# The Effect of Public School Lockdown Measures on Test Scores

Dui J. Davidsson

April 22, 2025

**Abstract**

This paper examines the impact that COVID-19 school lockdown measures on district-level standardized math test scores in the United States. Using newly released 2024 data from the Stanford Education Data Archive (SEDA), I combine grade-year standardized (GYS) test score measures with district-level instructional-mode data from the COVID-19 School Data Hub to estimate the marginal effects of different remote learning exposure during the 2020–2021 school year. Two empirical strategies are employed. First, a two-way fixed effects OLS model estimates the effect of additional months spent in remote learning. Second, I use propensity score matching combined with a differences-in-differences (DiD) framework that compares school districts exposed to remote learning with socio-economically similar districts that remained primarily in-person during the school year. The OLS results indicate that each additional month in remote learning was associated with approximately a 0.044 grade-level decline in math achievement growth relative to the 2019 national average. The DiD estimates suggest that districts exposed to lockdown measures experienced an average post-pandemic decline of approximately 0.160 grade levels, with substantially larger effects observed for Black and minority students. These findings suggest that pandemic-related instructional disruptions negatively affected student achievement and may have contributed to pre-existing inequalites in educational achievement between marginalized and non-marginalized students. Our results further highlight the importance of considering long-term academic externalities when evaluating large-scale educational disruptions.

Following the global outbreak of the COVID-19 pandemic and lockdown orders, schools across the United States considered different teaching methods for the 2020/2021 school year. With concerns about the risks of spreading COVID-19, many schools closed down, adopting a fully online or hybrid classroom environment. Since then, there has been a considerable public debate about the effects that online learning has had on students. Many argue that online learning environments negatively affected student learning experiences. This is aided by studies surveying U.S students and teachers that have found online learning created for increased classroom stress (Reynolds et al., 2022), as well as decreased socialization and learning satisfaction (Cingel et al., 2022). For this paper, we study the effect of school COVID-19 learning measures on test scores. Our research is guided by the question of whether COVID-19 online learning measures were detrimental to the education of students, and whether a supposed detriment exaggerated existing inequalities between students of different socio-economic backgrounds, including ethnicity and poverty levels. In doing said research, this paper contributes to the literature in two notable ways. Firstly, this research uses newly released test score from the Stanford Educational Data Archives (SEDA) that includes test scores for 2024. Secondly, while studying whether there exists a statistically significant effect of COVID-19 learning measures on test scores, I also exploit data on durations of remote learning to study the marginal effects that additional duration in lockdown had on test scores. This is important given the heterogeneity of remote learning measures: districts that closed down during the 2020-2021 school year were closed down for differing durations of time, and it's reasonable to assume that the effect of remote learning would be stronger for schools that spent a longer duration in remote learning conditions.

To accomplish this research, I primarily combine standardized test score data from SEDA and district-level learning model tracking data to observe what districts spent time in lockdown over the 2020-2021 school year, and link both remote learning and the duration of remote learning to district standardized test scores. To control for various school characteristics, the effects of Covid-19, and political demographics, I further merge test scores data with Common Core Data (CCD) from the National Centre for Educational Statistics (NCES), Covid-19 tracer data from the *New York Times*, and finally 2020 county election data from the MIT data labs.

This paper studies the effect of COVID-19 learning measures on test scores through two separate research designs. Firstly, I employ an ordinary least squares (OLS) regression with district controls and two-way fixed effects to study the marginal effect of an additional month in remote learning on test scores. As will later be noted, one of the limitations of this design is that whether a district had most of its schools in lockdown is not randomly assigned: schools that were in lockdown for most of the 2020-2021 school year were more likely to be larger districts with higher levels of poverty, single-motherhood, and other social issues that are associated with test scores. Thus, to observe the absolute effects of COVID-19 lockdown measures[1] on test scores, I use propensity score matching to match districts with similar socio-economic characteristics, and then employ a differences-in-differences (DiD) design to compare differences in test scores for districts who took remote learning measures and those who didn't, before and after the 2020-2021 school year.

Using the OLS regression, we find that an additional month of remote learning had a statistically significant negative effect on a districts test score. In particular, an additional month a district spent in lockdown for the 2020-2021 school year was associated with 0.047 grade level

[1] For this paper, we use "lockdown" and 'remote learning" interchangeably.

decrease in math test score growth relative to the 2019 national average across all students. For Black, Hispanic, and economically disadvantaged students, we do not find a statistically significant decrease in test scores, though it's hypothesized that this explained by the lower sample sizes for these regressions, as well as the downwards bias of our model.

For the DiD regression, we find that across all students there is a statistically significant difference in test scores between lockdown and non-lockdown schools post-COVID-19. We observe that districts in lockdown for any period during the 2020-2021 school year had a 0.175 lower math grade level post-COVID-19 relative to comparable districts that didn't spend time in lockdown. Similar statistically significant differences are also observed for Black students, Hispanic students, and economically disadvantaged students. Black students in particular had a disproportionately larger decrease, having a 0.290 lower grade level in math post-COVID-19 relative to comparable districts that didn't spend time in lockdown. This would appear to suggest that not only COVID-19 remote learning orders had a negative effect on test scores broadly, but may have exacerbated existing inequalities in academic achievement between minority and non-minority students.

The implications of paper should hope to illustrate the potential far-reaching consequences of lockdown measures on child education and learning outcomes. While we make no comment on the justification or ethics of school lockdown measures or other COVID-19 health measures during the pandemic, policy makers should be aware of the externalities to education that can be caused by major class disruptions, such that decisions on school learning measures for a future pandemic or other major domestic event can be best weighed.

## Literature Review

This project follows work done by Dee et al. (2023), which studied the effects of COVID-19 school learning measures on enrollment levels. Constructing data on district enrollment levels from the National Centre for Educational Statistics (NCES) and state government sources, Dee et al. uses a differences-in-differences design to compare enrollment levels for districts with differing COVID-19 learning measures before and after the 2020-2021 school year. Using this approach, Dee et al. found that districts with remote-only learning measures suffered a 1.1 percentage point decrease in their overall enrollment figures; specifically, a 42 percent increase in the disenrollment rate from the district. These effects were concentrated among kindergarten and elementary schools, with middle and high schools experiencing no statistically significant effect in enrollment. For test scores, there is Kuhfeld et al. (2022), which provides statistics on changes in test scores between pandemic lockdown measures, showing that pandemic in-person school disputed lead to an average .20-.27 and .09- .18 standard deviation decrease in math and reading scores nation-wide. One such paper studying the effect of lockdown measures on test scores is Heise (2024). Using test score data from the Stanford Education Data Archive (SEDA), and data on district COVID-19 learning measures, Heise uses a similar differences-in-differences design to study the effect of the lockdown measures on test scores.

**Data Sources**

For this project, we utilize the following data sources for estimates and controls in our methodology.

*Stanford Educational Data Archives (SEDA) 2024*. — Complied by Educational Opportunity Project at Stanford University, SEDA2024 contains district-level data on standardized test scores by grade, subject (math, reading), and year. As required by U.S federal law, each state is required to have a standardized testing program through grades 3-8 in both math and the reading language

arts (RLA) (Reardon et al., 2025). Through the EDfacts data system, which compiles state accountability reports across U.S schools districts, SEDA2024 takes district-level test score data and supplements it with state assessment data from the National Assessment of Educational Progress (NAEP) to construct standardized subject-grade-year mean estimates for test scores across all available U.S school districts (Reardon et al., 2025). For this paper, we exploit SEDA2024's long grade-year-standardized (GYS) dataset, which consists of panel data for each district and their GYS test scores between 2009-2024, seperated by grade and subject. GYS estimates measure the amount of which a student's test score in the district is expected to grow on the NAEP relative the 2019 national average, with one unit representing a growth of a single grade-level (Reardon et al., 2025). Not only does SEDA2024 contain GYS estimates for all students of a particular grade and subject, but also includes GYS estimates for various subgroups, such as black students, white students, Hispanic students, and economically disadvantaged (ED) students. All these GYS estimates are used to estimate the effect of remote learning measures on these subgroups, along with the general students population.

Along with GYS estimates, SEDA2024 also includes a separate dataset containing data on various district covariates (number of students, percent black, percent Hispanic, district poverty rates, etc.) between 2009-2024. This dataset is merged to the primary SEDA2024 GYS estimates dataset to provide various district-levels controls for our estimating equation. Taking SEDA2024, we include only the years around the 2020-2021 school year (2016-2024), and districts which contain at least 1000 students. For our regressions, we focus on GYS estimates for math scores.

*Covid School Data Hub (CSDH) District In-Person, Hybrid, or Virtual* — The CSDH contains district level monthly data on the percentages of schools in a district that were teaching remotely/virtually, in-person, or hybrid classes during the 2020-2021 school year (CSHD 2023). Lead by Dr. Emily Oster at Brown University, the CSDH collected information on district learning measure statuses through individual requests to state education agencies. Using the percentages of schools under remote learning, we count the number of months any school district had over 80% of schools in remote learning to provide an estimate on the total duration of time a district spent in lockdown. These estimates are then linked to districts in SEDA2024 by district name and state abbreviation.

*New York Times COVID-19 Data Repository* — Spanning from the beginning of the COVID-19 Pandemic to March 23rd, 2023, the COVID-19 Data Repository consists of county-level data tracking COVID-19 cases and deaths, including 7-day rolling average case and deaths numbers per 100,000 residents. Created by the New York Times, this data repository was created by requests to a variety of government sources at all levels (NYT, 2023). Given endogeneity concerns regarding the ever-evolving nature of the COVID-19 pandemic and it's effect on district learning measures and test scores, we utilize the 7-day rolling averages of COVID-19 cases and deaths as a crucial control variable for our research designs. Taking the 7-day rolling averages for each county, we calculate a pandemic-long average representing the degree to which a certain county was impacted by the COVID-19 outbreak during the pandemic. In doing so, we choose to include rolling averages before the 2020-2021 school year, given that increased COVID-19 cases and deaths pre-school year may impact the initial decision-making of districts to quarantine. These new pandemic-long averages are then linked to districts by county names,

which are derived from the common core data (CCD) of the National Center for Educational Statistics (NCES). Data is then linked to SEDA2024 by state and county name.

*Common Core Data, National Center for Educational Statistics* — A research agency part of the U.S federal government, the NCES collects, analyses, and reports data related to U.S education. For this paper, we exploit the CCD, which provides district-level yearly data on district demographics and characteristics. Specifically, we select for district names and county names, in order to link SEDA2024 with county-level data, along with certain covariates not included in the SEDA2024, such as the average pupil-teacher ratio and institutional expenditures per student.[2] Given the possibility that certain districts may have changed counties over the time period, we take county names for each district between 2017-2024 and coalesce them to link all districts with a singular county. To link the NCES to SEDA2024 by district, we further removed the words "district" from all NCES district observations. Our final CCD was then linked to SEDA 2024 by district and state name.

*MIT Elections Lab Presidential Election Data.* — Given the hot political climate that existed around the COVID-19 pandemic and ensuing policy, it seems plausible that the political leanings of a district and it's elected leaders played a role in whether a district went into remote learning or not, as well as other educational and economic policies that affect test scores. Thus, our final dataset comes from the MIT Election Labs, who complied county-level data on the 2020 U.S presidential election, including numbers of voters who voted for either the Democratic or the Republican Party, collected from official election state election data records. For this paper, we

[2] Data on pupil-teacher ratio and institutional expenditures per student are only available for a limited number of years, so we utilize data observations from the 2021-2022 school year as an estimate for these values.

calculate the share of voters in each county that voted for the democratic party (Joe Biden) in the 2020 presidential election and link this data to SEDA2024 by county name in order to control for political demographics. Given that not all counties in the data set report their total vote count, but instead report their vote count by method of voting (election day, mail-in ballots, early voting, etc.), we sum all vote counts for each district to come up with a total vote count, before taking the number of votes for the democratic party dividing by the derived total vote count to get our percent share of democratic voters in each county.

Having received our final data set, we filter out school districts with less than 1000 total students. Our final merged dataset consists of 195,062 observations from 5,504 U.S school districts in 40 states plus the District of Colombia.

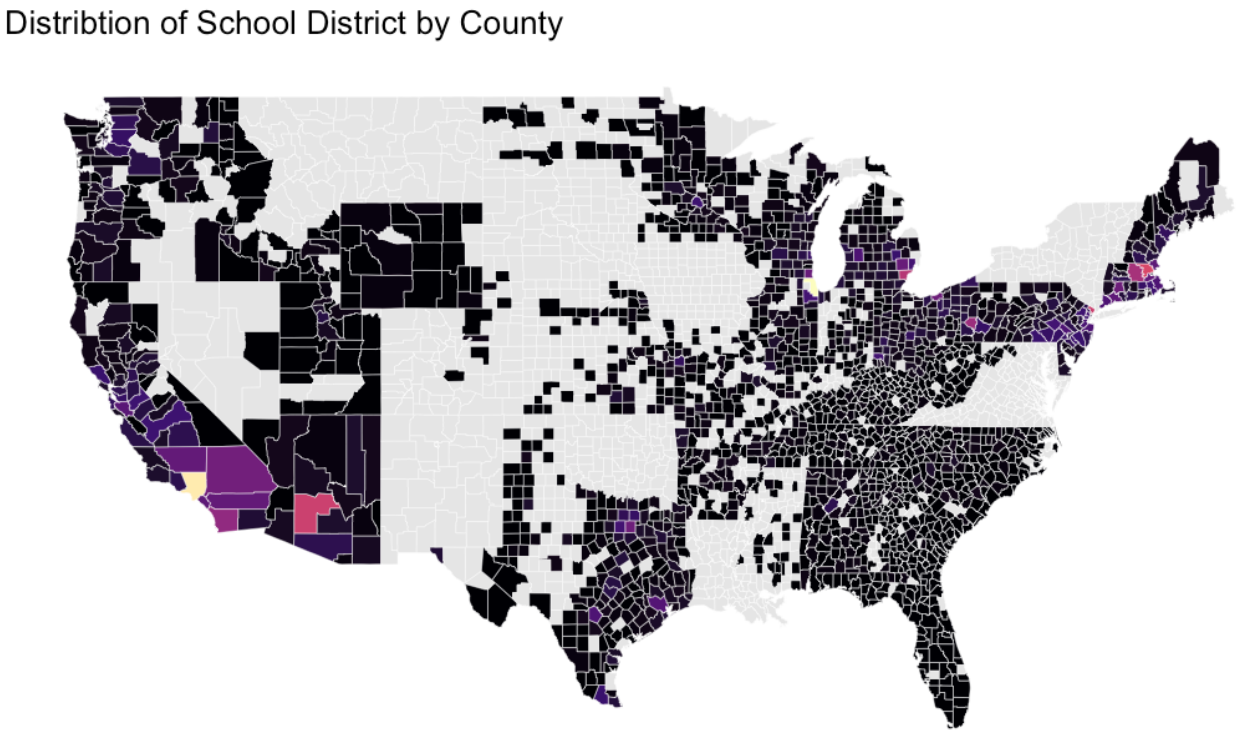


*Figure 1: A map of our sampled U.S school districts and their distribution across U.S counties in the final merged dataset, with brighter colour indicating the location of multiple school districts within a singular county. For our merged dataset, the following states have no observed school district: Alaska, Iowa, Louisiana, Maryland, Montana, New Mexico, New York, Oklahoma, South Dakota, Virginia. Plot does not include districts in Hawaii.*

Four states were lost in merging our SEDA2024 data with the CSHD, as the CSHD only contains data on districts in 48 states. Another 3 states were lost when cleaning SEDA2024, and finally 3 states were lost due to not having a single school district with over 1000 students in the final dataset.

**Methodology**

To test for the marginal effect of duration of remote learning on tests, we employ an OLS model with two-way fixed effects and district controls. Our estimating equation is of the form

$$GYS_{dst} = \beta Months_{ds} + X_{dst} + \delta_d + \mu_t + \epsilon_{dst}$$

Where $GYS_{dst}$ is the grade-year standardized test score for district $d$ in state $s$ in year $t$. $Months_{dst}$ represents the total number of months a district had over 80% of schools in remote learning. $X_{dst}$ is a vector of covariates, which includes percentages of Black and Hispanic students, percentage of students on the free-reduced lunch (FRL) program, COVID-19 average cases and deaths per 100,000 residents, pupil teacher ratio, instructional expenditures per student, democratic vote share, total number of grade 3-8 students, and indicators for whether the district is in an urban area. $\delta_d$ and $\mu_t$ represents the district and year fixed effects in our two-way fixed effects model, and finally $\epsilon_{dst}$ is our error term. Our coefficient of interest is $\beta$, which represents the effect of an additional month in lockdown/remote learning on grade-year standardized test scores. The identification assumption for this model is that the exact number of months a district spends in lockdown is uncorrelated with the error term. This identification assumption will be analyzed using a balance test, comparing the mean characteristics of districts which spent at or over 6 months in lockdown and those that did not.

Next, we employ propensity score matching (PSM), whereby districts are matched for comparison by specific characteristics. Propensity scores were estimated using a logistic

regression model, where the treatment group (i.e, districts that spent any time in lockdown) was regressed on a set of observed socio-economic covariates, including democratic vote share, pupil-teacher ratio, single mother household rate, percent of students on FRL, poverty rate, and an indicator for whether the district is urban. We choose these covariates as they are believed to influence both months a district spent in lockdown, along with test scores We then applied nearest-neighbor matching without replacement, with matching performed using the MatchIt package in R. Using this matched data, we run a DiD regression, taking districts of similar socio-economic characteristics who either went into lockdown or did not, and comparing their test scores before and after the 2020-2021 school year. Our DiD regression is of the form

$$GYS_{dst} = \beta_1 Remote_{ds} + \beta_2 Post_{dst} + \beta_3 Remote_{ds} * Post_{dst} + X_{dst} + \delta_d + \epsilon_{dst}$$

Where $Remote_{ds}$ is an indicator for whether a district d in state s spent any months with over 80% of schools in remote learning, and $Post_{dst}$ is an indicator for whether the observation in test score is after the 2020-2021 school year. Our DiD coefficient is $\beta_3$, which represents the effect of being in lockdown on test scores after 2020-2021 school year. The identification assumption for this model is that trends in test scores between districts of similar socio-economic who either went into lockdown or did not are parallel before the 2020-2021 school year. This will be verified by looking at the pre-trend results of test scores for each group before and after the 2020-2021 school year, as well as perfoming a paired t-test on the matched and unmatched covariates to test whether their difference is statistically significant.

**Results**

In this section, we present the findings from the above dataset and research designs. On table 1, we compare the mean estimates in our cleaned and merged final dataset to the mean estimates in our raw files. As observed from table 1, districts in our raw data on average spent

more months in lockdown, had higher democratic vote shares, and noticeably lower institutional expenditures per student. While these characteristics would initially be indicative of larger districts in more dense urban areas, we find that districts in our merged dataset on average are less likely to be urban, and have higher percentages of white students. Another noticeable difference is that districts in our final dataset have notably lower GYS test scores across all subgroups besides black students.

Table 1: Summary Statistics: Before and After Cleaning By Dataset

| Dataset | Variable | Raw Data | Final Dataset |
|---|---|---|---|
| COVID-19 School Data Hub | Months in lockdown | 1.971 | 1.722 |
| MIT Election Labs | Democratic Vote Share | 0.334 | 0.433 |
| NCES Common Core Data | Pupil-Teacher Ratio | 14.294 | 14.864 |
| | Institutional Expenditure per Student | 9758.805 | 9053.471 |
| NYT COVID-19 Data Repository | Average Cases per 100k | 24.857 | 23.132 |
| | Average Deaths per 100k | 0.431 | 0.398 |
| SEDA2024 | GYS Test Scores, All Students | 5.429 | 5.170 |
| | GYS Test Scores, Black Students | 3.843 | 3.384 |
| | GYS Test Scores, Hispanic Students | 4.447 | 4.193 |
| | GYS Test Scores, White Students | 6.011 | 5.749 |
| | GYS Test Scores, ED Students | 4.464 | 4.186 |
| SEDA2024 Covariates | Total Students | 2782.055 | 3984.561 |
| | Total Students, Grades 3-8 | 1247.250 | 1846.817 |
| | Percentage of Black Students | 0.120 | 0.098 |
| | Percentage of Hispanic Students | 0.191 | 0.189 |
| | Percentage of FRL Students | 0.531 | 0.497 |
| | Percentage of White Students | 0.590 | 0.635 |
| | Urban District | 0.179 | 0.118 |

*Table 1: Summary statistics table, displaying mean estimates for key variables in each used dataset before and after cleaning/merging into final merged dataset. Note that raw data files are observed after basic cleaning, such as renaming of columns and calculations of certain aggregate variables not presented in the raw file (months in lockdown, democratic vote share, cases per 100k, deaths per 100k).*

These differences in test scores are likely the result of removing previous testing years well before COVID-19. Finally, we see a notable increase in total students and grades 3-8 students in our merged dataset. This is likely the result of our filtering out of smaller districts. Overall, changes in mean estimates suggest that our merged dataset may have omitted some poorer, urban districts with more months spent in lockdown. Given the plausibility that these districts may be

more vulnerable and thus more likely to face decrease in test scores from COVID-19 and lockdown measures, the omission of these districts from our final dataset may downwards bias our estimates.

Moving to figures 2 and 3, we plot unweighted and weighted binned scatter plots of GYS district test scores on months spent in remote learning. For both the unweighted and weighted plots, we see a noticeable negative correlation of months spent in remote learning on GYS test scores for a district: that is, schools that spent more time in lockdown tended to have lower GYS test scores. There are two important notes regarding these figures.

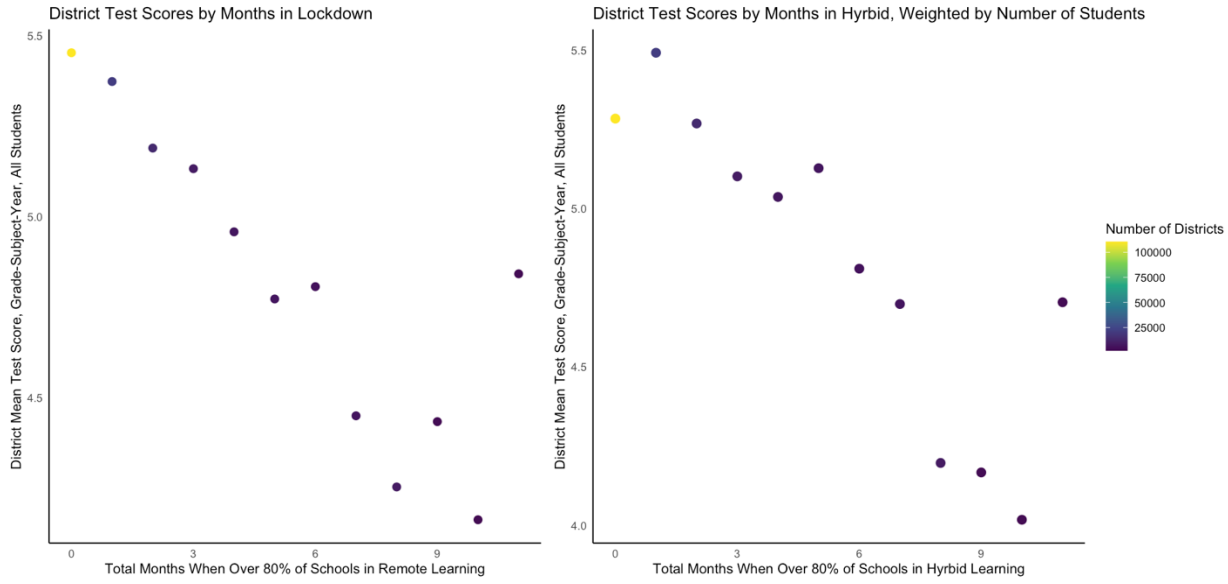


*Figure 2&3: Two binned scatter plots, displaying the unweighted and student-weighted average GYS test scores by number of months spent in remote learning. Colours on scatter points indicate the number of districts within each binned scatter point.*

Firstly, it's clear from the colour density on Figures 2 & 3 that a disproportionate number of the districts spent zero months in lockdown relative to the other bins. Secondly, in both plots it appears that districts that spent the whole year (11 months) in lockdown appear to be an outlier, having significantly higher test scores on average than district that spent 10 or less months in lockdown. One plausible reason for this is that districts that spent the whole of the 2020-2021 school year in lockdown may have more school and parental resources, whereby remote learning

conditions are more comfortable for students and there is less of an incentive to leave lockdown. To test this, we plot the months in remote learning on both the average pupil-teacher ratio and single mother household rate in Figures 4 & 5.

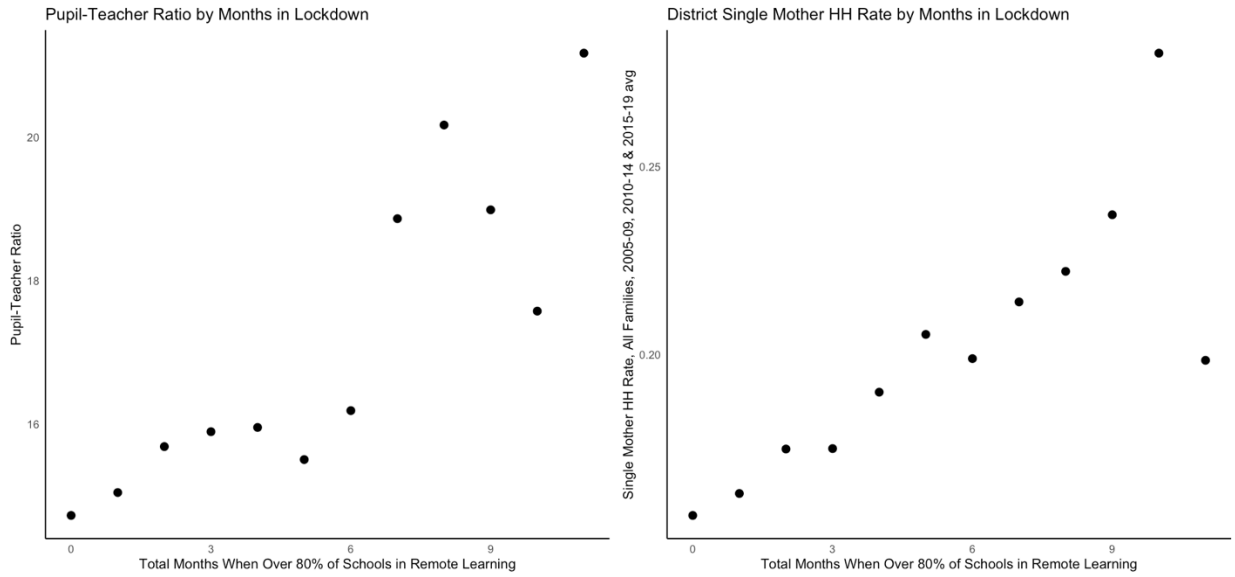


*Figure 4&5: Two binned scatter plots, displaying unweighted average pupil-teacher ratio and single mother household rate by months spent in remote learning. In creating both scatter plots, 38 districts that did not have data on average pupil-teacher ratios, and 50 districts that did not have data on single mother household rates have been omitted.*

Results show that with districts that spent 11 months in lockdown having significantly larger class sizes and less single mother households than districts that spent 10 months in lockdown. This would indicate that districts with high class sizes had less incentive to move back to in-person learning, potentially given the greater probability of contracting COVID-19 or the easier management of large classes on online platforms, while lower single mother household rates

would've left remote learning more comfortable for students at home.

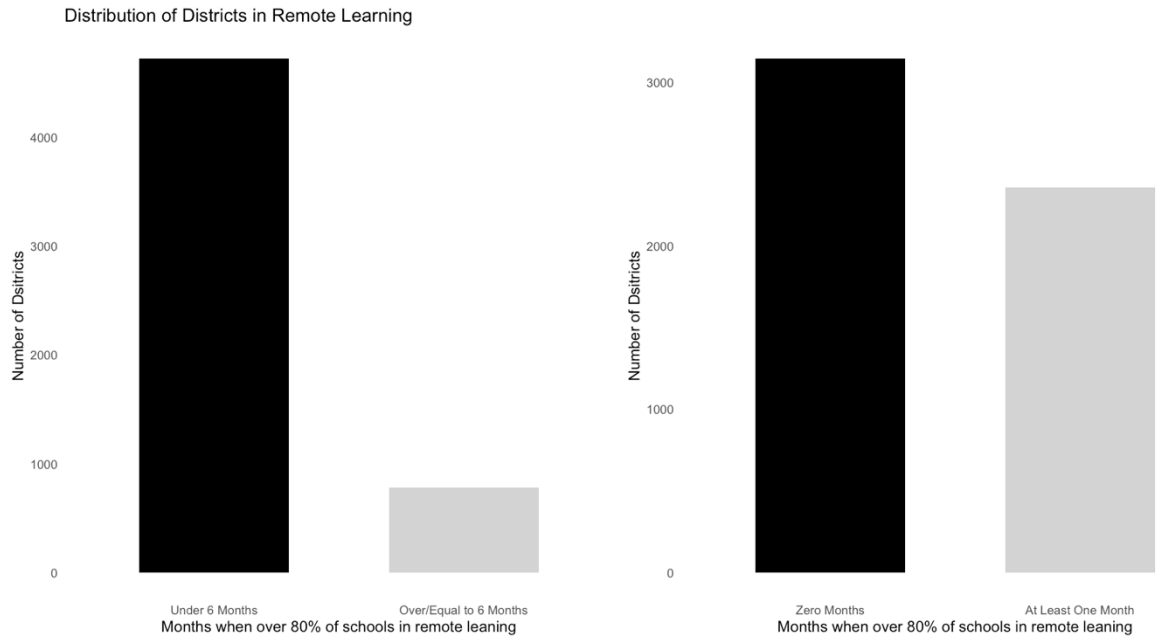


*Figure 6&7: Density distributions of displaying the number of districts in the sample that spent under 6 months in lockdown versus over/equal to 6 months in lockdown, as well as the number of districts that spent at least one months in lockdown versus zero months in lockdown.*

Finally, on Figures 6 & 7 we plot density distributions of districts by months spent in lockdown. While 2358 districts (approx. 44% of sample) spent at least one month in lockdown, only 738 districts (approx. 14% of sample) spent over 6 months in lockdown. Thus, while there existed a significant amount of heterogeneity in the duration of time spent in lockdown for the 2020-2021 school year, the majority of district did not spend any time in lockdown, either hosting in-person classes or hybrid classes. Given the greater balance in sample sizes between districts with no time in lockdown and districts with at least one month in lockdown, we elected to use these as the treatment and control group for our DiD regression design.

Having visualized the makeup of our data, we now turn to our two regression models. We start our analysis first with the OLS model. Table 2 reports result from the OLS regression on grade 3 students.

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | Dependent Variable: GYS Mean Test Score | | | | | |
| | *OLS* | | *panel linear* | *panel linear* | *panel linear* | *panel linear* |
| | No Controls | District Controls | DC + TW-FE | Black | Hispanic | ECD |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Months Spent in Remote Learning | −0.139*** | −0.048*** | −0.044*** | 0.016 | 0.116*** | −0.010 |
| | (0.002) | (0.002) | (0.011) | (0.024) | (0.041) | (0.013) |
| Cases Per 100k | | 0.010*** | 0.026*** | −0.033** | −0.035** | 0.037*** |
| | | (0.001) | (0.005) | (0.014) | (0.016) | (0.005) |
| Deaths Per 100k | | 0.400*** | −0.217** | −0.222 | 0.815** | 0.078 |
| | | (0.033) | (0.110) | (0.212) | (0.357) | (0.126) |
| Urban District | | 0.013 | 0.109 | −0.520*** | −0.343** | −0.009 |
| | | (0.016) | (0.072) | (0.168) | (0.140) | (0.083) |
| Democratic Vote Share | | 0.281*** | 1.089*** | 1.333*** | −1.658*** | 0.775*** |
| | | (0.042) | (0.211) | (0.463) | (0.433) | (0.252) |
| Percent FRL Students | | −3.593*** | 0.281*** | 0.721*** | 0.689*** | 1.105*** |
| | | (0.032) | (0.058) | (0.122) | (0.092) | (0.069) |
| Percent Black Students | | −0.742*** | −3.291*** | −2.411*** | −1.407*** | −3.006*** |
| | | (0.038) | (0.138) | (0.268) | (0.296) | (0.162) |
| Percent Hispanic Students | | −0.172*** | −3.411*** | −4.686*** | −3.933*** | −3.322*** |
| | | (0.028) | (0.108) | (0.254) | (0.167) | (0.130) |
| Constant | 3.268*** | 4.752*** | | | | |
| | (0.007) | (0.048) | | | | |
| Observations | 36,200 | 35,850 | 35,764 | 11,394 | 18,415 | 32,180 |
| $R^2$ | 0.091 | 0.526 | 0.060 | 0.047 | 0.040 | 0.042 |
| Adjusted $R^2$ | 0.091 | 0.526 | −0.108 | −0.165 | −0.169 | −0.143 |

*Table 2: A plot of the results from the OLS two-way fixed effects (TWFE) regressions. Test scores are measured in grade level growth relative to the 2019 national average, and our coefficient of interest "Months in Remote Learning" is measured in months where over 80% of schools in a district were in lockdown. (1) is displays results for the OLS regression with no controls or fixed effects. (2) displays results for the OLS regression with district controls but no fixed effects. Panels (3)-(6) display results for the OLS regression with district control and TWFE on the GYS test scores of all students, black students, Hispanic students, and economically disadvantaged (ECD) students respectively. Below each regression results are the number of observations used in each regression (district, year, and grade), as well the R squared and adjusted R squared values. For statistical significance, * *p<0.1; ∗∗p<0.05; **p<0.01.*

From table 2, we observe a negative and statistically significant effect of months spent in remote learning on GYS test scores among all students. Specifically, we observe that each additional month a district spends in lockdown is associated with 0.044 less of a grade level in test growth, relative to the 2019 national average. Intuitively, this means that a district which tests at the grade-level will score 4.4 percent below grade level for each month spent in lockdown. This is a significant figure, for it predicts that students who went to a school that spent all of the 2020-2021 school year in lockdown and originally tested at grade level will score close to half a grade

behind the 2019 national average. However, this result is for all grade 3 students. If we observe columns (4)-(6), which show the effect on GYS test scores for black students, Hispanic students, and economically disadvantaged students, we observe no statistically significant effect of months in lockdown on test scores. Interestingly, there exists a statistically significant and *positive* effect of months in lockdown on GYS test scores for Hispanic students, suggesting that lockdowns actually *improved* test scores for Hispanic students. One possibility to explain this discrepancy is the aforementioned outlier discussed with figures 4&5. Looking at figure 8, we see a similar high percentage of Hispanic students in districts with 11 months in lockdown that was observed with pupil-teacher ratios and single mother household rate.

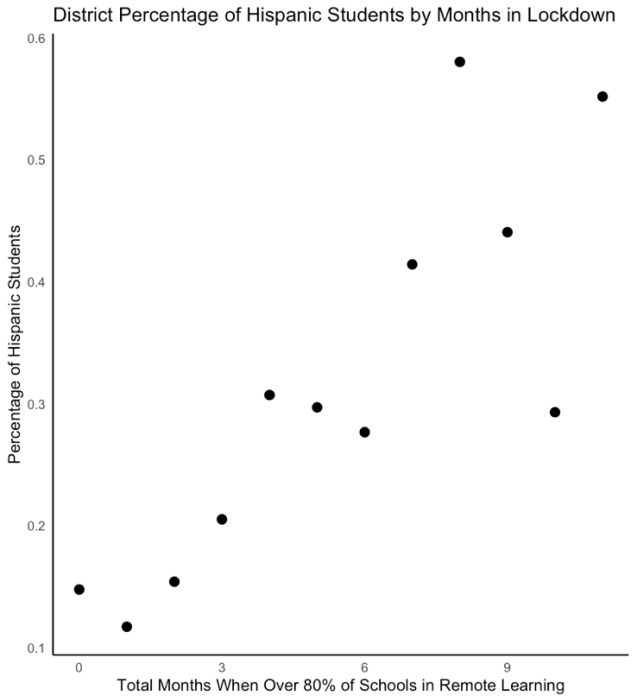


*Figure 8: Binned scatter plot of months spent in remote learning by the percentage of Hispanic students in a district.*

To verify the the validity of our results, we use a balance test comparing the mean covariate estimates of districts who spent under six months in lockdown, and those that spent six months and over. Observing table 3, we note that our data is highly unbalanced. Specifically, districts with that spent at or over 6 months in lockdown are more likely to have higher pupil-

teacher ratios, higher percentages of minority students, and more students overall, and score worse in various socio-economic indicators. Given that all these same covariates are presumed to be Negatively correlated with test scores, and are correlated with high duration in lockdown,

| | Under 6 Months in Lockdown | Over/Equal to 6 Months in Lockdown | p-test | SMD |
|---|---|---|---|---|
| Observations | 31438 | 4762 | | |
| GYS Test Score, All Students (mean (SD)) | 3.16 (1.19) | 2.16 (1.38) | <0.001 | 0.776 |
| Pupil-Teacher Ratio (mean (SD)) | 15.01 (3.31) | 18.71 (4.33) | <0.001 | 0.961 |
| Cases Per 100k (mean (SD)) | 23.62 (5.59) | 20.84 (5.69) | <0.001 | 0.494 |
| Deaths Per 100k (mean (SD)) | 0.41 (0.17) | 0.35 (0.17) | <0.001 | 0.336 |
| Institutional Expenditure Per Student (mean (SD)) | 8764.29 (2712.40) | 9914.41 (6337.31) | <0.001 | 0.236 |
| Percent Black (mean (SD)) | 0.09 (0.15) | 0.16 (0.25) | <0.001 | 0.326 |
| Percent Hispanic (mean (SD)) | 0.17 (0.21) | 0.44 (0.30) | <0.001 | 1.037 |
| Total Students, Grades 3-8 (mean (SD)) | 2613.35 (6084.93) | 4884.02 (10503.17) | <0.001 | 0.265 |
| Percent FRL (mean (SD)) | 0.47 (0.22) | 0.62 (0.24) | <0.001 | 0.660 |
| Percent Asian (mean (SD)) | 0.03 (0.06) | 0.08 (0.13) | <0.001 | 0.524 |
| Percent Native (mean (SD)) | 0.01 (0.03) | 0.02 (0.10) | <0.001 | 0.149 |
| Percent Other Ethnicity (mean (SD)) | 0.04 (0.03) | 0.04 (0.04) | <0.001 | 0.168 |
| Percent White (mean (SD)) | 0.67 (0.26) | 0.26 (0.24) | <0.001 | 1.640 |
| Total Students (mean (SD)) | 5729.37 (13474.59) | 10234.01 (22940.60) | <0.001 | 0.239 |
| Socio-Economic Status Composite (mean (SD)) | 0.27 (0.92) | -0.11 (1.15) | <0.001 | 0.369 |
| Log Income (mean (SD)) | 10.90 (0.34) | 10.92 (0.38) | 0.001 | 0.050 |
| Bachelor's Degree + Rate (mean (SD)) | 0.26 (0.15) | 0.26 (0.16) | 0.876 | 0.002 |
| Unemployment Rate (mean (SD)) | 0.07 (0.02) | 0.09 (0.03) | <0.001 | 0.789 |
| Snap Recipiant Rate (mean (SD)) | 0.11 (0.07) | 0.12 (0.08) | <0.001 | 0.230 |
| Poverty Rate (mean (SD)) | 0.13 (0.07) | 0.16 (0.09) | <0.001 | 0.435 |
| Single Mother HH Rate (mean (SD)) | 0.16 (0.06) | 0.22 (0.09) | <0.001 | 0.728 |
| Urban (mean (SD)) | 0.09 (0.29) | 0.28 (0.45) | <0.001 | 0.503 |
| Suburb (mean (SD)) | 0.35 (0.48) | 0.50 (0.50) | <0.001 | 0.296 |
| Town (mean (SD)) | 0.27 (0.44) | 0.14 (0.35) | <0.001 | 0.321 |
| Rural (mean (SD)) | 0.29 (0.45) | 0.08 (0.27) | <0.001 | 0.555 |

*Table 3: Balance table results, comparing covariates mean estimates for districts under 6 months in lockdown and districts at over 6 months in lockdown. Columns 1-3 present the various covariates found the merged dataset, as well as their mean estimates and standard deviations for both groups of the balance test. Column 4 presents p-values for a statistically significant difference between the two balance test group, and column 5 presents standard mean differences (SMD), which measure the difference in means between two groups relative to the variability of the data.*

it would appear that our analysis has some downwards bias, downplaying the marginal effect that high duration in lockdown had on test scores.

To combat endogeneity in our OLS, we turn to our second research using a DiD regression with propensity score matching. Figure 9 presents the covariates used in our propensity score matching, before and after the matching. Using a paired t-test for covariates before and after matching, we derive a p-value of 0.0133, indicating the difference in SMD's from our unmatched and matched covariates is statistically significant from zero and thus is more balanced.

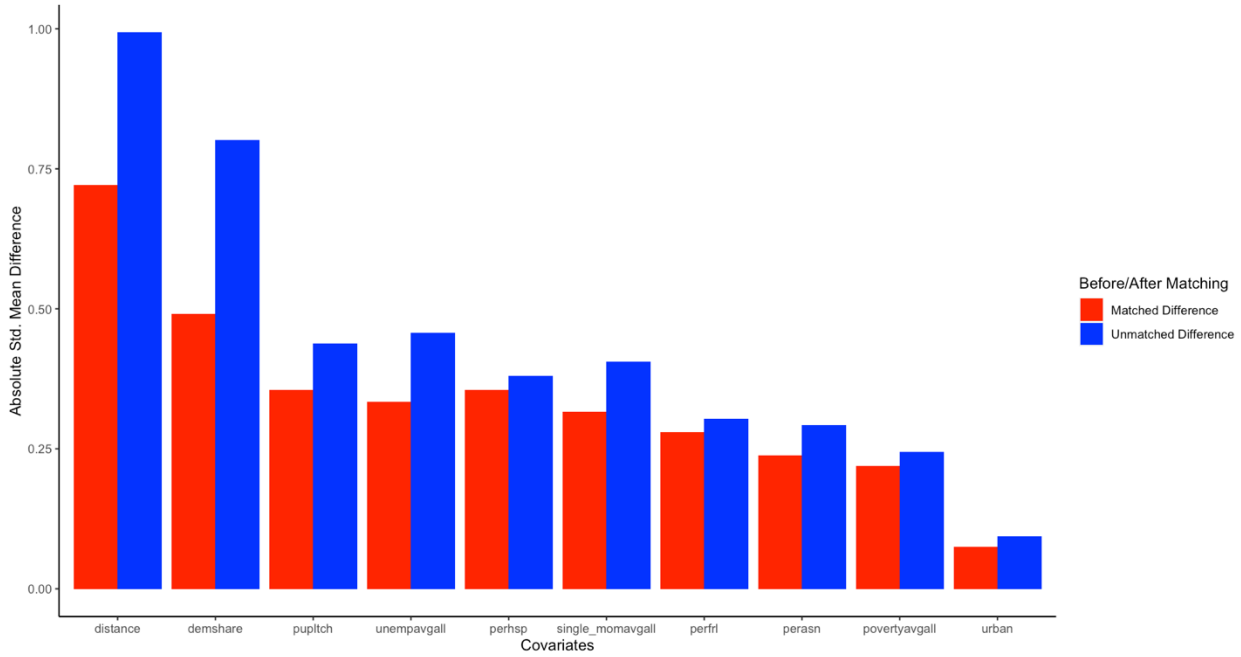


*Figure 9: Comparison of SMD's for covariates used in propensity score matching before and after matching.*

Having matched districted by covariates, we next present the trends of the treatment and control group over time. Looking at figure 10, we see that pre-lockdown trends in test scores

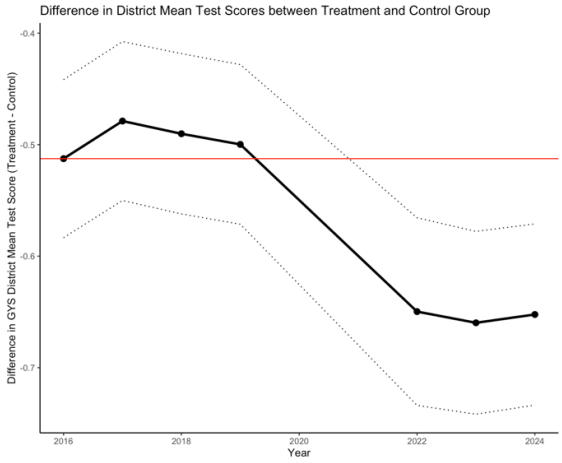


*Figure 10: Binned scatter plot of mean differences between the treatment group (districts that spent at least one month in lockdown) and the control group (districts that spent zero months in lockdown) on GYS test scores. The dotted lines display 95% confidence intervals for results, while the red line indicates the initial difference in GYS test score. Jump in data points indicates gap in data for the 2020 and 2021 years.*

experienced little change, before significantly dipping in our first observed year after the 2020-2021 school year. This initial stability in differences between 2016-2019 indicates that the trends in test score between our treated and untreated districts were parallel before COVID-19, and is

given as evidence for our parallel trends assumption in using this research design. On table 4, we present results for our DiD regressions. Observing the DiD estimate on table 4,

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Dependent Variable: GYS Mean Test Score | | | | |
| | *OLS* | | *panel linear* | *panel linear* | *panel linear* |
| | No Controls | District Controls | DC + District FE | Black | Hispanic |
| | (1) | (2) | (3) | (4) | (5) |
| Remote | −0.495*** | −0.156*** | −0.097*** | −0.574*** | −0.506*** |
| | (0.019) | (0.014) | (0.036) | (0.108) | (0.088) |
| Post | −0.052** | −0.026* | −0.005 | −0.119*** | −0.096*** |
| | (0.021) | (0.014) | (0.009) | (0.020) | (0.016) |
| Cases Per 100k | | 0.015*** | 0.014*** | 0.015 | 0.044*** |
| | | (0.001) | (0.004) | (0.013) | (0.012) |
| Deaths Per 100k | | 0.324*** | −0.200* | −1.059*** | −0.562 |
| | | (0.036) | (0.113) | (0.235) | (0.361) |
| Urban | | 0.007 | 0.119* | −0.835*** | −0.380*** |
| | | (0.016) | (0.068) | (0.129) | (0.106) |
| Democratic Vote Share | | 0.359*** | 0.317 | 1.076** | 0.025 |
| | | (0.048) | (0.205) | (0.477) | (0.404) |
| Percent FRL | | −3.647*** | 0.204*** | 0.822*** | 0.458*** |
| | | (0.035) | (0.062) | (0.123) | (0.096) |
| Percent Black | | −0.808*** | −2.178*** | −0.679*** | 0.032 |
| | | (0.041) | (0.116) | (0.224) | (0.247) |
| Percent Hispanic | | −0.314*** | −2.223*** | −1.402*** | −2.126*** |
| | | (0.031) | (0.120) | (0.314) | (0.210) |
| DiD Estimate ($\beta_3$) | −0.160*** | −0.160*** | −0.175*** | −0.290*** | −0.158*** |
| | (0.030) | (0.020) | (0.012) | (0.024) | (0.019) |
| Constant | 3.273*** | 4.832*** | | | |
| | (0.014) | (0.055) | | | |
| Observations | 29,740 | 29,740 | 29,686 | 10,578 | 16,215 |
| $R^2$ | 0.050 | 0.553 | 0.078 | 0.111 | 0.077 |
| Adjusted $R^2$ | 0.050 | 0.553 | −0.108 | −0.090 | −0.136 |

*Table 4: Results from differences-in-differences (DiD) regression. Test scores are measured in grade level growth relative to the 2019 national average, and our coefficient of interest "DiD Estimate" is an indicator for districts who had 80% of schools go into lockdown for some duration, and their test score estimates post 2020-2021. (1)-(3) present results for are for all students, without controls, with district controls, and with district controls plus district fixed effects. Columns (4)-(5) present results with districts controls and fixed effects on Black and Hispanic GYS test scores. Results on ECD students has been omitted from graph due to compiling errors but can be viewed in the replication package located in the appendix.*

we observe negative and statistically significant effects of lockdown measures on test scores for all students, as well as our chosen subgroups of students. Overall, districts that had over 80% of schools in lockdown for any period of time during the 2020-2021 school year had 0.160 grade levels worse testing scores post 2021 than similar districts that did not undergo lockdown.

Looking to the subgroups, we note a particularly large and statistically significant effect on black test scores, with black students that spent time in lockdown having 0.290 grade-level lower test scores, post 2021, compared to black students in similar districts that did not spend time in lockdown. Given existing inequalities in test scores between black and non-minority students (Fryer Jr. & Levitt, 2004), and the higher proportion of black students in lockdown schools relative to non-lockdown schools (as shown in table 3), this would indicate that lockdown measures may have worsened existing inequalities in education between black and non-minority students.

To explain the statistical significance of our two regressions, we propose the following mechanism to explain the reduction in test scores for districts in lockdown. Firstly, there is single motherhood. As was seen in figure 5, a districts time spent in lockdown was positively correlated with a higher single motherhood rate; besides for 11 months which we've already discussed. Since single motherhood is negatively correlated with test scores, given the lack of an additional parent to provide adequate resources and help with schoolwork, the further isolation of a student from outside class resources plausibly had a negative effect on learning outcomes over this time. Another plausible mechanism are Free-reduced lunch programs (FRL). Numerous studies (Cohen et al., Kleinmann et al.) have found that nutrition and access to school lunches is positively associated with test scores and other learning outcomes. During the COVID-19 pandemic, many districts saw increased challenges of sustaining breakfast programs and free-reduced lunch programs for kids while schools were in lockdown (Jabbari et al.), resulting in numerous districts losing access to school meal programs (Plank et al.). Thus, it's plausible that a lack of access to FRL programs reduced learning outcomes of students in lockdown districts,

with figures 11 & 12 highlighting the negative associations of GYS test scores.

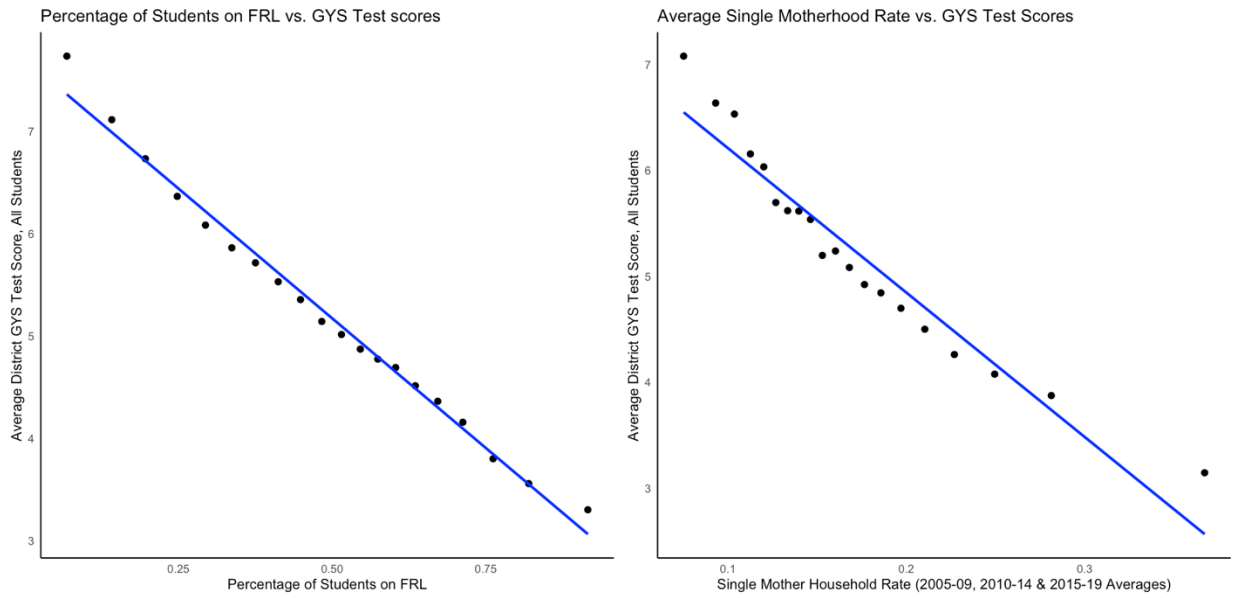


*Figure 11&12: Binned scatterplots of FRL percentage and single mother household rates on GYS test scores.*

**Discussion**:

In this paper, we ran two regression models to highlight the effect of school lockdown measures on test scores. In our OLS regression, we found statistically significant and negative effects of lockdown duration on test scores for all students, but found limited results (and even statistically significant positive results) for test scores on subgroups such as Black students, Hispanic students, and economically disadvantaged students. For our differences-in-differences regression, we found that districts that spent any duration of time with over 80% of schools in lockdown had lower test scores post 2020-2021 than district of comparable socio-economic characteristics that did not spend time in lockdown. While there remains some conflicting results between subgroups in both regressions, the general negative statistically significant effects found in both regression models suggests that lockdown measures did impact education and learning outcomes for all students. Furthermore, the larger effects found in DiD on black test scores

provides evidence for the hypothesis that lockdown orders exacerbated existing inequalities in education and learning outcomes between black and non-minority students.

As noted in the results, each model suffers from certain limitations. Firstly, in the OLS regression we found significant differences in the characteristics of districts that spent long durations of time in lockdown versus those that did not. While we partially mediated this through propensity score matching in our DiD regression, there remains issues of balancing results given the correlation of lockdowns and lockdown duration to numerous district covariates. However, as also noted, many of the district covariates positively correlated to lockdown duration, such as pupil-teacher ratios and the district poverty rate, are presumed to be negatively correlated with test scores. This leads us to believe that any biases resulting from our analysis are primarily downwards, and would thus dampen the observed effects of lockdown duration seen in this paper rather than exaggerate them.

Another endogeneity issue involves the prevalence of missing data figures. Not only were numerous U.S schools districts lost during the cleaning/merging of datasets, but many districts don't report GYS test score data for subgroups. As a result, it's observed on tables 3 & 4 that our regression results on Black students, Hispanic students, and economically disadvantaged students use only a fraction of the samples that are used in our regressions on all students. Assuming the missing of data points is potentially correlated with district characteristics, such as district funding, then the issue of sample selection bias remains a pressing issue in any analysis of Black, Hispanic, or any other subgroup of students.

**Exercise**:

In an ideal scenario, we would run a randomized control trial (RCT), independent of COVID-19 or other socio-economic factors, to determine the exact effect of how lockdowns or remote

learning can effect test scores and education. In this RCT, students would be randomly assigned to go to a certain school for a year, of which students/parents have no information on and cannot transfer or change from. These assigned schools would then be randomly assigned a lockdown time, with certain schools spending no time in lockdown/remote learning, other schools spending some time in lockdown, and then some schools spending the whole year in lockdown. At the beginning and end of this hypothetical school year, students from all schools would be tested on a singular standardized test, relative to their grade level. We could then compare the changes in test scores for schools that spent no duration in lockdown versus those that did to derive a causal estimate. Our final dataset would thus consist of cross-sectional data that contains the duration of time spent in lockdown (either in months or weeks), the average test scores of students before the beginning of this hypothetical school year experiment, and the test scores at the end of the school year.[3] Given the random assignment of students and lockdown measures to schools, our final dataset should contain no possible covariates that are correlated with both duration in lockdown and test scores. Our regression model would thus be of the form.

$$TS_s = \beta_1 Lockdown_s + \epsilon_s$$

Where $TS_s$ represents change in test score for a particular school s, and $Lockdown_s$ measures the duration of time that district spent in lockdown. The parameter of interest is $\beta_1$, which represents the marginal effect of an additional month/week spent in lockdown on the changes in test scores. Our identification assumption is that the randomness of school assignment and lockdown measures implies that no other variable is correlated with lockdown duration and test

[3] If we wanted to study Long-term effects of lockdowns on test scores, we ideally could also utilize panel data that studies test scores for individual students at these hypothetical schools over multiple years.

scores, implying that there is no need for any further controls or other elements to our estimating equation.

**Conclusion:**

In future, one of the main improvements that could be made to this paper would be getting more detailed and holistic data on school lockdown measures. For this paper, we only were able to study the effect of lockdown measures taken during one school year across a reduced sample of school districts. Other data sources, such as those complied by Burbio (2025) and used in provide more time-accurate and expanded data on school learning models during both the 2020-2021, and 2021-2022 school years; though due to limited funding these sources could not be exploited. Furthermore, different methods of balancing data, such as synthetic control method, might have proven to be better at balancing our results and reducing the differences in mean estimates between districts of different lockdown duration. Finally, another future direction of this research could be to study lockdown measures on post-secondary education and outcomes. Given the necessity of high-quality post-secondary education to improve human capital, it would be interesting to extend this analysis and see whether university aged adults experienced similar reductions in learning outcomes during the COVID-19 Pandemic.

**Appendix:**

All data sources, results, and figures can be found in the replication package attached to this paper.

# Bibliography

Burbio (2025). Burbio's K-12 School Opening Tracker**.** Retrived From https://about.burbio.com/school-opening-tracker.

Cohen, J. F. W. & Hecht, A. A. & McLoughlin, G. M. & Turner, L. & Schwartz, M. B. (2021) Universal School Meals and Associations with Student Participation, Attendance, Academic Performance, Diet Quality, Food Security, and Body Mass Index: A Systematic Review. *Nutrients*,13(3):911.

CSDH. (2023). *District-Monthly Percentage In-Person, Hybrid, or Virtual.* Retrieved from COVID-19 School Data Hub: https://www.covidschooldatahub.com/data-resources

Cingel, A. R. & Drew, P. (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. *17*(10).

Fryer Jr., R. G. & Levitt, S. D. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School**.** *The Review of Economics and Statistics,* 86(2): 447–464.

Heise, M. (2024). Educational Rights and Wrongs: How COVID-Related Student Math Achievement Losses Distributed and Implications for Equal Educational Opportunity. *Journal of Education Finance, 49*(3), 291-323.

Jabbari, J. & Chun, Y. & Nandan, P. & McDermott, L. & Frank, T. & Moreland-Russell, S. & Ferris, D. & Roll, S. (2021). How Did School Meal Access Change during the COVID-19 Pandemic? A Two-Step Floating Catchment Area Analysis of a Large Metropolitan Area. *Int J Environ Res Public Health*, 18(21).

Kleinman, R.E. & Hall, S. & Green, H. & Korzec-Ramirez, D. & Patton, K. & Pagano M.E. & Murphy, J. M. (2002). Diet, breakfast, and academic performance in children. *Ann Nutr Metab,* 46 Suppl 1(0 1):24-30.

MIT Election Data and Science Lab. (2018). County Presidential Election Returns 2000-2020. Harvard Dataverse. Retrieved from https://doi.org/doi:10.7910/DVN/VOQCHQ.

National Centre for Education Statistics. (2024). *Common Core Data* . Retrieved from https://nces.ed.gov/ccd/elsi/tableGenerator.aspx

Plank, K. & Hewawitharana, S. & Talmage, E. & Rauzon, S. & Woodward-Lopez, G. (2022). School meal access and changes in meal participation during COVID-19 school closures: A stratified random sample of CalFresh Healthy Living Eligible school districts in California. *Preventive Medicine Reports*, 28:101794.

Reardon, S. F. & Fahle, E. M. & Shear, B. R. & Ho, A. D. & Saliba, J. & Min, J. & Shim, J. & Kalogrides, D. (2025). *Stanford Education Data Archive Technical Documentation (Version SEDA 2024).* Retrieved from https://purl.stanford.edu/pt329xg7054.

Rebecca Reynolds, J. A. (2022). Digital divide, critical-, and crisis-informatics perspectives on K-12 emergency remote teaching during the pandemic. *Journal of the Association for Information Science and Technology*(73), 1665–1680.

The New York Times. (2023). *COVID-19 Data Repository.* Retrieved from https://github.com/nytimes/covid-19-data

Thomas Dee, E. H. (2023). The Revealed Preferences for School Reopening: Evidence from Public-School Disenrollment. *American Educational Research Journal, 60*(5), 916-940.