Significant declines in standardised test scores due to COVID-19 school closures disproportionately affect vulnerable students: A replication analysis using data from the United States and a case study on the Netherlands*

Krishiv Jain, Julia Kim, Abbass Sleiman

February 15, 2024

Abstract

School closures due to the COVID-19 pandemic have raised significant concerns about the consequences on student learning and achievement gaps. This paper replicates a data analysis of the scope of school closures in the United States in the 2020-2021 school year, the inequitable distribution of such closures by demographic characteristics and the resultant declines in pass rates on standardised tests in mathematics and English Language Arts (ELA) of students in grades 3-8 across 11 states. We apply secondary research regarding the scope of school closures in the Netherlands in same year and the declines in national examination scores in reading, spelling and mathematics of Dutch students in grades 4-7. Although the Netherlands is regarded as 'best-case' scenario, with a short lockdown, equitable school funding, and high degree of technological preparedness, we find that Dutch students still experienced a significant learning loss, equivalent to one-fifth of that of the United States, with a disproportionate impact on vulnerable students. These findings suggest that COVID-19 school closures imposed significant costs of learning loss and widened inequality gaps in the United States and the Netherlands, with impacts likely larger in countries disenfranchised by weaker infrastructure or longer closures.

Table of contents

*Code and data are available at: https://github.com/AbbassSleiman/COVID19-US-School-Closures; Replication on Social Science Reproduction platform available at: https://www.socialsciencereproduction.org/reproductions/8aab4425-63ad-47bc-93ef-82076d6e49cc/index

1 Introduction

On March 11, 2020, the World Health Organisation (WHO) declared the COVID-19 outbreak a global pandemic, urging all countries take action in detecting infection and limiting spread (WHO 2020). Shortly after the announcement, many countries began to implement measures to reduce COVID-19 contagion and transmission. In spite of the general effectiveness of such measures in saving lives (Wu et al. 2020), they imposed severe disruptions to the lives of children and adolescents. These included delayed healthcare provision, decreased contact with peers and relatives, routine modification, and diminished sense of well-being (Richard et al. 2023). Among the most significant disruptions, and the topic of this paper, was the closure of schools in favour of virtual learning.

Before the pandemic, there was a modestly-sized economic literature on the impact of school closures on learning, with data generally drawn from short-term episodes such as teacher strikes (R. Jack and Oster 2023). It was widely observed that time outside of school caused loss of literacy and numeracy skills, more so among the marginalised, and widened student achievement gaps (Aurini and Davies 2021). These findings are certainly suggestive of the possible effects of COVID-19 school closures on student learning. However, as the pandemic's disruptions typically surpassed previous episodes of school closures, the scope for inference from earlier works is limited (R. Jack and Oster 2023). A paper by Jack and Oster, "COVID-19, School Closures, and Outcomes" (2023) published in American Economic Association's *Journal of Economic Perspectives*, sought to address this gap, by focusing on emerging short-term evidence of the impact of COVID-19 school closures on academic performance.

In studying the scope of school closures across the United States (US), they reported that the country experienced, on average, one of the highest number of days of school closed between January 2020 to December 2021 worldwide (R. Jack and Oster 2023). Jack and Oster (2023) also found that the duration of school closures varies with local demographic characteristics, with more disadvantaged districts facing more prolonged closures. The study population was grade 3 to 8 students across 11 major American states, being Colorado (CO), Connecticut (CT), Massachusetts (MA), Minnesota (MN), Mississippi (MI), Ohio (OH), Rhode Island (RI), Virginia (VA), Wisconsin (WI), West Virginia (WV) and Wyoming (WY). The estimand was the impact of COVID-19 closures on academic performance. They found considerable declines in test scores during the 2020-2021 academic year, with declines larger in school districts with less in-person instruction (R. Jack and Oster 2023). On average, pass rates fell from 2019 to 2021 by an average of 12.8 percentage points in mathematics and 6.8 in English Language Arts (ELA) (R. Jack and Oster 2023).

Our paper seeks to assess the robustness and generalisability of Jack and Oster's work. We successfully replicate and extend three of its major research claims:

(1) With respect to the world, the United States faced the second highest number of days of school-closed between January 2020 to December 2021, with a mean number of days of 667.

(2) Access to in-person education was unequal across various demographic characteristics, being significantly correlated with an area's ethnic composition and broadband usage. In general, more disadvantaged districts spent more of the school year in virtual schooling.

(3) School closures contributed significantly to the decline in students' pass rates in mathematics and ELA in the 2020-2021 school year. Changes in ELA scores were smaller than mathematics scores, but significantly larger in districts with larger populations of students who are Black, Hispanic, English Language Learners (ELL) or eligible for free and reduced price lunch (FRPL).

We also apply a Dutch-facing lens, examining the extent of school closures and the impact on student learning, in a country whose pandemic circumstances were regarded as most favourable for students (Engzell, Frey, and Verhagen 2021). In spite of a short lockdown, equitable school funding and high degree of technological preparedness, we discuss how Dutch students still suffered a loss of the equivalent of one-fifth of a year of education and how the effects were distributed unequally according to student demographics. Throughout, our investigation is conducted using the open-source statistical programming language `R` (R Core Team 2022), with functionalities from `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2021), `knitr` (Xie 2023), `here` (Müller 2020), `kableExtra` (Zhu 2021), `DiagrammeR` (Iannone and Roy 2024), `readxl` (Wickham and Bryan 2023), `patchwork` (Pedersen 2024) and `gridExtra` (Auguie 2017).

This paper is structured, as follows. In Section 2, we address the sources of the data sets in the original paper, the methodologies used to collect them, and highlight key features. In Section 3, we conduct a replication of the original study to confirm the robustness of its findings and to design scientific figures that permit a more effective data visualisation. In Section 4, we assess the relevance of these findings to the Netherlands. We compliment this discussion by addressing ethical implications and limitations of the original research, and by giving directions for future analysis.

2 Data

2.1 Source

The paper and raw data used for replication is obtained from "COVID-19, School Closures, and Outcomes" (R. Jack and Oster 2023), published in the American Economic Association's *Journal of Economic Perspectives* (AEA 2024). To investigate the degree of school closures in the United States, and their attendant effects on children, they pose the following three questions:

(1) Relative to the global picture, what was the extent of school closures in the United States in 2020-2021?

(2) What are the demographic patterns underlying these closures? Namely, how are the number of days of school closed correlated with ethnicity and broadband usage at the district- or county-level?

(3) What is the short-term impact of school closures on student academic performance and learning inequalities?

Whilst the work by Jack and Oster (R. Jack and Oster 2023) uses number of data files, we require only a subset of them. The specific files used to address each of the three major research questions, including their sources and method of data collection, is given below.

2.1.1 Extent of School Closures in the United States and the World

`region_data.csv`

This dataset is obtained from the Oxford COVID-19 Policy Tracker, a project that collected information on COVID-19 policy measures for the years 2020, 2021 and 2022 (Hale et al. 2021). For 185 countries, the dataset assigns a daily Tracker score for each day between January 1st, 2020, and December 31st, 2022. Note that a Tracker score is defined as an integer value between 0 and 3, with a value of 2 or 3 denoting a school closure (R. Jack and Oster 2023).

`country_region.csv`

This dataset was manually created by ourselves for the purpose of replicating "FIGURE 2" in `99-replications.R`. It includes the same countries found in `region_data.csv` as well as the region to which they correspond, as classified by the World Bank (World Bank Group 2023). These regions are sub-Saharan Africa, Europe and Central Asia, East Asia and Pacific, Middle East and North Africa, Latin America and Caribbean, South Asia, and North America.

2.1.2 Average Number of School Closures by Local Demographic Characteristics

`District_Overall_Shares.csv`

This dataset is obtained from the COVID-19 School Data Hub (CSDH), a database created by Jack and Oster to understand how the COVID-19 pandemic shaped American students' modes of learning in 2020-2021 (Hub 2021). For each district in the United States, it provides information on the percentage of the school year spent in each type of schooling mode – in-person, hybrid and virtual – and the state of which they are a part. To collect such data, the CDSH team submitted data requests to state education agencies, asking for their record of learning models used by schools and districts in 2020-2021 (Hub 2021). They requested the data to be given at the school or district level and at the most frequent reporting interval available.

`nces_district_directory_2018.dta`

This dataset is obtained from the National Center of Education Statistics (NCES 2021). For each district, it provides information as to the county and state to which they belong, and their NCES district ID.

`nces_district_enrollment_2018_2020.csv.zip`

This dataset is obtained from the National Center of Education Statistics (NCES 2021). For each district and for each of the years 2018, 2019 and 2020, it reports the total student enrollment, split by grade (ranging from Kindergarten to Grade 12), sex (Male and Female), and ethnic class (American Indian or Alaska Native, Asian, Black, Hispanic, Native Hawaiian or other Pacific Islander, or Two or more races). The data is collected via surveys and supplied by state agencies to NCES (U.S. Department of Commerce n.d.).

Note the original file was a `Stata` file of very large size. To reduce its size, 2018 and 2019 data were removed, being unnecessary for our paper. The file was then converted to a `.csv` file and compressed into a `.zip` file, before being processed in R for analysis.

`broadband_data_2020October.csv`

This dataset is available from Microsoft's "United States Broadband Usage Percentages Dataset" GitHub Repository (Kahan 2020). The data consists of broadband usage percentages at the US county-level in October 2020, obtained from anonymised data Microsoft collects. In less technical terms, this means that every time a device receives an update from or connects to a Microsoft service, Microsoft servers make an estimate of its throughput speed and determine its location at the zip-code level (Kahan 2020). After making such measurements, noise is added to data aggregations, preventing leakage about specific individuals in the dataset and ensuring differential privacy.

2.1.3 Standardised Test Assessment Scores

`scores_lm_demographics.csv`

This dataset is a csv version of `scores_lm_demographics.dta` obtained from the final dataset found in the data repository for "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts" (C. H. Jack Rebecca and Oster 2023). This dataset contains extensive information on various school districts across the aforementioned 11 states, including information on the shares of students by race, the share of schooling done in-person, and changes in average test scores in both Math and ELA from the previous year across the years 2018-2019 and 2021-2022 (no information is present regarding the year 2020 as all standardised assessments were cancelled due to the pandemic).

2.2 Methodology

2.2.1 Extent of School Closures in the United States and the World

Figure 1 depicts the average number of school closures by region, as defined by the World Bank, between the start of January 2020 and the end of December 2021. In particular, it employs data on 185 countries, each of which is assigned its particular region, and provides the average number of school closures for nations in each region between the aforementioned time frame.

In order to achieve this, we first created the dataset `country_region.csv` which assigns each of the 185 countries present in the primary dataset, `region_data.csv`, its respective region. This was done since `region_data.csv` did not feature each nation's region, and as such, by creating the `country_region.csv` dataset we were able to 'merge' the two sets of information so that the final product featured all information regarding the Covid Tracker score of each country one each day, as well as the region the country is in.

We initially cleaned the data in `region_data.csv` by omitting information from the year 2022 in a bid to replicate and examine the internal validity of the work by Jack and Oster who also omitted the year 2022 from their figure (R. Jack and Oster 2023). From there, only the columns of interest were kept which detailed the country name, (empty) region name, date, and Covid Tracker score. We also found a minor spelling error of the Faroe Islands which were incorrectly spelled as Faeroe Islands which we also amended in the cleaning process. From here, we simply merged the cleaned `region_data.csv` with the regions listed in `country_region.csv` in order fill in the empty region names present in the original dataset.

With the cleaned dataset, we were able to create the figure detailing the average days closed by region by first grouping the cleaned data by region name and country, then summing the number of instances where a Tracker score of 2 or 3 was present (as this defines a school closure as per Jack and Oster's work (R. Jack and Oster 2023)). From there, we calculated the mean number of days closed by region by computing the average number of days closed for all nations in a given region.

Then, in an attempt to garner a deeper understanding of how the US compares to other nations with regards to their COVID-19 schooling policies and measures, we employ Table 1 which provides a snippet of the top 5 countries by number of school closures between the start of January 2020 and the end of December 2021. Due to the US featuring the second-highest number of school closures, Table 1 specifically depicts the top 5 nations in an attempt to evidence the US' large number of closures relative to the rest of the world. In order to generate this figure, we employed the same cleaned dataset used to create Figure 1. In this particular case, we simply grouped the data by country and summed the number of instances where a country registered a Covid Tracker score of 2 or 3 and reported a snippet of the generated table detailing each country and the number of school closures it faced in the aforementioned time frame.

2.2.2 Average Number of School Closures by Local Demographic Characteristics

The purpose of Figures 2 and 3 is to explore the correlations between certain district and county characteristics, and the average number of virtual school days for the respective district or county. Specifically, we look at broandband usage by county, and the shares of four races (White, Black, Asian, Hispanic) by district. For each county or district, we assign "high" or "low" depending on if their level for the characteristics being explored is above or below their state's median, respectively. By doing so, we allow for the results to be driven by a variation within states, as opposed to variation across states.

We start by using `nces_district_enrollment_2018_2020.csv.zip`. As previously mentioned, this data set splits the total enrollment into various categories. We start by converting the columns into the numeric class as appropriate. Some enrollment cells did not have a value, indicating that data was not collected for that specific metric. These occurred a handful of times for certain races within certain districts. These NAs were replaced with 0s. For our paper, we are concerned specifically with race. Thus, for each district, we calculate the total number of White, Black, Asian, and Hispanic students enrolled. Given the structure of the data, this is done by summing the total enrollment per race for Male and Female for each district. Next the total enrollment for the whole district was calculated, before then calculating the share of total enrollment that each race comprises.

The data above was then separated into different tables for each race. For each race, we then classified the district as either "High" or "Low" depending on if their share was above or below their respective state's median. Using this data, we were now interested in the number of virtual days that each of these districts had. This information was available in the `District_Overall_Shares.csv` data set. Given that the US average number of school days per year is 180, the percentage for share_virtual was multiplied by 180 for each district.

Next, the aforementioned data for each race that separated the districts into quantiles, and the data for the number of virtual days per district were merged. This results in some NAs being produced since the data set on race had more districts than the data on the number of virtual days. Given the lack of data, all NAs were droppped. Finally, the number of virtual days for the "High" quantile and the "Low" quantile were averaged. This results in four data sets, one for each race considered.

To calculate the information regarding broadband, we started with the `broadband_data_2020October.csv` data set. Only columns that were necessary were kept, and the columns were converted to the numeric class as appropriate. Each county was then classified as either "High" or "Low" depending on if their share was above or below their respective state's median. A single county contains multiple districts, and since our data regarding the number of virtual days was based on district, the data was averaged across districts for each county. `nces_district_directory_2018.dta` provided information on which county each district belonged to, and was merged with the previous data that calculated the average number of virtual days for each district, before then calculating the average number of virtual days for each county. Finally, we average the number of virtual days for the counties labelled "High" and "Low."

2.2.3 Standardised Test Assessment Scores

3 Results

Data from the Oxford COVID-19 Policy Tracker (Hale et al. 2021) and World Bank (World Bank Group 2023) were used to determine the average number of school closures by regions worldwide between January 2020-December 2021. We reconstruct this result in Figure 1 and extract the top 5 countries with the highest average number of closures in Table 1. Figure 1 shows that North America had the highest at 535 days, which is around 73% of the two-year period, whilst sub-Saharan Africa had the least at 300 days, or around 38% of the two-year period. Thus, countries with the longest closures spent more than twice the amount of time in virtual schooling compared to those with the shortest closures.

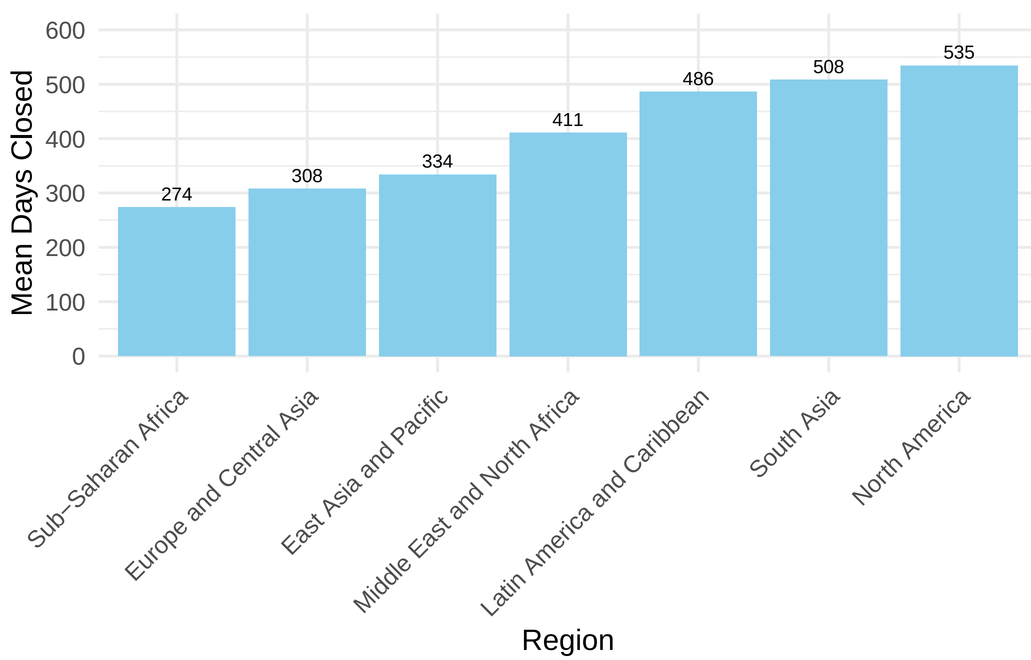

Figure 1: Average number of days of school closed, by region, between January 2020–December 2021.

In studying individual countries, Table 1 shows that the United States faced the second highest number of closures worldwide during this two-year period, with a mean number of 667 days. This is, in fact, significantly higher than the mean number of closures worldwide, determined to be 356 ± 175 days. Thus, relative to the global picture, the extent of school closures in North America and, in particular, the United States appears to be significantly more prolonged.

Table 1: Countries with Highest School Closures from Jan 2020-Dec 2021.

Country	Days Closed
Azerbaijan	669
United States	667
Romania	661
Brazil	660
Panama	660

Moreover, data from CSDH (Hub 2021), NCES (NCES 2021) and Microsoft (Kahan 2020) were used to delve deeper and analyse correlations between the average number of virtual schools days, and certain demographic characteristics, and Figures 2 and 3 were created. As discussed previously, Figure 2 shows the average number of virtual days for counties that were classified as “Low” and those classified as “High,” whilst Figure 3 shows the average for districts. The classification of districts and counties was explained in Section 2.2.2.

We built upon the research done by Jack and Oster (2023) in their paper by firstly utilising a different methodology. Jack and Oster had weighted their results by total district enrollment, thus, districts with larger enrollments will have a greater impact on the overall average than districts with smaller enrollments. In our paper, we do not weight it by enrollment, which allowed us to treat each district equally, regardless of its size. This offers a more balanced view of the educational landscape without overemphasising larger districts. We also enriched the analysis on the average virtual days by race by also considering the shares of White and Asian students, with the goal of better analysing the nuanced impact of racial demographics. Moreover, it is important to note that Jack and Oster did not mention their methodololgy for finding the virtual days for each county, given that data was only available for district. As previously mentioned, we found this metric by averaging across the districts found in each county.

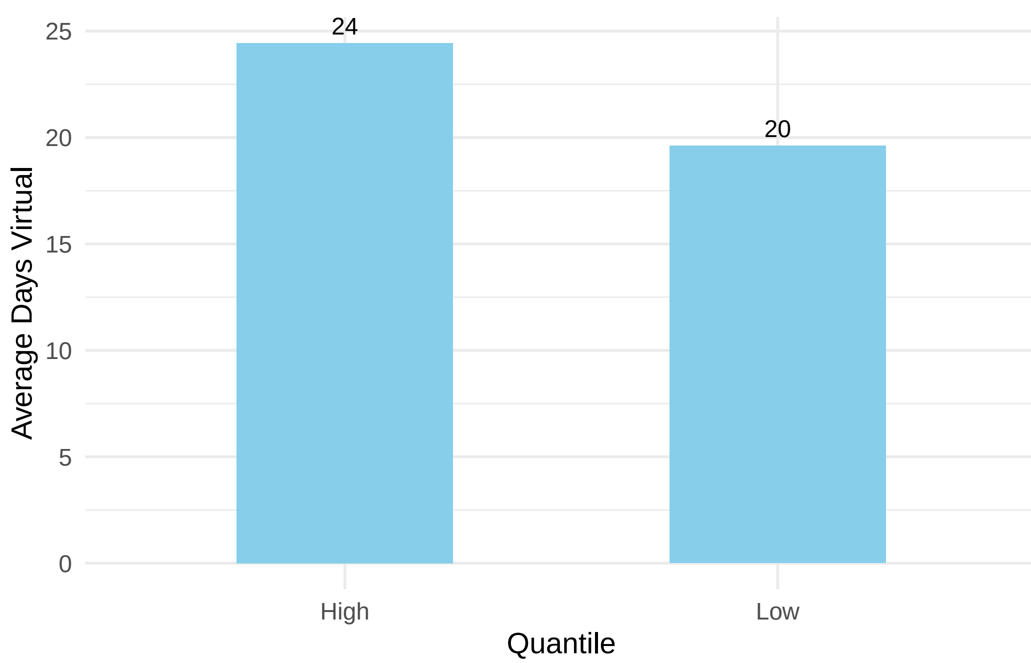

Figure 2: Average days in virtual school by county broadband usage

Figure 2 shows that the average number of virtual days for counties that were on the lower end of broadband usage for their respective states was 20 days, whilst the average number of virtual days for the "High" counties was 24 days. This result is quite different from the one obtained by Jack and Oster (2023), whereby they found that the average for "Low" counties was 55 days, and "High" counties was 48 days. Thus, not only are the number of days for the quantitles different, but our analysis found that the counties with lower broadband had less virtual days. The discrepancy in data can arise due to numerous reasons. Firstly, we did not weight the data by enrollment. Secondly, we neglected the cells without any data, whereas Jack and Oster may have dealt with these in a different manner. Lastly, their method of finding average virtual days for a county may be different. The observed differences in virtual learning days between counties with varying broadband usage levels highlight potential disparities in access to and the usage of online education resources. This suggests that regions with lower broadband usage may face additional challenges in facilitating effective virtual learning experiences for students.

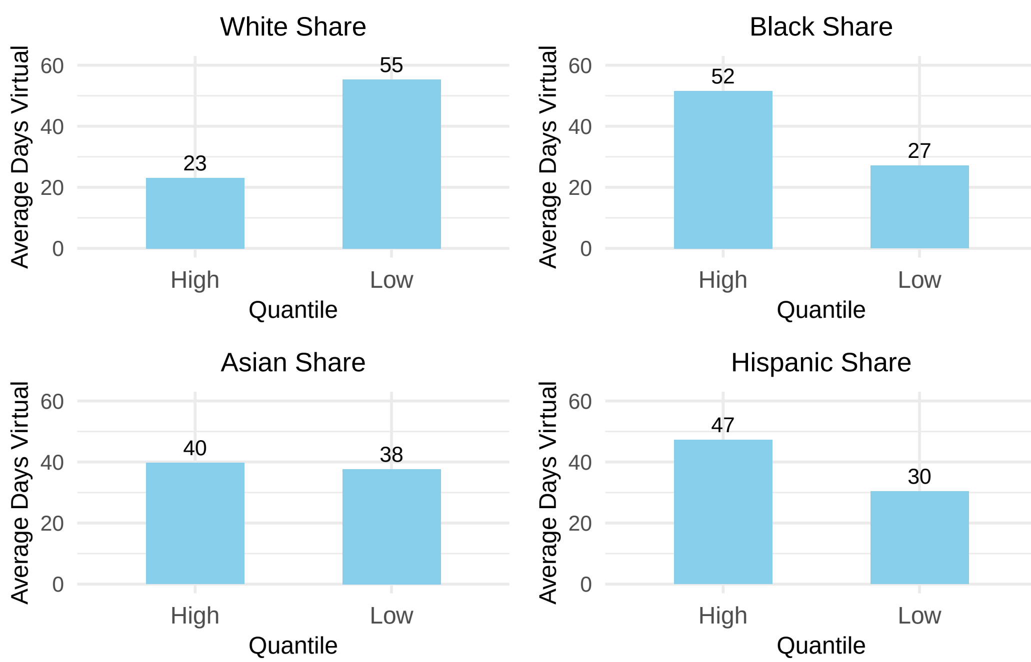

Figure 3: Average days in virtual school by district-level race shares

Figure 3 shows that the average number of virtual days for the districts classified as having a "High" share of black students spent 52 days in virtual schooling, and the "Low" districts spent 27 days. Jack and Oster (2023) found the number to be 56 days and 36 days respective. Our results for Hispanic students show that "High" districts had an average of 47 virtual days, and "Low" districts had an average of 30 virtual days. Jack and Oster found 53 days and 43 days respectively. Thus, our results are consistent with Jack and Oster in that we also observed that districts with a higher share of black students and a higher share of Hispanic students spent more days in virtual schooling. We find a larger discrepancy between "High" and "Low" district for both shares, primarily driven by the lower number of virtual days we calculated for the "Low" districts.

We extended Jack and Oster's research by also looking at the shares of White and Asian students. We found that districts with "High" shares of white students had an average of 55 days virtual, whilst the "Low" districts had 23 virtual days. For Asian students, we found "High" districts had an average of 40 days virtual, whilst "Low" districts had an average of 38 days virtual. This variation can be considered insignificant.

Taken together, in-person education was inequitably distributed across demographic characteristics, with more disadvantaged districts spending more time in virtual schooling (R. Jack and Oster 2023). In particular, districts with less technological preparedness and higher shares of minorities were further disenfranchised by less access to in-person education. This pattern likely reinforced educational inequities between vulnerable and more privileged groups.

Finally, data from … was used to examine changes in district-level pass rate data from 11 state standardised assessments, taken in the spring of 2018, 2019 and 2021. We again reconstruct and extend the data…

[ADD DETAIL].

Panels A and B show the scores in mathematics and in ELA, respectively. [EXPLAIN FIGURE & SYMBOLS]

Consistent with Jack and Oster (2023), our analysis shows a significant decline in pass rates from the 2018-2019 to the 2020-2021 school year. In particular, pass rates in mathematics and ELA declined by 12.8 and 6.8 percentage points, respectively. The discrepancy between these two subjects is consistent with much of the current economic literature, which shows mathematics scores to be more sensitive to schooling differences (Betts and Tang 2011; Angrist, Pathak, and Walters 2013).

… Thus, we find that in-person learning has significantly higher benefits for districts with larger shares of Black and Hispanic students.

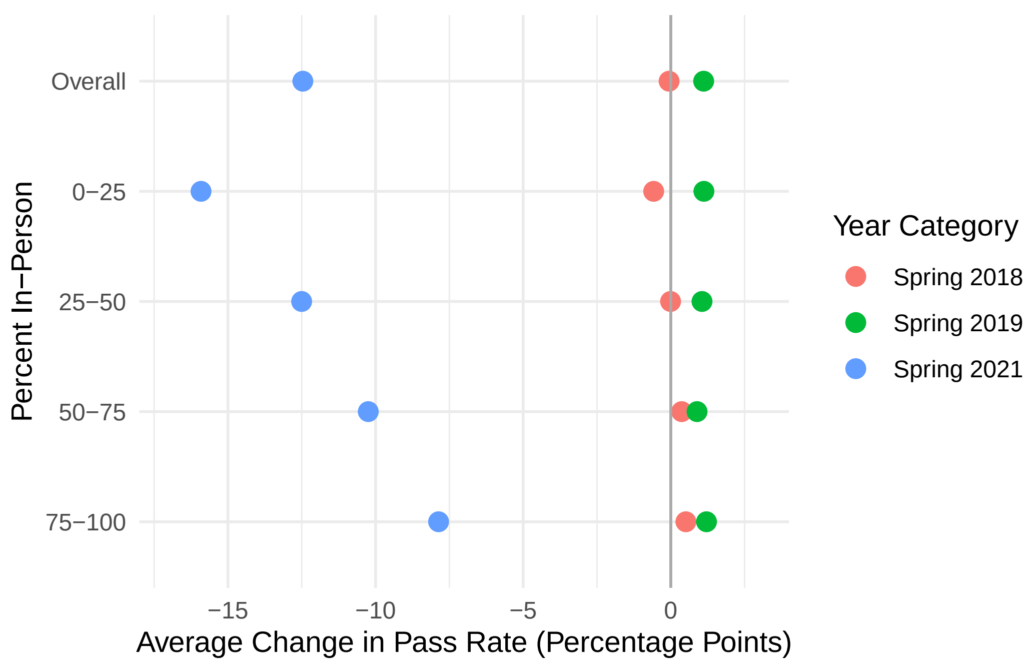

Figure 4: Test score by percent online (change)

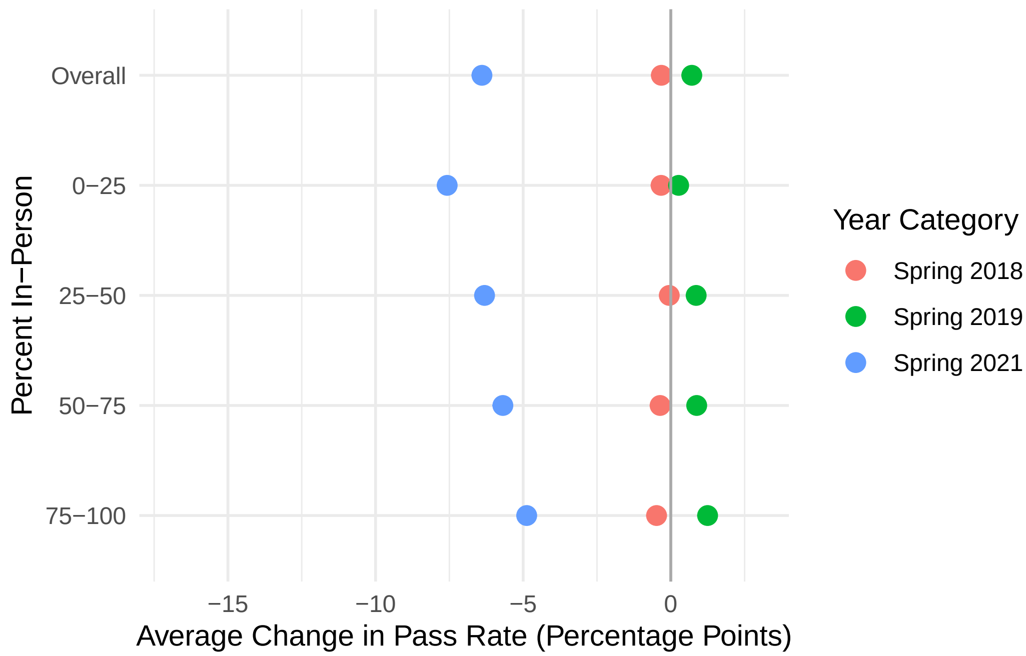

Figure 5: Test score by percent online (change)

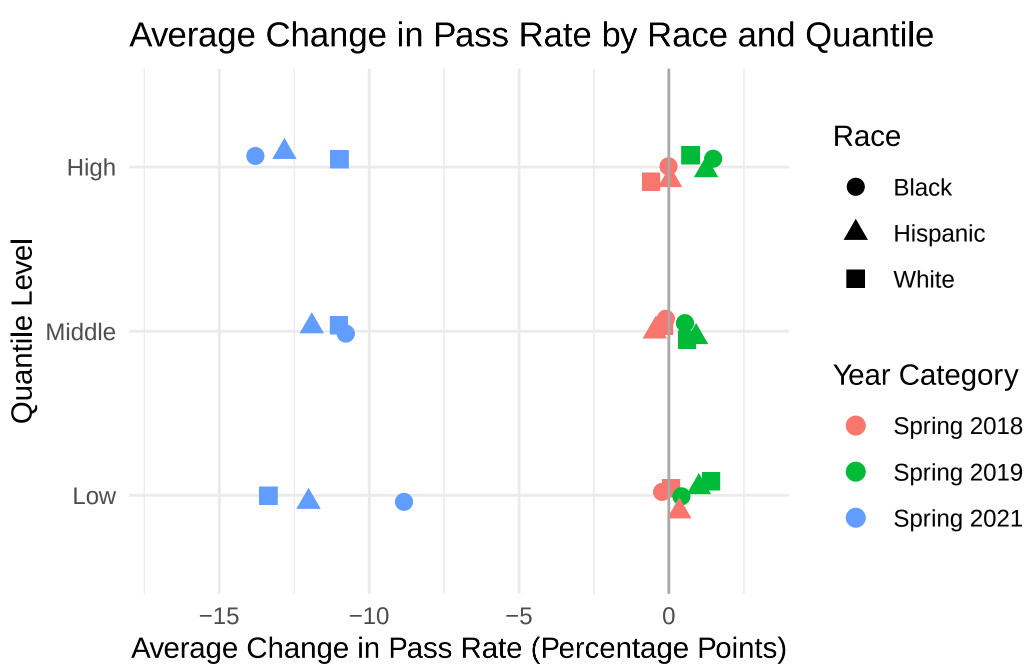

Figure 6: Update

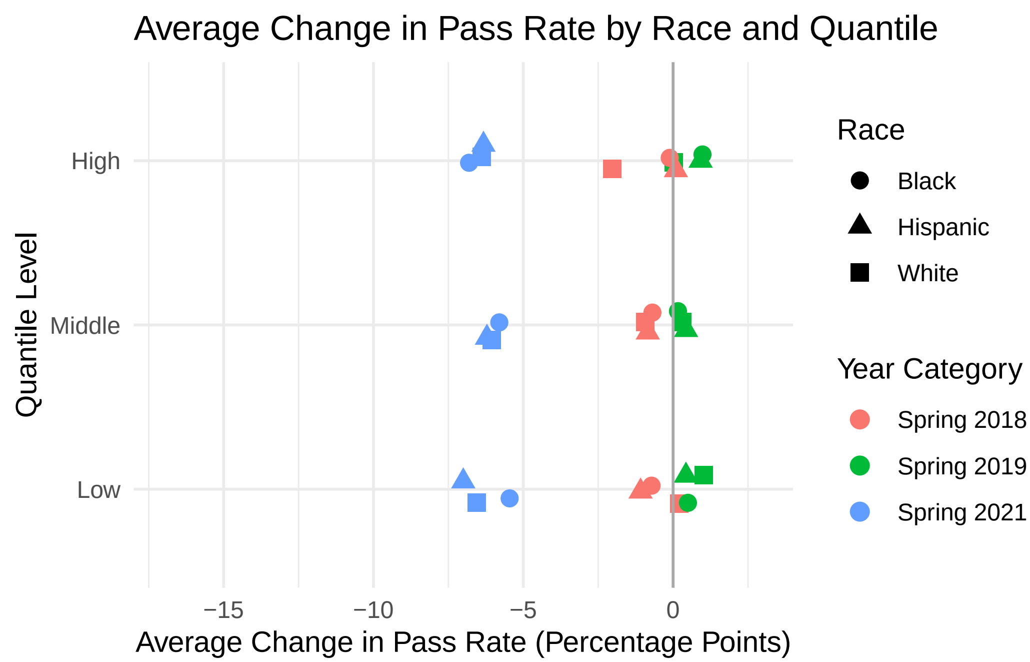

Figure 7: Update

4 Discussion

4.1 Findings

The analysis by Jack and Oster seeks to define the scope of COVID-19 school closures in the United States with respect to the global picture, and explore the correlation between the number of days of school closed with various demographic characteristics of 11 of its major states. The authors also compare the average change in pass rates in mathematics and ELA with the number of days of in-person schooling in Spring 2021 versus Spring 2018-2019. Our paper has successfully replicated and extended an important part of three of their major findings:

(1) Between January 2020 and December 2021, the United States experienced the second-highest number of days of school closed worldwide, with an average of 667 days. In fact, it falls second only to Azerjiban, which recorded an average of 669 days.
(2) Access to in-person education was unequal by demographic characteristics. More disadvantaged districts, *i.e.*, those with higher proportions of racial minorities and lower broadband usage spent more days in virtual schooling.
(3) During the 2020-2021 school year, students' pass rates on standardised assessments declined by 12.8 percentage points in mathematics and 6.8 in ELA. Declines were more severe in more disadvantaged districts, *i.e.*, those with higher proportions of racial minorities, ELL and FRPL students.

As noted by Jack and Oster (2023), it can be seen that the equity impacts are two-fold. First, schools (and students) with lower resources are less likely to have access to in-person instruction. The significant differences in virtual schooling days across districts categorized by racial demographics underscore systemic inequalities in access to in-person education. Our findings reveal that districts with higher proportions of Black and Hispanic students spent substantially more time in virtual schooling compared to those with lower proportions. This highlights a disparity wherein more disadvantaged districts, often populated by minority students, face prolonged periods of virtual learning, potentially exacerbating existing educational inequities. Second, for such groups, the consequences of a lack of in-person instruction are more severe. This may be due to a smaller investment by such schools in effective remote learning programs or due to the challenges disadvantages students face at home, such as less reliable Internet connection or fewer community resources. Accordingly, while COVID-19 school closures imposed learning loss costs on students as a whole, they affected most negatively those who were already most vulnerable. This exacerbates the already large inequality gap.

Having replicated their work, we now turn to examine its relevance to the Netherlands. We hope to gain insights into the extent of school closures in the Netherlands and its effect on student performance, with an especial focus on students from disadvantaged homes.

4.2 Relevance to the Netherlands: a Case Study

Efforts have been made worldwide to study the impacts of school closures on student learning, typically revealing significant test-score losses (R. Jack and Oster 2023). The Netherlands is interesting as a "best-case" scenario for students, with a short lockdown of 8 weeks, equitable school funding and world-leading rates of broadband access (Engzell, Frey, and Verhagen 2021). According to the Organisation for Economic Cooperation and Development (OECD) (2023), the Dutch government also entered the pandemic with strong public finances, extending generous support measures for individuals and families. The Netherlands also scores close to the OECD average in school reading and places among its top performers in mathematics (Schleicher 2018). For these reasons, the Netherlands may likely provide a lower bound on student learning loss elsewhere in Europe and the world (Engzell, Frey, and Verhagen 2021).

Between January 2020 and December 2021, the Netherlands faced 55 days of school closed, considerably shorter than the global average (362 days) and the United States (677 days) (R. Jack and Oster 2023). Key to our discussion is that the Dutch school system provides compulsory standardised assessments in mathematics, spelling and reading twice a year: halfway into to the school year in January and at the end of the school year in June. In 2020, these testing dates occurred just before and after nationwide school closures. Access to biannual test scores of 15% ($n \approx 350,000$) Dutch students between grades 4 to 7 from 2017 to 2020, published by the Student Monitoring System (LVS), thus provides a natural benchmark against which to assess learning loss. To study the data, Engzell, Frey, and Verhagen (2021) transformed these test scores into percentiles by imposing a uniform distribution separately by subject, grade and testing occasion for each of the years 2017 to 2020. In adjusting for sample composition, their results revealed a learning loss of about 3.1, 3.0 and 3.3 percentile points in mathematics, spelling and reading, respectively. The effect was equivalent to a loss of one-fifth of a school year, the same period that Dutch schools remained closed. By contrast, declines in test scores of American students between grades 3 to 8 were equivalent to a loss of about one year of learning (Binkley 2022).

The LVS also collects data on student demographics and school characteristics, as part of the national system of weighted student funding (Sanoma 2024). In comparing this data with test scores, Engzell, Frey, and Verhagen (2021) found learning losses to be up to 60% larger among students from less-educated homes, confirming the unequal costs of the pandemic on children. Moreover, they report learning losses vary considerably by school, with some schools seeing a loss of 10 percentiles or more, and some recording no loss or small gains. However, they discovered a concentration of these losses in schools with a high proportion students who are socioeconomically disadvantaged or of non-Western immigrant background. These patterns are consistent with those observed in the United States, where learning loss was also more pronounced among students of more vulnerable groups. Whether in the

Netherlands or in the United States, school closures appear to have caused socioeconomic gaps to widen.

In short, despite favourable conditions, the Netherlands appeared to have made little to no progress in learning from home. However, the impact on student learning appears to be milder than that in the United States, with Dutch students losing the equivalent of one-fifth rather than a full year of learning. These findings by Engzell, Frey, and Verhagen (2021) suggest learning losses even larger in countries with weaker infrastructure or longer lockdowns. Efforts should be made to investigate the impact of school closures on learning outcomes in these countries. Moreover, to remediate consequences on student learning, social investment strategies should be designed and implemented to enhance resilience and equity in education.

4.3 Ethical Concerns

There are ethical concerns with the collection of data on racial identity, socioeconomic status and COVID-19 cases in surveys, including data sensitivity and security. Collecting personally-identifying information (PII) may rouse feelings of anxiety and distrust, and raise important concerns about privacy and confidentiality (Ontario Human Rights Commission 2009). Individuals from vulnerable populations are particularly susceptible to such feelings of distrust in researchers (Corbie-Smith, Thomas, and St George 2002): anxiety about the research protocol, stigma, fear of disclosure, historical exploitation may be some reasons motivating a hesitancy to participate.

Moreover, stratifying test assessment scores by district-level race, ELL and FRPL shares may be problematic. The underperformance of vulnerable populations in relation to more privileged populations may reinforce discriminatory notions about their intellectual ability. These notions are harmful, having been found to impair even further the academic performance of minority groups (Steele and Aronson 1995). Indeed, Steele and Aronson (1995) have found that, when an individual is conscious of the negative stereotypes of their group, they face *stereotype threat*, becoming fearful of confirming or being judged by those stereotypes through their actions or performance. This may lead to increased self-doubt and anxiety, and thus poorer performance on assessments.

4.4 Accounting for Bias

Perhaps the most important threat to the validity of the results is the “endogeneity in schooling mode”: this is the concern that districts with more days of school closed were more affected by other pandemic-related stressors (R. Jack et al. 2023). For instance, more school closures in an area may arise in response to higher COVID-19 rates, and higher COVID-19 rates may in turn worsen the health and, thus, academic performance of students. The directed acyclic graph in Figure 8 demonstrates this phenomenon, whereby COVID-19 rates serve as a confounding variable, creating a possibly spurious relationship between school closures and academic performance. However, the number of school closures were shown to have very low correlation with COVID-19 rates (R. Jack and Oster 2023). This shows that COVID-19 rates may not be as significant of a confounding variable as initially presumed.

It is also possible that other types of lockdowns — such as the closure of counselling services or after-school activities — might have impacted student learning. This considerable difficulty of separating the effect of school closures from other pandemic-related stressors may bias the results and so affect their validity. One way in which we have helped disentangle these factors is to stratify the test assessments data by race, ELL and FRPL shares per district. By means of this “restriction” technique, we could compare test scores between more homogeneous demographic subgroups and better control for racial and demographic differences that may have impacted student learning outcomes.

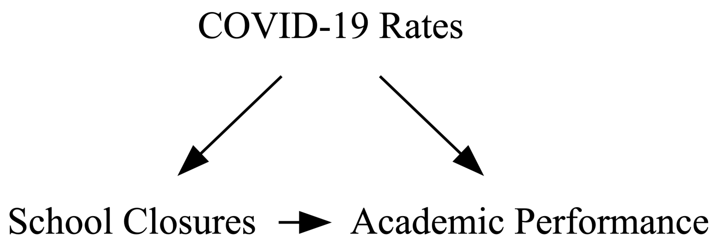

Figure 8: COVID-19 rates as a confounder, affecting the relationship between school closures and academic performance.

Important sources of bias may have also arisen from the demographic data sets obtained from the US Department of Education's Common Core of Data (Irwin et al. 2022), the US Bureau of Labour Statistics (US Bureau of Labor Statistics 2021) and the US Broadband Usage Percentages data set (Kahan 2020). In particular, there are various errors that may exist in census data, collected to determine district-level race shares across the United States. These include *coverage errors*, when a respondent is missed or counted more than once, *non-response errors*, when some information about a respondent is missing, *response errors*, where a question is misunderstood or misreported by a respondent and *processing errors*, when data is incorrectly processed ("Guide to the Census of Population, 2021 / Chapter 9 – Data Quality Evaluation" 2021). On the other hand, the US Broadband Usage Percentages data set determines broadband usage by combining data from Microsoft services (Kahan 2020). In failing to account for other popular forms of internet services, such those provided by Google, Facebook, Amazon and Apple, this data set may skew the representation of broadband usage patterns in the United States.

4.5 Limitations

Being a reproduction, much of the limitations of our analysis arise from those in the original work of Jack and Oster (R. Jack and Oster 2023). In particular, the correlations of school closures to local demographic variables, while insightful, are not comprehensive. For instance, it would have been beneficial to determine the average number of days spent in virtual school, weighted by district enrollment, and the share of the school population that is of other ethnicities — namely, Native Americans or Alaska Natives, and Pacific Islanders — so that comparisons can be made across all important ethnic classes. Other demographic variables might include the share of students experiencing homelessness, who have limited English proficiency or who are of low socioeconomic status. Determining such correlations would more rigorously test the claim that social inequity existed in the distribution of school closures.

What is more, Jack et al.'s (2023) standardised assessments data excludes information about the corresponding broadband usage and unemployment rates by district. It instead includes data of the corresponding shares of ELL and FRPL students by district. This discrepancy prevents us from telling a full story as to how the pass rates of schools located in districts with lower broadband usage and employment rates, known to face more days of school closed than the national average, were affected. We also question whether the shares of FRPL students are a good estimator of the shares of students with low socioeconomic status. Though FRPL enrollment are widely used in education research as an indicator of student poverty, there is emerging literature that it serves as a poor proxy for poverty (Fazlul, Koedel, and Parsons 2023). In fact, using multiple data sources external to the American school system, Fazlul et al. (2023) show that FRPL rates greatly exceed what would be expected from stated income thresholds for program participation by 35 to 50%. This suggests that the shares of FRPL students per district may not be the most optimal estimator of student poverty, and that more insightful conclusions could have been obtained in examining instead, for instance, the shares of students in low-income households per district.

Lastly, the state-level assessment data is obtained from only eleven states of the United States, for which district-level data is available. This data set may not be representative of the academic performance of American students as a whole for two reasons. First, state-standardised assessments remain controversial, with critics questioning whether they accurate indicate academic ability (Yang, n.d.). Certainly, they cannot capture the ways in which students learned that were not directly given in the assessment. Second, the sample of eleven states may be too small to be generalisable. Indeed, Parolin and Kee (2021) showed that wide geographic disparities existed in school closures across the country: those concentrated on the West Coast faced declines of >75% in-person visits to schools, whereas those in the Midwest faced declines of <25%. Both of these factors put into question the validity and generalisability of the state-assessment data, and could present a valuable direction for future work.

4.6 Future Research

Reports of poor mental health (MH) in youth have been on the rise in the United States for the past decade (U.S. National Library of Medicine 2022). The COVID-19 pandemic appears to have exacerbated this trend, in significantly disrupting the lives of youth in one of their most critical periods of psycho-social development (U.S. National Library of Medicine 2022). However, the evidence on the impact of school closures — as opposed to other pandemic-related stressors — on student MH in the short- and long-term remains mixed, limited or else unknown (R. Jack and Oster 2023). For instance, Bacher-Hicks et al. (2022) used internet search data to track in-person bullying and cyberbullying patterns as COVID-19 shifted to remote learning in Spring 2020. Surprisingly, they reported that *both* forms of bullying dropped by 30-35%, citing transitions to virtual learning to be likely a major explanation for these trends. In contrast, Cingel et al. (2022) used online survey data, in which over 1,000 American adolescents aged 14-16 reported their classroom learning experiences and MH symptoms during the 2020-2021 school year. In this data, virtual as opposed to in-person learning was associated with reduced MH, with those online reporting less satisfaction and social connection than those in-person. In particular, problems were particularly pronounced for students identifying as transgender or gender non-conforming.

Much more research remains to understand the full impacts of the pandemic, and of school closures, in particular, on youth MH. Whilst collecting data on MH, and differentiating between the effects of school closures and other pandemic stressors on MH, are notoriously difficult, these issues remain crucial to investigate. Findings of future search will help guide policy to remediate the consequences and provide more tailored support for affected youth that promote their well-being.

References

AEA. 2024. *The American Economic Association Portal*. *The American Economic Association*. https://www.aeaweb.org/.

Angrist, Joshua D, Parag A Pathak, and Christopher R Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5 (4): 1–27.

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*.

Aurini, Janice, and Scott Davies. 2021. "Covid-19 School Closures and Educational Achievement Gaps in Canada: Lessons from Ontario Summer Learning Research." *Canadian Review of Sociology*, May.

Bacher-Hicks, Andrew, Joshua Goodman, Joshua Green, and Matthew Holt. 2022. "The COVID-19 Pandemic Disrupted Both School Bullying and Cyberbullying." *American Economic Review: Insights* 4 (3): 353–70. https://doi.org/10.1257/aeri.20210456.

Betts, Julian R, and Y Emily Tang. 2011. "The Effect of Charter Schools on Student Achievement: A Meta-Analysis of the Literature." *Center on Reinventing Public Education*.

Binkley, Collin. 2022. "Test Scores Show How COVID Set Kids Back Across the u.s." *PBS*. https://www.pbs.org/newshour/education/test-scores-show-how-covid-set-kids-back-across-the-u-s.

Cingel, Drew P., Alexis R. Lauricella, Leslie B. Taylor, Haley R. Stevens, Sarah M. Coyne, and Ellen Wartella. 2022. "U.s. Adolescents' Attitudes Toward School, Social Connection, Media Use, and Mental Health During the COVID-19 Pandemic: Differences as a Function of Gender Identity and School Context." *PLOS ONE*. https://doi.org/10.1371/journal.pone.0276737.

Corbie-Smith, G, SB Thomas, and DM St George. 2002. "Distrust, Race, and Research." *Arch Intern Med* 162 (21): 2458–63. https://doi.org/10.1001/archinte.162.21.2458.

Engzell, Per, Arun Frey, and Mark D Verhagen. 2021. "Learning Loss Due to School Closures During the COVID-19 Pandemic." *Proceedings of the National Academy of Sciences* 118 (17). https://doi.org/10.1073/pnas.2022376118.

Fazlul, Ishtiaque, Cory Koedel, and Eric Parsons. 2023. "Free and Reduced-Price Meal Enrollment Does Not Measure Student Poverty: Evidence and Policy Significance." *Economics of Education Review* 94: 102374. https://doi.org/10.1016/j.econedurev.2023.102374.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

"Guide to the Census of Population, 2021 / Chapter 9 – Data Quality Evaluation." 2021. https://www12.statcan.gc.ca/census-recensement/2021/ref/98-304/2021001/chap9-eng.cfm.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, and et al. 2021. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behaviour* 5 (4): 529–38.

Hatch, Thomas. 2023. "Initial School Closures and Suspensions of Exams in the Netherlands: An Interview with Melanie Ehren on the Educational Response to the COVID-19 Pandemic in the Netherlands (Part 1)." International Education News. https://internationalednews.com/2023/09/27/initial-school-closures-and-suspensions-of-exams-in-the-netherlands-an-interview-with-melanie-ehren-on-the-educational-response-to-the-covid-19-pandemic-in-the-netherlands-part-1/.

Hub, COVID-19 School Data. 2021. *Percentage of School Year Spent in-Person, Hybrid, or Virtual, District Overall Shares*. https://www.covidschooldatahub.com/data-resources.

Iannone, Richard, and Olivier Roy. 2024. *DiagrammeR: Graph/Network Visualization*. https://rich-iannone.github.io/DiagrammeR/.

Irwin, Véronique, Josue De La Rosa, Ke Wang, Sarah Hein, Jijun Zhang, Riley Burr, Ashley Roberts, et al. 2022. "Report on the Condition of Education 2022 (NCES 2022-144)." Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Jack, Clare Halloran, Rebecca, and Emily Oster. 2023. *"Replication Data for: Pandemic Schooling*

Mode and Student Test Scores: Evidence from US School Districts. https://doi.org/10.3886/E168843V1.

Jack, Rebecca, Clare Halloran, James Okun, and Emily Oster. 2023. "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts." *American Economic Review: Insights* 5 (4): 173–90. https://doi.org/10.1257/aeri.20210748.

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.

Kahan, John. 2020. "United States Broadband Usage Percentages Dataset." Github Repository.

Leonhardt, David. 2022. "'Not Good for Learning'." *The New York Times*. The New York Times. https://www.nytimes.com/2022/05/05/briefing/school-closures-covid-learning-loss.html.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

NCES. 2021. *National Center for Education Statistics Common Core of Data*. https://nces.ed.gov/ccd/.

OECD. 2023. *OECD Economic Surveys: Netherlands 2021 / Chapter 1: Key Policy Insights*. OECD iLibrary. https://www.oecd-ilibrary.org/sites/a807eed2-en/index.html?itemId=%2Fcontent%2Fcomponent%2Fa807eed2-en.

Ontario Human Rights Commission. 2009. "Count Me in! Collecting Human Rights-Based Data." Ontario Human Rights Commission. November 29, 2009. https://www.ohrc.on.ca/en/book/export/html/2494.

Parolin, Zachary, and Emily K. Lee. 2021. "Large Socio-Economic, Geographic and Demographic Disparities Exist in Exposure to School Closures." *Nature News*. https://www.nature.com/articles/s41562-021-01087-8.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. https://patchwork.data-imaginist.com.

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

Richard, Viviane, Roxane Dumont, Elsa Lorthe, Andrea Loizeau, Helene Baysson, Maria-Eugena Zaballa, Francesco Pennacchio, et al. 2023. "Impact of the COVID-19 Pandemic on Children and Adolescents: Determinants and Association with Quality of Life and Mental Health-a Cross-Sectional Study - Child and Adolescent Psychiatry and Mental Health." *BioMed Central*. BioMed Central. https://capmh.biomedcentral.com/articles/10.1186/s13034-023-00563-5.

Sanoma. 2024. "The LVS Bes Project (English)." Bureau ICE; https://www.bureau-ice.nl/lvs-bes-engels/.

Schleicher, Andreas. 2018. *PISA 2018: Insights and Interpretations*. OECD Publishing.

Steele, Claude M, and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811. https://doi.org/10.1037//0022-3514.69.5.797.

U.S. Department of Commerce. n.d. "Non-Fiscal Surveys of the Common Core of Data (CCD)." https://www.census.gov/econ/overview/go2000.html.

U.S. National Library of Medicine. 2022. "Child and Adolescent Mental Health." 2022 National Healthcare Quality and Disparities Report [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK587174/.

US Bureau of Labor Statistics. 2021. "County-Level Unemployment Data (La.data.64.county)." US Department of Labor.

WHO. 2020. "WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020." https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

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, and Jennifer Bryan. 2023. *Readxl: Read Excel Files*. https://readxl.tidyverse.org.

World Bank Group. 2023. "World Development Indicators." https://databank.worldbank.org/source/world-development-indicators.

Wu, Di, Jianyun Lu, Yanhui Liu, Zhoubin Zhang, and Lei Luo. 2020. "Positive Effects of Covid-19 Control Measures on Influenza Prevention." *International Journal of Infectious Diseases*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7151395/.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Yang, Quanya. n.d. "Should Standardization Tests Be Used to Assess Student Ability? – an Evaluation on Effects of Standardization Tests." *Journal of Education, Humanities and Social Sciences*. https://drpress.org/ojs/index.php/EHSS/article/view/4268.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.