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) done by students in grades 3-8 across 21 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

---

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 (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 (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 (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 is primarily grades 3 to 8 American students from 21 US states. The estimate which we intend to replicate is the impact of the pandemic on academic performance of these students on standardised assessments. Their findings suggest that considerable declines in test scores occurred during the 2020-2021 academic year, with declines larger in school districts with less in-person instruction. 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) (Jack and Oster 2023).

Our paper seeks to assess the robustness and generalisability of Jack and Oster's work. Excepting one aspect noted here, 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 or Hispanic.

The only main discrepancy between our analysis and the original paper is in the second research claim, where we found districts with lower broadband usage to in fact spend less of the school year in virtual schooling. We accounted for factors which may have produced this discrepancy, and concluded it to be a consequence of their lack of technological preparedness.

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" (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 (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 made available by the Oxford COVID-19 Policy Tracker, a project that collected information on COVID-19 policy measures in 185 countries for the years 2020, 2021 and 2022 (Hale et al. 2021). Data were collected by a team of over 1500 volunteers at Oxford's Blavatnik School of Government from publicly available sources– such as government press releases, briefings, international organisation reports, and credible news articles. For 185 countries, this dataset assigns a daily Tracker score between January 1st, 2020, and December 31st, 2022. This daily Tracker score is a composite integer measure of nine response metrics that evaluate the stringency of a country's COVID-19 response (Hale et al. 2021); in particular, a value of 2 or 3 denotes a school closure (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. Note that unique NCES ID numbers are automatically assigned to districts upon being reported to the US Department of Education by their state education agency via the platform EDFacts (NCES 2021).

`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). Thanks to partnerships with state education authorities and organisations, NCES is authorised to collect such data by conducting numerous surveys with public schools and local education agencies (Encyclopedia 2024). Its National Forum on Education Statistics then collects, processes and reports elementary and local education statistics (Encyclopedia 2024).

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). According to Kahan (2020), 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. In using the Federal Communications Commission definition of broadband, being 25 megabits per second per download (FCC 2021), a count is then made of the number of devices that connect to the Internet at broadband speed per zip code. After making such measurements, noise is added to data aggregations using the open-source SmartNoise platform, thereby preventing leakage about specific individuals in the dataset and ensuring differential privacy ("SmartNoise" 2023).

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 replication package of "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts" (Jack et al. 2023a), published in the American Economic Association's *AER: Insights* (AEA 2024). Data are obtained from the Departments of Education for the 21 major American states, being Arkansas (AR), Colorado (CO), Connecticut (CT), Georgia (GA), Idaho (ID), Kansas (KS), Louisiana (LA), Massachusetts (MA), Minnesota (MN), Mississippi (MS), Missouri (MO), New Hampshire (NH), Ohio (OH), Pennsylvania (PA), Rhode Island (RI), South Carolina (SC), South Dakota (SD), Virginia (VA), Wisconsin (WI), West Virginia (WV) and Wyoming (WY). Each state collects data via surveys from constituent schools, school boards and school authorities (US Department of Education 2023). The dataset contains extensive information on school districts across these 21 states, including shares of students by race, shares of schooling done in-person, and changes in average test scores in both Mathematics and ELA from the previous year across the years 2018-2019 and 2021-2022. Note no information on test scores are available for 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 shows the average number of school closures by region between the start of January 2020 and the end of December 2021. The Oxford COVID-19 Policy Tracker (Hale et al. 2021) assigns a score to each of the 185 countries and the World Bank (World Bank Group 2023) provides a one-to-one correspondence between the country and the broader region to which it belongs: sub-Saharan Africa, Europe and Central Asia, and so forth. Details on how this figure was achieved follow.

Since `region_data.csv` did not feature the region to which each of the 185 countries belonged, we created the dataset `country_region.csv`. This assigned to each country its corresponding region, as defined by the World Bank (World Bank Group 2023). In doing so, we could have information both on the COVID-19 Tracker score of each country for each day, as well as the region in which the country is located.

We followed the workflow of Jack and Oster (Jack and Oster 2023), by likewise omitting in `regions.csv` all information from the year 2022. Only columns of interest to the analysis were then kept, including the country name, (empty) region name, date, and COVID-19 Tracker score. We also identified and amended a minor spelling error in the name of `Faroe Islands`, originally misspelled

as `Faeroe Islands`. Thereafter, we 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.

We regrouped the cleaned data by region name and country, summed 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 (2023)). We then calculated the mean number of days closed by region by computing the average number of days closed for all countries in a given region.

To gain an understanding of how COVID-19 schooling measures implemented in the United States compared with those elsewhere in the world, we created Table 1. This table provides a snippet of the top 5 countries with the highest number of school closures in 2020-2021, featuring the US in second place. We used the same cleaned dataset as before, regrouped data by country, summed the number of instances where a country registered a COVID-19 Tracker score of 2 or 3 and isolated the top 5 countries where these numbers were the largest.

2.2.2 Average Number of School Closures by Local Demographic Characteristics

Figures 2 and 3 explore correlations between the average number of school closures with two county- and district-level characteristics, these being county-level broadband usage and district-level race shares of four major ethnic classes– White, Black, Asian and Hispanic. For both demographic characteristics, we followed Jack and Oster's (2023) work, by assigning to each county or district a quantile of "High" or "Low", depending on whether its levels for said characteristic were above or below the state's median, respectively. By doing so, we allowed the results to be driven by a variation within states, as opposed to variation across states.

We drew on the `nces_district_enrollment_2018_2020.csv.zip` dataset, which reports the total student enrollment by grade, sex and ethnic class. A handful of enrollment cells had no value, which were subsequently replaced by 0s. Being specifically concerned with race data, we calculated for each district the total number of White, Black, Asian, and Hispanic students enrolled. We then separated data according to race, classifying districts as "High" or "Low", depending on whether their race share was above or below their state's median. Thereafter, we matched each district with those contained in `District_Overall_Shares.csv`, so as to obtain information on its average number of days closed. We found there was some discrepancy, with the number of districts in `nces_district_enrollment_2018_2020.csv.zip` being larger than that in `District_Overall_Shares.csv`, which we resolved by dropping all instances of districts not present in both datasets. After merging these two datasets, we averaged the number of virtual days across districts, classified in the "High" and "Low" quantiles. This resulted in four data sets, one for each race considered.

Here, we differed from the methodology employed by Jack and Oster (2023) in one specific way. In their workflow, Jack and Oster (2023) also weighted the average number of days of virtual school by total district enrollment. In this way, districts with larger enrollments had greater impact on the overall average than those with smaller enrollments. In our analysis, we omitted such weighting, instead treating each district equally, regardless of size. This offered a more balanced view of the educational landscape, without overemphasising larger districts.

Next, we drew on `broadband_data_2020October.csv` and `nces_district_directory_2018.dta` to determine correlations of days spent in virtual schooling with broadband usage. With the former dataset, we obtained the broadband usage across all US districts; with the latter, we regrouped all districts belonging to each US county and averaged their broadband usages. We then assigned to each county a quantile of "High" or "Low", according to whether their broadband usage was above or below their respective state's median. Jack and Oster (2023) did not mention their methodology to find the number of virtual days for each county, given that data were available only at the district-level. To remedy this, we simply regrouped the districts in `District_Overall_Shares.csv` into their

respective counties and averaged the number of virtual days across all districts for each county labelled as "High" and "Low."

2.2.3   Standardised Test Assessment Scores

Figure 4 and Figure 5 explore the relationship between the share of in-person learning and change in pass rates in standardised tests in Mathematics and ELA respectively. In particular, they examine changes in test pass rates across 3 years (2021, 2019, and 2018) from the year prior (excepting the case of 2021, where changes are relative to 2019).

To achieve this, we made use of the `scores_lm_demographics.dta` dataset. However, we first converted the dataset into a `.csv` file, `scores_lm_demographics.csv`, to be used in the cleaning and analysis processes. From there, we cleaned the data by only selecting the columns of interest, those being the share of in-person learning, the subject of the test, the changes across each of the three years of interest, the number of students enrolled in each school district, the state each schooling district is in, and the shares of each of the three races of interest (White, Black, Asian and Hispanic). To generate the aforementioned figures, we first calculated the total number of students enrolled across all school districts. We then filtered by subject, so as to deal with one subject at a time when creating the two figures. We proceeded by defining four in-person categories of 0-25%, 25-50%, 50-75%, and 75-100%, after which we mutated our filtered data by assigning each school district one of the four categories. We then proceeded to group the data by year (one of the three previously mentioned), and by in-person category so that we were able to calculated the weighted average change in test pass rates for each in-person category in each year. Finally, we ungrouped the data to calculate the overall, and thus un-weighted, average change in test pass rates for each of the three years. This was repeated for both Mathematics and ELA test pass rates by simply changing the filtering of the data to only include one of the two.

Figure 6, Figure 7 examine these test pass rate changes from a different lens. Instead of considering the share of in-person learning, it explores the possibility of a systematic difference in the changes in test pass rates based on the relative share of various races. In particular, the figures examine the correlation between changes in test pass rates and the relative quantile share of White, Black, and Hispanic students which is one of "Low" or "High". As before, this is also done for each of the three years of interest for each of the three races.

To generate this graph, we once again employed the cleaned `scores_lm_demographics.csv` file as was used to generate Figure 4 and Figure 5. We first filtered by subject to only consider one at a time, after which the total enrollment across all school districts was calculated as well as the two quantiles, "Low" and "High", for each of the three races relative to the average shares across all school districts. From there, we grouped the data by year and quantile for each of the three races, then calculated the weighted average change in test pass rates for each subject and quantile, across all races and years. From there, we simply combined the summary statistics for each of the races and their respective quantiles, and showcased them in the figures, differentiating by race through color, and year through shape.

3 Results

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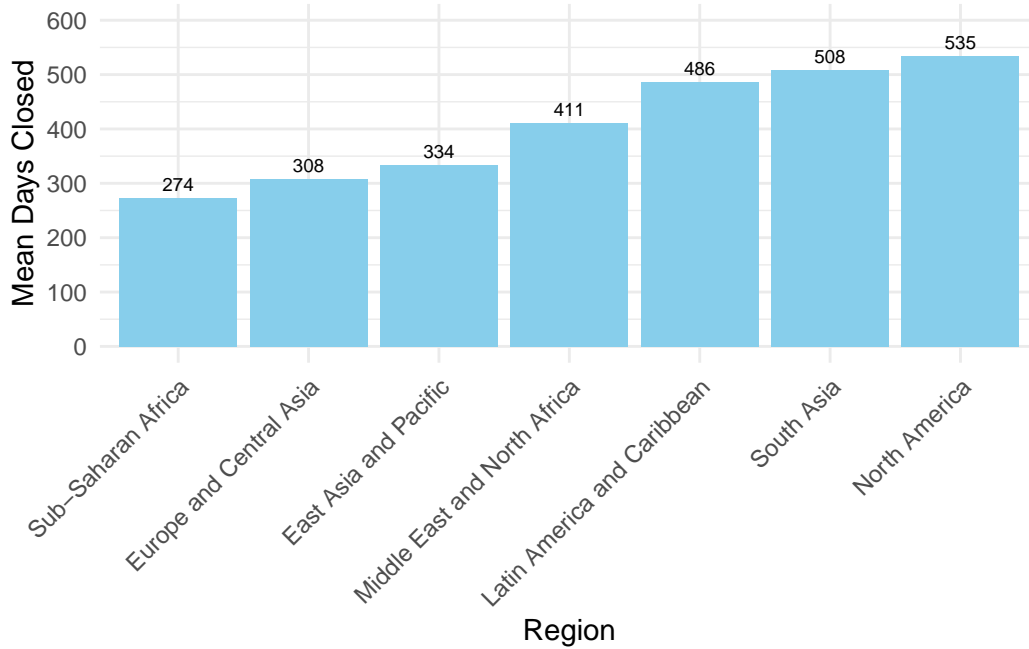

Figure 1: Average number of days of school closed, by region, between January 2020–December 2021.

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.

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

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.

3.2 Average Number of School Closures by Local Demographic Characteristics

Data from CSDH (Hub 2021), NCES (NCES 2021) and Microsoft (Kahan 2020) were used to determine correlations between the average number of virtual schools days, and certain demographic characteristics. We reconstruct these results in Figures 2 and 3. These show the average number of virtual days for counties or districts that have "High" or "Low" levels of broadband usage and various race shares, respectively. Note that we enriched Jack and Oster's (2023) analysis on race by also considering the shares of White and Asian students, so as to better understand the nuanced impact of racial demographics.
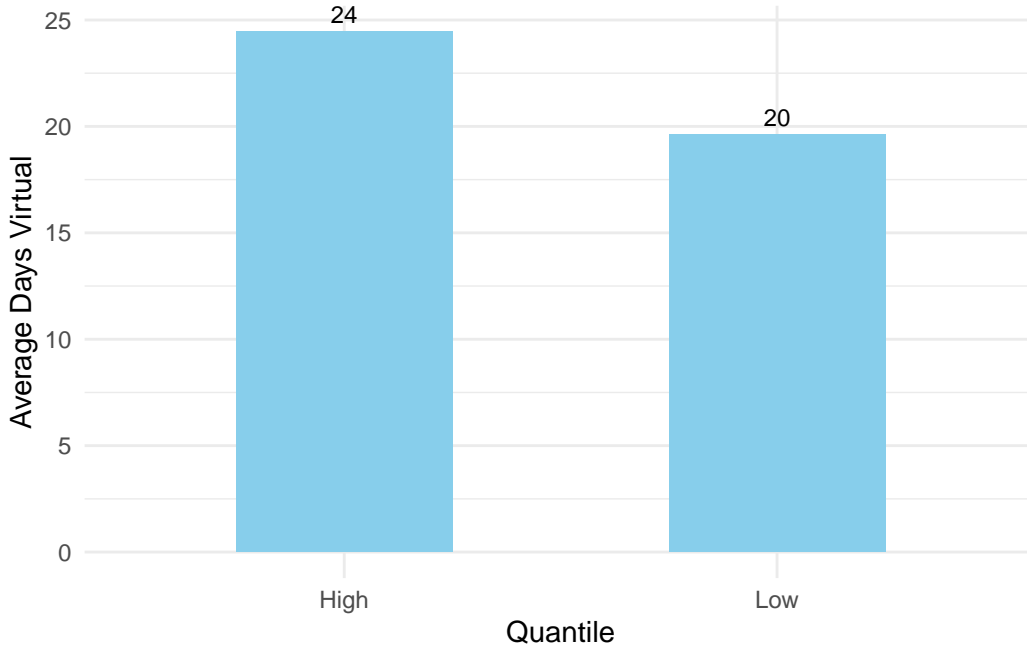

Figure 2: Average days in virtual school by county broadband usage.

Figure 2 shows the average number of days with school closed in counties with "Low" and "High" broadband usage to be 20 and 24, respectively. Note this result significantly differs to the one obtained by Jack and Oster (2023), who determined the average for "Low" and "High" counties to be 55 and 48 days, respectively. Moreover, our numbers point to an opposite trend– that counties with lower broadband usage had fewer virtual days. The discrepancy in data can arise due to a number of important reasons. First, we did not weight the data by enrollment. Second, whilst we chose to neglect all cells with missing data, Jack and Oster (2023) may have approached this problem in a different manner, which they did not document. Lastly, our methods to find the average number of virtual days for a county may differ to theirs. As noted previously, we regrouped districts into counties, averaging across the number of virtual days for districts into counties, whereas Jack and Oster (2023) left no note to document their own approach.

Our conclusions still point to the fact that potential disparities in access to and the usage of online education resources exist. The observed differences in virtual learning days with the level of broadband usage suggests that regions with less technological preparedness may face additional challenges in accessing effective virtual learning experiences.
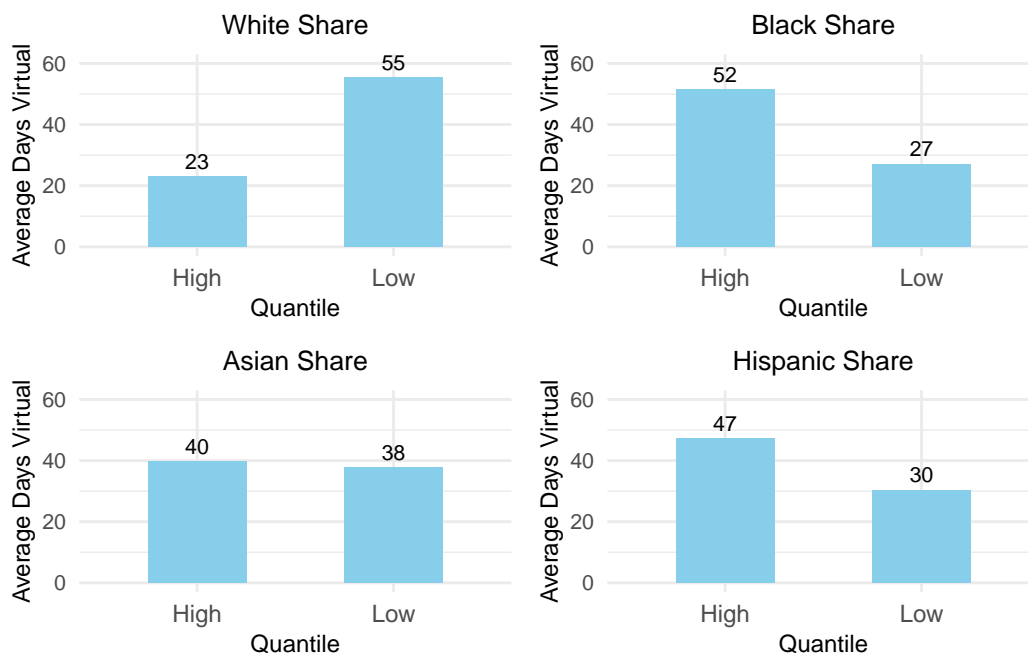
Figure 3: Average days in virtual school by district-level race shares.

Figure 3 shows how the average number of days with school closed in districts vary "Low" and "High" shares of Black, Hispanic, Asian and White students. For Black students, these numbers were 52 and 27, respectively, differing slightly to those from Jack and Oster (2023) who determined 56 and 36 days, respectively. For Hispanic students, we found an average of 47 and 30 days, respectively, again differing to those from the original paper, which reported 53 and 43 days, respectively. In spite of the numerical differences, our results point to conclusions consistent with those from Jack and Oster (2023), who also observed districts with higher shares of Black and Hispanic students to spend more of the 2020-2021 school year in virtual schooling.

We extended Jack and Oster (2023)'s research by also looking at the shares of White and Asian students. We found that districts with "High" and "Low" shares of White students had an average of 55 and 23 days in virtual schooling, respectively. Likewise, we found that districts with "High" and "Low" shares of Asian students had average of 40 and 38 days in virtual schooling, respectively. These results point to the following: White students appear to be further enfranchised by having greater access to in-person learning, in contrast to historically and currently disadvantaged Black and Hispanic students. No such conclusions can be drawn for Asian students, for whom the variation in the number of virtual days between "High" and "Low" districts is too small to be considered.

Taken together, in-person education was inequitably distributed across demographic characteristics, with racially disadvantaged districts spending more time in virtual schooling (Jack and Oster 2023). By contrast, districts with less technological preparedness spent less time in virtual schooling, which may arise as a consequence of a lack of ability to provide quality on-line learning resources. For these differing reasons, both groups can be said to have been further disenfranchised by the pandemic. This pattern likely reinforced educational inequities between vulnerable and more privileged groups.
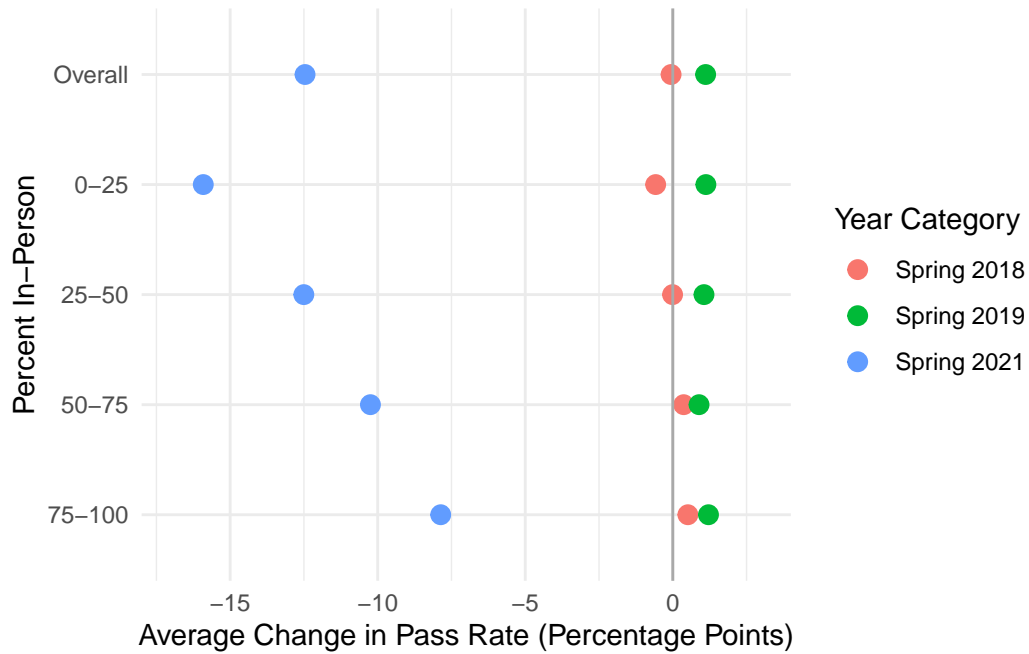
3.3   Standardised Test Assessment Scores

Figure 4: Enrollment-weighted average change in pass rates in Math standardised tests by the percent of in-person learning.
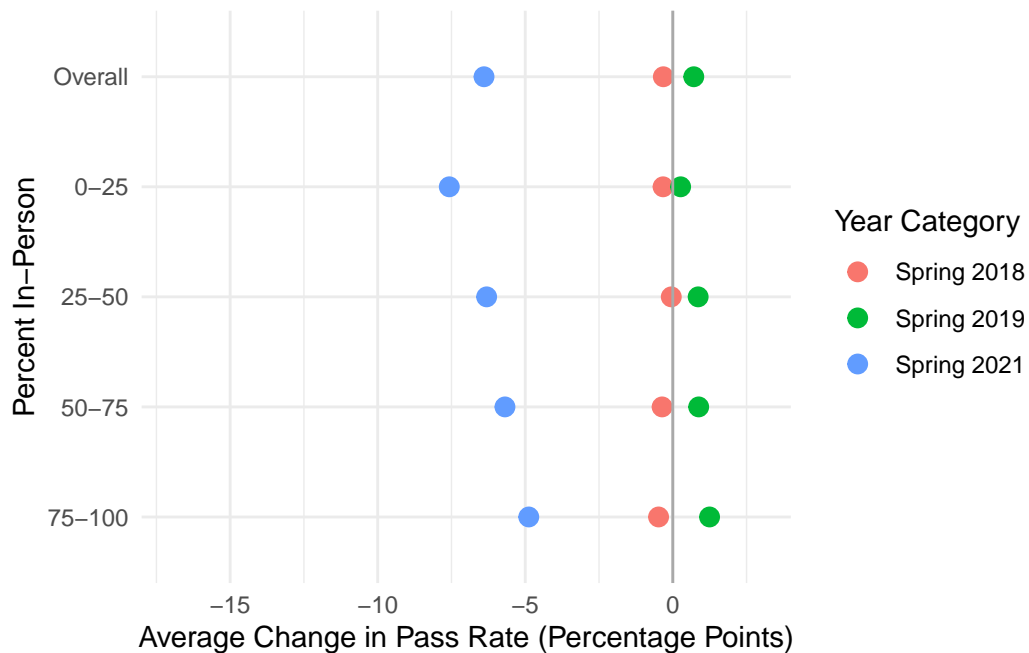

Figure 5: Enrollment-weighted average change in pass rates in ELA standardised tests by the percent of in-person learning.

Finally, data from "Pandemic Schooling Mode and Student Test Scores: Evidence from US School

11

Districts" (Jack et al. 2023a) was used to examine changes in district-level pass rate data from 21 state standardised assessments, taken in the spring of 2018, 2019 and 2021. Both Figure 4 and Figure 5 both provide information on the change in pass rates in Math and ELA standardised tests from the previous year, respectively, taking into account the share of in-person learning done by various school districts and across the three years of interest, each indicated by its respective color. 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. We see that in Spring 2021, there is a clear fall in the pass rates overall in both subject areas from 2019 (note that there were no standardised assessmments in 2020), though this fall is much more pronounced in Math than ELA (-12.8 percentage points compared to the -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).

We also see evidence that a lack of in-person learning has a detrimental effect on student's learning, as seen by the positive correlation between the share of in-person learning and the average change in pass rates in both Math and ELA in Spring 2021. One can note the lack of significant change and correlations between in-person learning and changes in pass rates for both subjects in Spring 2019 and Spring 2018, possibly due to schooling districts with lower shares of in-person learning being better equipped for this learning style given that this was prior to the pandemic, meaning schools were less likely to have been forced to switch to online or hybrid schooling, and instead were able to do so on their own accord.
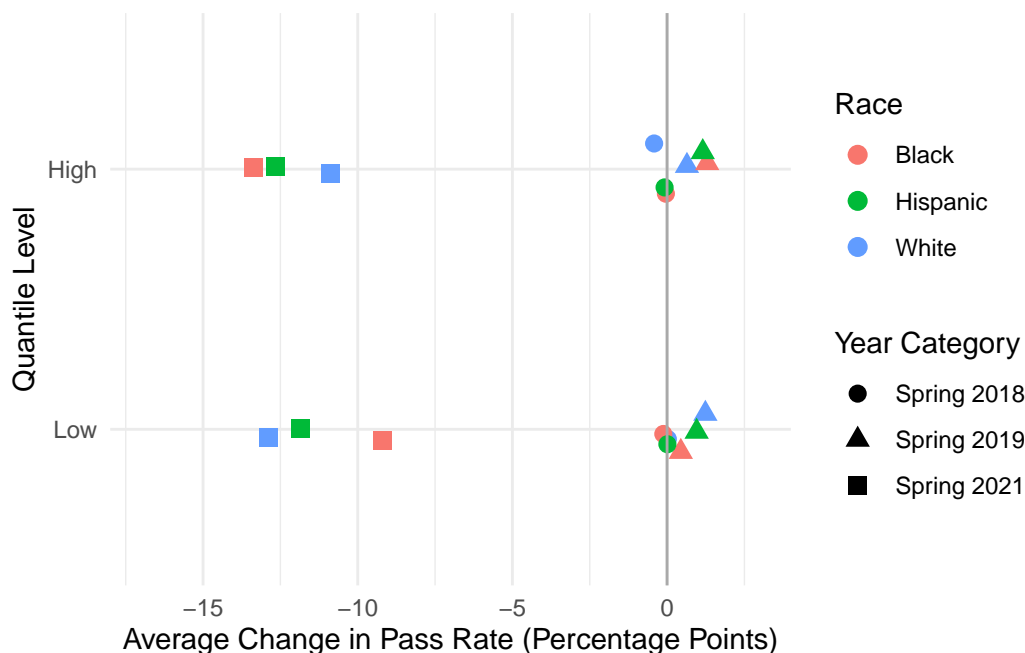

Figure 6: Enrollment-weighted average change in pass rates in Math standardised test scores by the quantile share of White, Black, and Hispanic students

In line with some of our previous discussions, we looked towards building upoing Jack and Oster's work by examining the potential systematic disparities between relative student racial compositions and the average percentage point change in test pass rates. In particular, both Figure 6 and Figure 7 provide information on the changes in test pass rates across the three years of interest from the year prior as a function of the race quantiles that each schooling district belonged to for each of the share of White, Black, and Hispanic students. By grouping the test score data by state then calculating the

Figure 7: Enrollment-weighted average change in pass rates in ELA standardised test scores by the quantile share of White, Black, and Hispanic students

average share of each of the three races mentioned, we were able to assign each schooling district a value of "High" or "Low" for each of the three races, detailing whether the particular schooling district had a share of students of a particular race greater than, or lower than, the median share across the 21 states, respectively. The figures distinguish between the year and race in question by shape and color, respectively.

By examining both figures, we find evidence that schooling districts with a relatively higher portion of either Black or Hispanic students featured greater drops in test pass rates on average, in both subjects, relative to schooling districts with lower portions of Black or Hispanic students in Spring 2021. Since we have established in the previous Results section that schooling districts with relatively higher shares of Black or Hispanic students typically had a higher number of virtual days than schooling districts with relatively lower shares, we effectively find evidence that in-person learning has greater benefits for districts with larger shares of Black and Hispanic students.

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 21 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. With one exception of note, 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. In particular, disadvantaged districts with higher proportions of racial minorities spent more days in virtual schooling. In contrast to Jack and Oster (2023), districts with lower broadband usage were found in our analysis to spend less time in virtual schooling, likely due to their diminished ability to deliver quality online resources.

(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.

As noted by Jack and Oster (2023), it can be seen that the equity impacts are two-fold. First, as shown in Figure 3, schools (and students) of more vulnerable racial backgrounds – especially, Black and Hispanic – are less likely to have access to in-person instruction. Second, as shown in Figure 6 and Figure 7, for such groups, the consequences of a lack of in-person instruction on academic performance 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 in educational outcomes.

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 (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) (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 may be problematic. The under-performance 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 (Jack et al. 2023b). 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 (Jack and

Oster 2023). This shows that COVID-19 rates may not be as significant of a confounding variable as initially presumed.

$$\text{COVID-19 Rates}$$
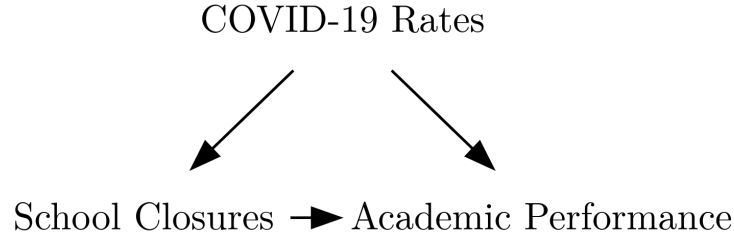
School Closures → Academic Performance

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

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 shares per district. By means of this "restriction" technique, we could compare test scores between more homogeneous demographic subgroups and better control for racial differences that may have impacted student learning outcomes.

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 (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. (2023a) standardised assessments data excludes information about the corresponding broadband usage and test scores of Asian students by district. It instead includes data of the corresponding shares of English Language Learners (ELL) and students eligible for Free and Reduced Price Lunch (FRPL) by district, two demographic characteristics we did not consider in our analysis. 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 higher shares of Asian minorities, the former of which is known to face more days of school closed than the national average, were affected. As an aside, we might also question whether the shares of FRPL students were a good estimator of the shares of students with low socioeconomic status in Jack et al. (2023a)'s paper. 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 in the Jack et al. (2023a) paper, and that more insightful conclusions could have been obtained in examining instead, for instance, the shares of students in low-income households per district.

Another important limitation is the discrepancy in the state information provided by the datasets and in the methodology used to analyse them. In particular, whilst the datasets to create Figure 3 contained data on all 48 states, Figure 6 and Figure 7 had data on only 21 states. Moreover, the results in Figure 6 and Figure 7 were calculated by weighting the averages by total district enrollment, whereas no such weightage was performed for Figure 3. These such differences limit our ability to compare the results represented in the figures.

Lastly, the state-level assessment data is obtained from 21 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, Jack and Oster (Jack and Oster 2023) gave no justification for their choice of these 21 states in particular, and it may be the case that the sample is not representative of the United States as a whole. 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%. It remains to examine whether these 21 states form a representative sample of the wide range of school closures across the country. 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 (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.

Encyclopedia. 2024. "National Center for Education Statistics." https://www.encyclopedia.com/education/encyclopedias-almanacs-transcripts-and-maps/national-center-education-statistics; Cengage. https://www.encyclopedia.com/education/encyclopedias-almanacs-transcripts-and-maps/national-center-education-statistics.

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.

FCC. 2021. "Fourteenth Broadband Deployment Report." https://www.fcc.gov/reports-research/reports/broadband-progress-reports/fourteenth-broadband-deployment-report.

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, Rebecca, Clare Halloran, James Okun, and Emily Oster. 2023b. "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.

———. 2023a. "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts." *American Economic Review: Insights* 5 (2): 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.

"SmartNoise." 2023. *SmartNoise*. SmartNoise. https://docs.smartnoise.org/.

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. 2000. "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.

US Department of Education. 2023. "An Overview of the u.s. Department of Education." https://www2.ed.gov/about/overview/focus/what.html.

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.