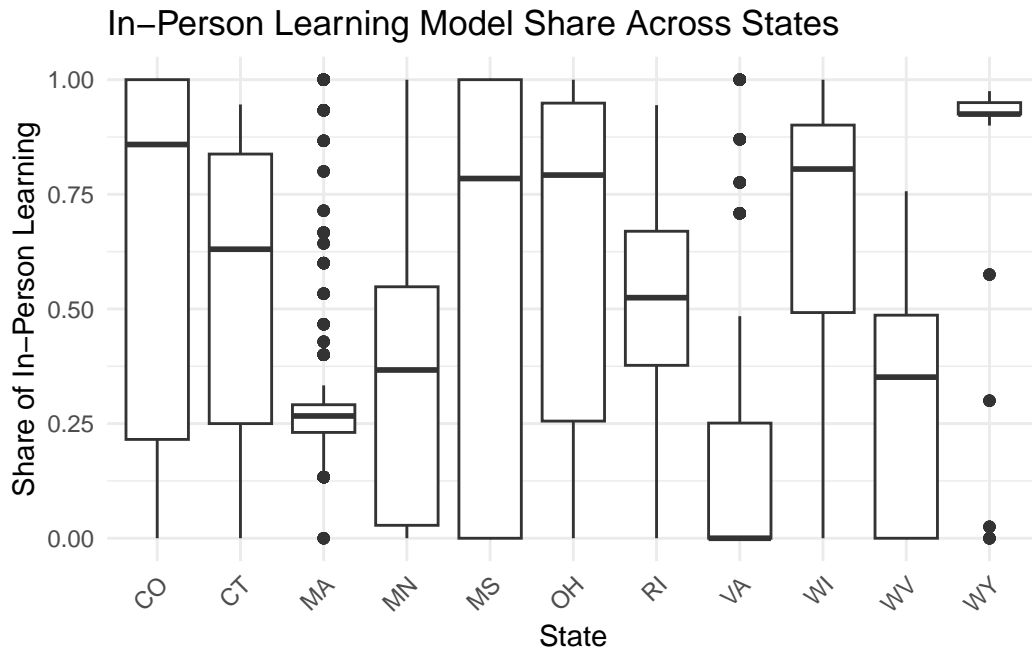


Figure 4: In-Person Learning Model

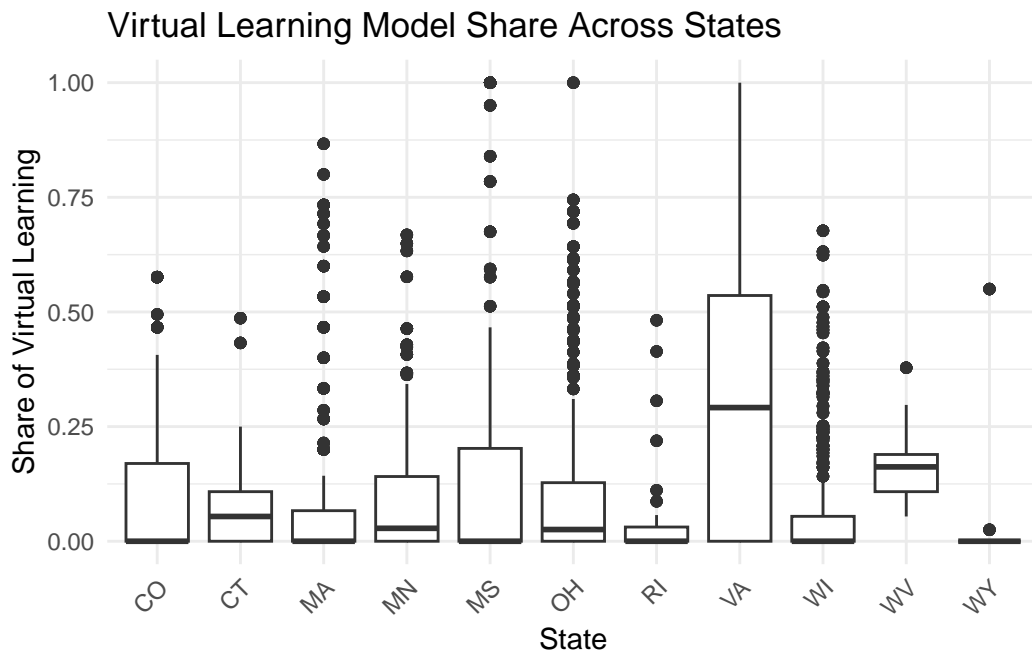


Figure 5: Virtual Learning Model



Figure 6: Hybrid Learning Model

Wyoming and Rhode Island had a median of 0 percent virtual learning whereas Virginia had the highest median of virtual learning. Lastly, the hybrid learning model Figure 6 box plot displayed variability, with some states having a wide interquartile range, indicating a diverse adoption of this learning model. However, Colorado had a median of 0 percent adoption of the hybrid learning model relative to other states. These visual insights set the foundation for further statistical examination in subsequent analyses.

After assessing the distribution of in-person, virtual, and hybrid learning models across states, we narrowed our focus to further investigate the impact of virtual learning, as this model showed a significant variation and prevalence during the pandemic period. We plotted a scatter plot Figure 7 and examined the relationship between the share of virtual learning and student passing rates. The scatter plot indicates a negative trend, which implies that an increase in the share of virtual learning correlates with a decrease in passing rates. The data points suggest that higher virtual learning is associated with lower academic performance of students.

### 3.2 Impact of COVID-19 Case Rates

When we look closer at how the pandemic affected student learning, we observe that the number of COVID-19 cases in an area seemed to have a correlation with student performance. We created box plots Figure 8 to observe the relationship between COVID-19 case rates per

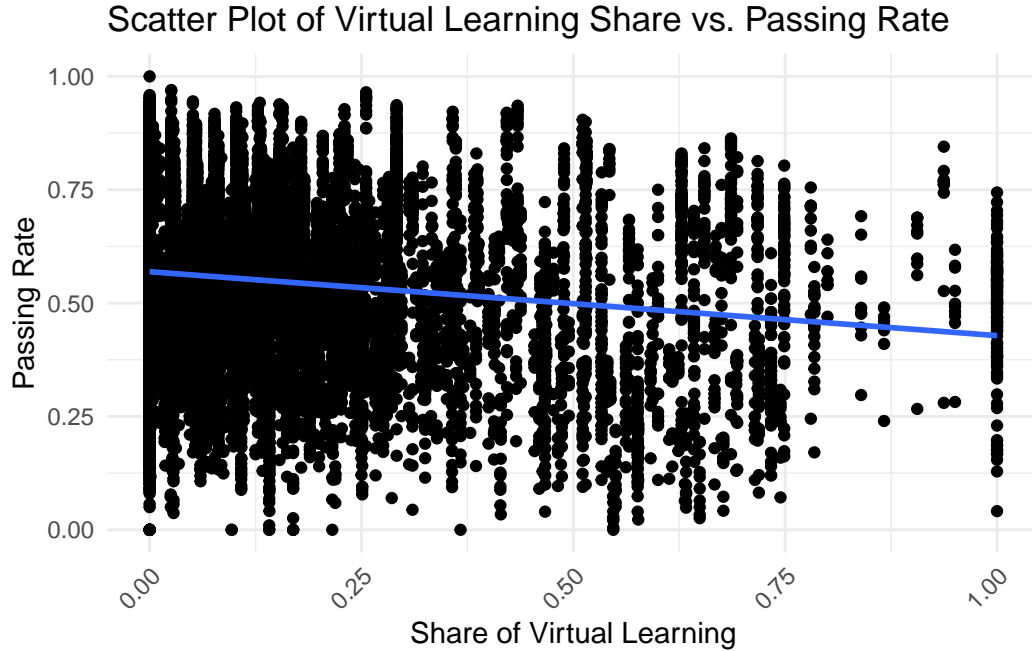


Figure 7: Scatter Plot of Virtual Learning Share vs. Passing Rate

100,000 and passing rates. We categorized case rates into ‘Low,’ ‘Medium,’ and ‘High’ based on the data’s quartile distribution. The plot reveals that districts with ‘Low’ COVID-19 case rates showed a median passing rate near the 50th percentile, while those categorized as ‘High’ showed a decline in median passing rates. Essentially, the trend suggests that places with fewer COVID-19 cases had average passing rates, but as the number of cases went up, the passing rates went down.

Referring to Figure 9, we further analyze the impact of the COVID case on the pass rate. We specifically use COVID cases measured by 100-thousand and the effect of case uptick on student pass rate. This allows us to have a closer look at the relationship between the variables, and the trend is negative as expected. Note that there is clustering on each end of the graph, this is likely explained by the time of data collection, i.e. when pass rates and student performance were measured. Given the rampant speed of COVID-19 cases during the pandemic, cases increased exponentially over different rates. However, the trend is clear and as expected despite the measurement gap. As the case rates go up, only points in lower passing rates are found, highlighting a negative trend. With this closer inspection of case rates, we can conclude that COVID-19 cases are indeed a significant factor contributing to the pass rates of students in the states we have assessed.

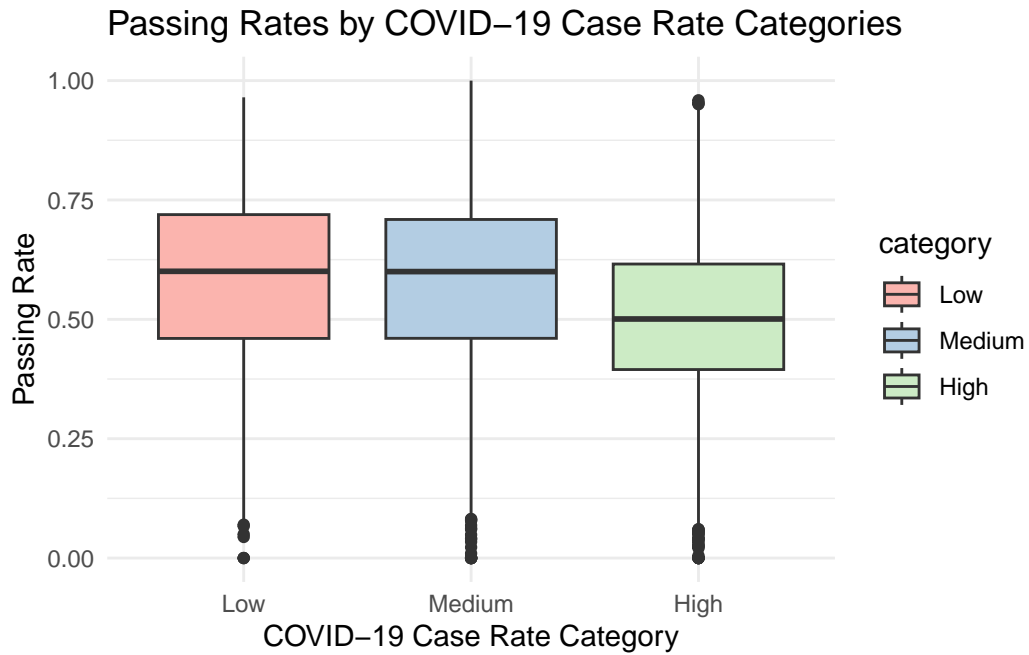


Figure 8: Passing Rates by COVID-19 Case Rate Categories

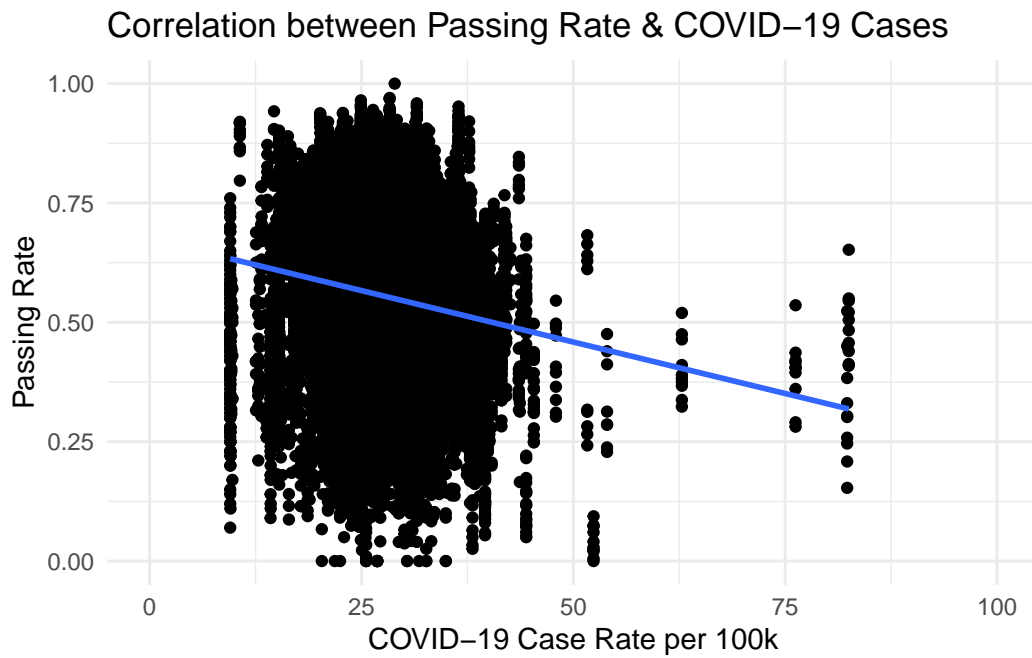


Figure 9: Correlation between Passing Rate and COVID-19 Case Rate per 100k

### 3.3 Impact of Staff Availability

We further expanded our analysis to examine the relationship between the average number of students per staff and the average pass rate for virtual learners across the 11 states in Figure 10.

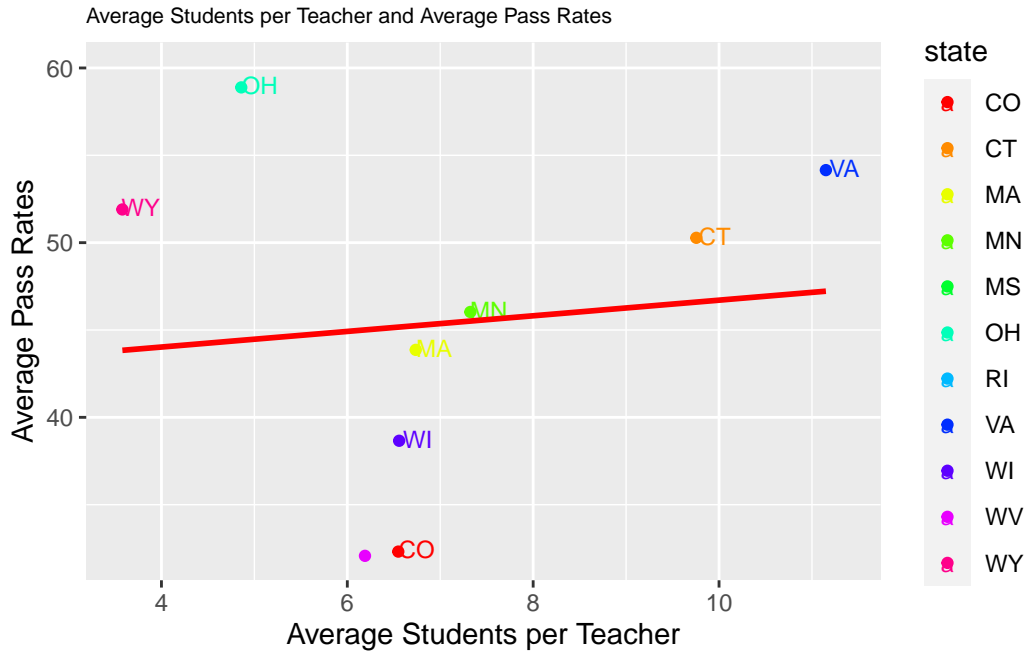


Figure 10: Relationship between teacher and pass rates

Due to insufficient data, Rhode Island (RI) and Mississippi (MS) were not graphed. The graph plots both these variables and shows an overall positive correlation between the number of students per staff on average vs the average pass rate. The trend indicates that the states with higher pass rates had more students assigned to a teacher. We note that Virginia (VA) had the highest student-to-staff ratio, while Wyoming (WY) had the lowest. While this may be not expected of a general learning environment, the virtual learning model may be a factor contributing to this positive relationship between the student-to-staff ratio and the average pass rates across the states.

Referring to Figure 10, we observe a positive correlation between the average student-staff ratio and the average pass rates across the 9 states with valid data. The correlation seems inconclusive towards pass rates in COVID-19, hence it is not a significant factor in our assessment.

### 3.4 Impact of Loss of Employment Income on Passing Rate

Every point on the Figure 11 represents the average passing rate of a particular state and their average loss of employment income. As seen in Figure 11, unemployment has a negative correlation with the passing rate of students across all the states in the U.S. As the loss of income increases, the average passing rate of the students across all the US states decreases and this establishes a negative correlation between these quantities.

The dataset by Jack and Oster (2023) also shows the negative correlation between unemployment and passing rates of students. This can be seen in Figure 12 where the graph is in the negative and which negative effects on the passing rate following unemployment in population of all the states.

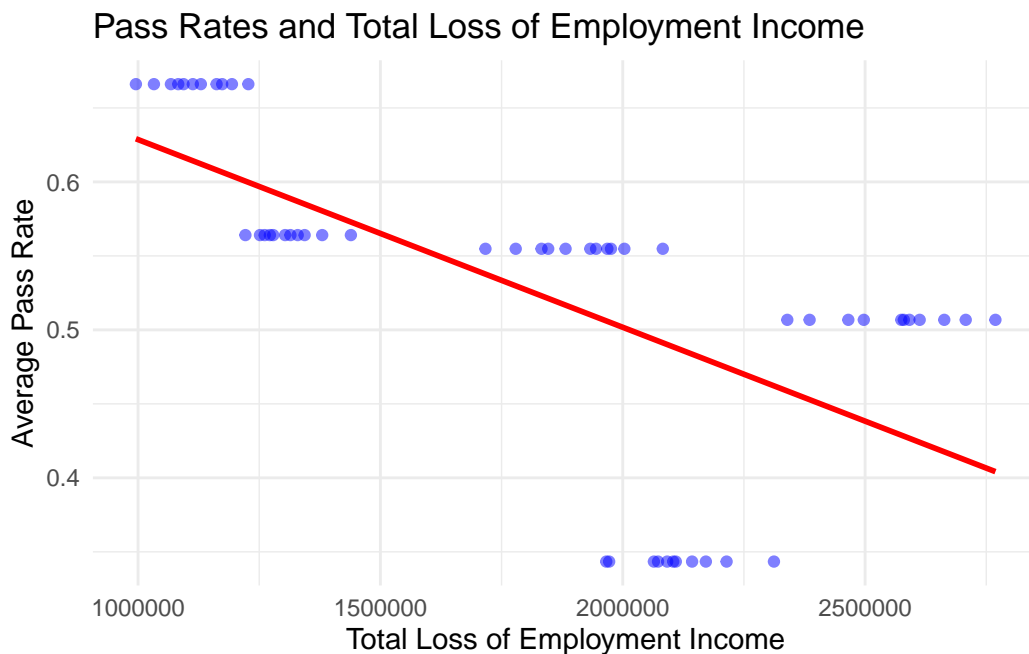


Figure 11: Correlation between Pass Rates and Total Loss of Employment Income

## 4 Discussion

### 4.1 Findings

In this paper, we have replicated the results found by Jack and Oster (2023). Their analysis sought to explore the impact of different learning models on COVID-19 for different states,

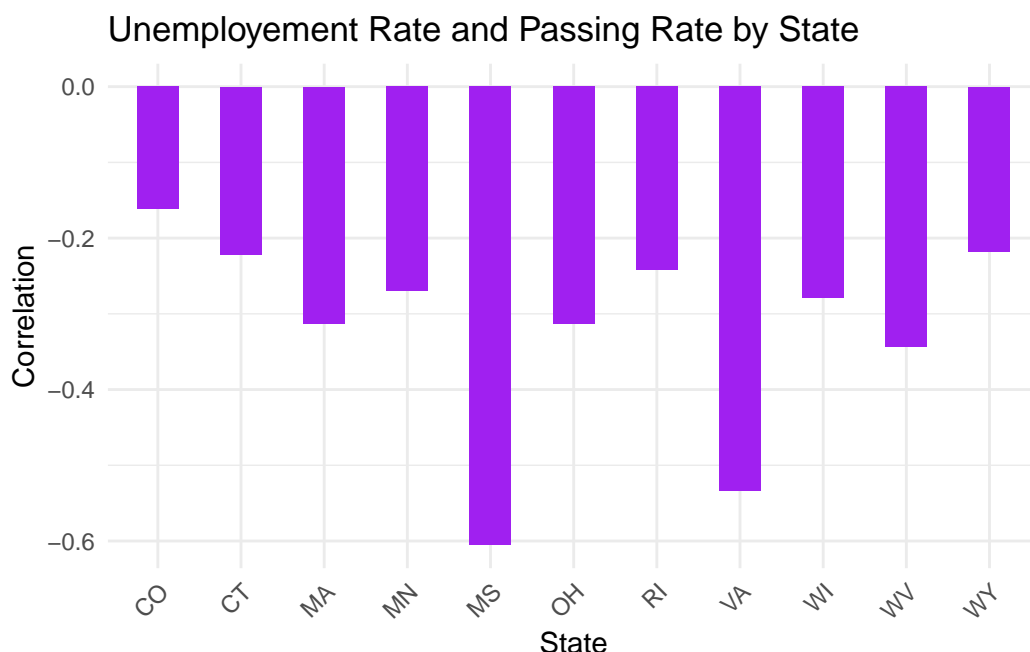


Figure 12: Correlation between Unemployment Rate and passing rate by State

and explored how certain modes of schooling were more or less effective with some focus on different racial groups. Our paper extends upon and replicates two of their major findings.

1. The virtual learning model had more variability in their results against pass rates, virtual learners had a decline in performance, measured by pass rates in our analysis. Specifically, we were able to quantify the amount of variability in the different learning models.
2. In exploring the learning loss subject wise, we consistently see across all groups that changes in pass rates for math are much larger than the changes in pass rates for ELA. Pass rate losses are approximately 13–14 percentage points in math and approximately 8 percentage points in ELA between 2019 and 2021.

While our analysis uses the general themes of the original claims regarding schooling mode, we have explored the decline of student test scores from the perspective of three other factors. These factors were not explored as a focused topic in the original work. As the paper has previously stated unemployment rates, staff count, covid case rates, learning mode and staff count are the focus of our analysis. Here are some of the key findings from our results.

In our study examining different teaching methods during the pandemic, we found that there isn't a one-size-fits-all approach to educating students when schools face disruptions. Some schools that continued in-person classes managed reasonably well, while others that transitioned to fully online instruction saw a decline in student grades. The effectiveness of the

hybrid approach, which combines online and in-person teaching, varied depending on how it was implemented in each location. We observed a trend suggesting that areas with lower COVID-19 case counts tended to have higher average passing rates, while areas with higher case counts experienced lower passing rates. This could indicate that increased illness rates may have impacted students' ability to perform academically, possibly due to concerns about health, disruptions in schooling, or other factors.

Additionally, our analysis considers the phenomenon of grade inflation during the pandemic years, which may have artificially boosted passing rates, particularly among lower-performing students (Sparks 2022). The presence of missing data from two states underscores the potential for bias in our assessment of this relationship, suggesting that other factors such as virtual learning modalities, grading policies during COVID-19, and regional differences may influence student pass rates.

The economic repercussions of job loss during the pandemic also play a significant role in students' academic outcomes. Families facing financial strain may struggle to afford essential educational resources, hindering students' ability to access materials, tutoring, or technology necessary for remote learning. This economic hardship exacerbates existing disparities in learning outcomes tied to socioeconomic status, particularly during the pandemic.

Furthermore, decreased access to educational resources due to financial constraints can impact students' preparation for standardized tests. As Vogels et al. (EMILY A. VOGELS 2020) highlight in their examination of U.S. adults surveyed between April 7th and 12th, 2020, approximately 36% expressed concerns about the possibility that their children might face challenges in completing school assignments due to the lack of access to a home computer for educational support, indicating that a significant portion of households faced challenges in completing school assignments due to the lack of a home computer.

Moreover, students from families experiencing job loss may encounter disruptions in their schooling, such as changes in attendance, increased mobility due to housing instability, or shifts between in-person and remote learning environments. These disruptions can impede academic progress and performance on standardized tests.

All these reasons are crucial to understanding the low passing marks of students during COVID-19 and present some context to the trend observed in our results above.

## 4.2 Accounting for Bias

The process of collecting quantitative and qualitative data, especially during a pandemic presents its unique challenges. Exploring the data provided in the paper we are replicating, we note that state data has been provided from individual state agencies, rather than from a centralized data source. As a result, we likely experience some degree of measurement bias as the measuring has not been stated to be standardized across the 11 states. Similarly, the pass rates data found in 'cleaned\_state\_scoredata' reports numbers district wise per state. The



data source did not provide insight into whether all districts in the states have been accounted for in our dataset, as such this introduces a selection bias it is likely more high performing or lower performing schools were allotted for our analysis. Upon this closer inspection of the replication data and resulting outcomes of our analysis, we note that the nature of data and their compatibility are the primary concern for the validity of our results.

The unemployment data obtained from Household Pulse Survey (Bureau 2020) reports data through survey sampling. The survey references terms like ‘loss of employment income’ and its definition may vary across respondents or regions to some degree. Similarly, the economic conditions and opportunities vary across regions, what is considered well may not be the case in other regions. Hence these may add to disparities in reported loss of employment income. As is with surveys there is likely a factor of non-response bias at play, certain groups of people may be less likely to respond to the survey, leading to under representation of their experiences.

In the staff-to-student ratio factor analysis, the student-staff ratio variable was newly calculated given the existing variables in our datasets. In this particular process, we noted that 2 states did not report staff counts throughout the COVID-19 schools years 2020-2021. Given we only have access to 11 states in total, out of those, two states had missing data. That proportion of missed data likely has neutralized what could’ve been a more definitive trend.

In general we must note that while we aim to assess how our chosen factors explored in the paper are relevant to student performance, we do not assess all such possibilities. The pandemic had

### **4.3 Limitations**

Many of limitations of our analysis stem from those in the data of the original work of Jack and Oster (2023). In the context of studying COVID-19 learning outcomes, we require data that is available across pre-covid and covid times to draw comparisons upon significance of certain factors. However, this lacks across datasets provided to us in the replication.

Specifically, our analysis upon student-to-staff ratio as a factor was affected by the discrepancy in the data. While we had availability of pass rate through the years 2016-2021, we also noted that there was no data available for staff counts prior to 2020. In order to correctly claim student-to-staff ratio as a factor contributing to student learning loss during COVID-19, a comparison to the years prior would be ideal in concluding it as a sure factor. By this restriction in our data, concluding staff count as a factor was inconclusive. These limitations likely explain why our student-to-teacher ratio did not develop as a strong indicator of student performance.

In another instance, our ‘learning model’ folder has data for each of the 11 U.S. states in their respective .csv files. Each file however reports data in different time intervals, some biweekly and some monthly, possibly introducing measurement bias to a certain degree to our overall state specific results. Our main measure of student performance, pass rates is found

in a different file, a quantity reported by various districts per state. Upon closer inspection, the pass rates are only reported at certain intervals in the pandemic. When we combined these variables to assess their relationship, we generally aggregated our data by obtaining a yearly average that encompasses all the intervals reported. As a direct consequence, we were restricted to exploring only the high-level picture of the factors, a yearly correlation with pass rates.

Lastly, the dataset employs records on 11 U.S states to generalize the COVID-19 learning outcomes across United States. Only 11 out of 48 states have been analysed, that is roughly one quarter of the country being represented, and no specific reasoning was provided for this choice. This discrepancy raises the question whether the states assessed were of significance or convenience. Neither California nor New York were accounted for, naturally these two areas likely had much different experiences in COVID-19 being key metropolitan hubs in America. These limitations hinder a broader image of COVID-19 learning trends, and our ability to synthesize accurate trends related to COVID-19 impact on learning that is representative of the entire country.

#### **4.4 Future Research**

The current works of Jack and Oster (2023) only touch upon a handful topics related to COVID-19 learning impacts. Often these results presented stem from incomplete, incompatible datasets. To gain a clear picture of the factors that affected student performance, the study needs to expand other aspects of the pandemic.

COVID-19 may be in the past, but its profound impact on society and education remains relevant today. Studies indicate that students faced significant challenges during the pandemic, leading to potential long-term learning gaps. Analyzing the performance of students in grades 3-8 during COVID-19 compared to their current performance could provide valuable insights into the lasting effects of pandemic-related disruptions. These grades mark a critical period of development and impressionability for students, making it crucial to understand how their experiences during the pandemic have influenced their learning.

Additionally, it is crucial to investigate whether students who experienced additional adversities, such as declines in mental health, continue to face challenges that affect their education today. Research suggests approximately 75% of young people in primary mental health care services experienced mental health issues post pandemic, with high levels of clinical depression, anxiety, and loneliness reported (Imogen H. Bell and Thompsona 2023). These findings highlight the need to explore the ongoing impact of mental health on learning and well-being.

By exploring these areas in greater detail, we can better prepare for future global health emergencies and develop guidelines to mitigate their widespread impact on younger populations. Addressing these issues is essential for preventing long-lasting epidemics related to mental health, promoting overall well-being among young people and minimizing overall learning loss.

## References

- AEA. 2022. *The American Economic Association Portal*. The American Economic Association. <https://www.aeaweb.org/>.
- Bureau, United States Census. 2020. *Household Pulse Survey*. The American Economic Association. <https://www.census.gov/data-tools/demo/hhp/#/?periodSelector=12&measures=JOBLOSS>.
- Card, Nations Report. 2023. “Policy Brief: education During COVID-19 and Beyond.” <https://docs.google.com/document/d/1Z8eZgmlh49z6ZXjF1-MF4SEJBWdliHV4mO-0Dc-76io/edit>.
- Clarke, Erik, Scott Sherrill-Mix, and Charlotte Dawson. 2023. *Ggbeeswarm: Categorical Scatter (Violin Point) Plots*. <https://github.com/eclarke/ggbeeswarm>.
- Education Statistics, Nations Center for. 2023. “Technology Support.” <https://nces.ed.gov/surveys/annualreports/topical-studies/covid/theme/elementary-and-secondary-education-technology-support/>.
- EMILY A. VOGELS, LEE RAINIE AND MONICA ANDERSON, ANDREW PERRIN. 2020. “53.” <https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-covid-19-outbreak/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Imogen H. Bell, Amy Broomhall, Jennifer Nicholas, and Andrew Thompsona. 2023. “The Impact of COVID-19 on Youth Mental Health: A Mixed Methods Survey.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9883078/>.
- Jack, James Okun, Clare Halloran, and Emily Oster. 2023. “Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts.” *American Economic Review* 5 (2): 173–90. <https://doi.org/10.1257/aeri.20210748>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://here.r-lib.org/>.
- Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. <https://tibble.tidyverse.org/>.
- Nations, United. 2020. “2022–2023 Long-Term Trend (LTT) Mathematics and Reading Assessments at Ages 9 and 13.” [https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg\\_policy\\_brief\\_covid-19\\_and\\_education\\_august\\_2020.pdf](https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf).
- Ozcan, Mehmet. 2021. “Factors Affecting Students’ Academic Achievement According to the Teachers’ Opinion.” <https://files.eric.ed.gov/fulltext/EJ1318651.pdf>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Slowikowski, Kamil. 2024. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. <https://ggrepel.slowkow.com/>.
- Sparks, Sarah D. 2022. “Pace of Grade Inflation Picked up During the Pandemic, Study Says.” <https://www.edweek.org/leadership/pace-of-grade-inflation-picked-up-during-the-pandemic-study-says/2022/05>.
- USA. 2022. *COVID-19 School Data Hub*. <https://www.covidschooldatahub.com/>.

- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 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 Golemund, 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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://readr.tidyverse.org>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.