Data and Code for Pandemic Schooling Mode and Student Test Scores: Evidence from U.S. School Districts

Rebecca Jack, Clare Halloran, James Okun, Emily Oster

Overview

This document describes the data and code contained in the replication archive used to produce the results in *Pandemic Schooling Mode and Student Test Scores: Evidence from U.S. School Districts*. The data used in this paper come from seven sources: (1) state test score data from 11 states, (2) schooling mode (learning model) data from the COVID-19 School Data Hub, (3) average COVID-19 case counts from USA Facts (4) National Center for Education Statistics Common Core of Data Demographic Data, (5) commuting zone labor market data, (6) Bureau of Labor Statistics unemployment data, and (7) Republican vote share in the 2020 election. Data were cleaned, combined, and analyzed in Stata 16 using the code provided and described below. Data variable labels are included in the final data file (state_score_data.dta) in place of a data dictionary.

Data Availability and Provenance Statements

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

Statement about Rights

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

☒ All data **are** publicly available.
☐ Some data **cannot be made** publicly available.
☐ **No data can be made** publicly available.

Details on each Data Source

State Test Score and Participation Data

State test score data are contained in the "Data/Raw/statescoredata" folder in individual state subfolders. Data on participation rates are either included in the test score data files or as a separate participation file in the state subfolder.

Data are publicly available and accessed from individual state websites as listed below:

- Colorado (CMAS) https://www.cde.state.co.us/assessment/cmas-dataandresults

  Citation: Colorado Department of Education (CDE). 2016–2021. "CMAS - Mathematics, English Language Arts, Science and Social Studies Data and Results". Accessed September 13, 2021, https://www.cde.state.co.us/assessment/cmas-dataandresults.

- Connecticut (SBAC)https://portal.ct.gov/SDE/Student-Assessment/Main-Assessment/Statewide-Summative-Assessment-Results-2020-21 and http://data.ctdata.org/data_by_topic

  Citation: Connecticut State Department of Education (CSDE). 2016–2021. "Smarter Balanced by All Students". Accessed October 31, 2021, http://data.ctdata.org/visualization/smarter-balanced-by- allstudents.

- Massachusetts (MCAS) https://www.doe.mass.edu/mcas/results.html

  Citation: Massachusetts Department of Elementary and Secondary Education (DESE). 2017–2021. "Massachusetts Comprehensive Assessment System Results". Accessed December 4, 2021, https://www.doe.mass.edu/mcas/results.html.

- Minnesota (MCA) https://public.education.mn.gov/MDEAnalytics/DataTopic.jsp?TOPICID=1

  Citation: Minnesota Department of Education (MDE). 2016–2021. "Data Reports Analytics: MCA Assessment Files". Accessed October 12, 2021, https://public.education.mn.gov/MDEAnalytics/DataTopic.jsp?TOPICID=1.

- Mississippi (MAAP) https://www.mdek12.org/OPR/Reporting/Assessment

  Citation: Mississippi Department of Education. 2016–2021. "Assessment Files". Accessed November 4, 2021, https://www.mdek12.org/OPR/Reporting/Assessment.

- Ohio (OAA) https://reportcard.education.ohio.gov/download

  Citation: Ohio Department of Education (ODE). 2016–2021. "Ohio School Report Cards: Download Data". Accessed October 28, 2021, https://reportcard.education.ohio.gov/download.

- Rhode Island (RICAS) https://www3.ride.ri.gov/ADP

  Citation: Rhode Island Department of Education (RIDE). 2018–2021. "Assessment Data Portal: RICAS". Accessed October 28, 2021, https://www3.ride.ri.gov/ADP.

- Virginia (SOL) https://www.doe.virginia.gov/statistics_reports/sol-pass-rates/index.shtml

Citation: Virginia Department of Education (VDOE). 2016–2021. "SOL Test Results". Accessed September 18, 2021, https://www.doe.virginia.gov/statistics reports/sol-pass-rates/index.shtml.

- West Virginia (WVGSA) https://zoomwv.k12.wv.us/Dashboard/dashboa rd/7301

  Citation: West Virginia Department of Public Instruction (WVDE). 2016–2021. "Assessment Achievement". Accessed October 28, 2021, https://zoomwv.k12.wv.us/Dashboard/dashboard/7301.

- Wisconsin (Forward) https://dpi.wi.gov/wisedash/download-files

  Citation: Wisconsin Department of Public Instruction (DPI). 2016–2021. "WISEdash Data Files by Topic: Forward". Accessed September 10, 2021, https://dpi.wi.gov/wisedash/download-files/type?field wisedash upload type value=Forward.22

- Wyoming (WY-TOPP) https://portals.edu.wyoming.gov/Reports/(S(wh lv4xvhuj5tmeiutlv2z1tn))/Public/wde-reports-2012/public-reports/asse ssment/pawsresultsdistrict-levelaggregated

  Citation: Wyoming Department of Public Instruction (WDE). 2018–2021. "WY-TOPP and WY-ALT Results grades 3 - 10 - District Level - Aggregated". Accessed October 12, 2021, https://portals.edu.wyoming.gov/Reports/(S(upztr0lvig411by0mehp0o4n))/Public/wde- reports- 2012/publicreports/assessment/ pawsresultsdistrict-levelaggregated.

COVID-19 School Data Hub Learning Model Data

Schooling mode (also referred to as learning model) data files are accessed through the COVID-19 School Data Hub (CSDH) and are available here: ht tps://www.covidschooldatahub.com/. Individual raw data files in both CSV and XLSX formats can be found on the CSDH website and are included with the replication archive. Data details describing the construction and categorization of schooling modes in each state file can be found on the CSDH website. The raw data files for the 11 states included in our analysis are appended together using code_to_combine_learning_model.do and the resulting data file (schooling_mode_data.dta) is included in the Data folder of the replication archive. The raw data files are in Data/Raw/learningmodel.

Citation: COVID-19 School Data Hub [CSDH]. 2022. School Learning Model Database. Accessed 1 April 2022. https://www.covidschoold atahub.com/for researchers.

COVID-19 Community Case Rate Data

We included COVID-19 Community Case Rate Data by District from the COVID-19 School Data Hub website accessed through the "Download Community Case Rate Data" link (https://downloads.ctfassets.net/9fbw4onh0qc1/1FyYF7Qqm

n2fXfWYqcqZUB/d2f9ec9d4a78bdedbc93869396393c09/Matched_Districts_and_Case_Rates.zip) in the *Community Case Rate Data by District* box on the *For Researchers* page of the CSDH website (https://www.covidschooldatahub.com/for_researchers). These community case rates are drawn from USA Facts (https://usafacts.org/visualizations/coronavirus-covid-19-spread-map) and are matched with school districts based on zip code. Data documentation can be found at the website above. These are included in the replication kit in the file "covid_case_rates.csv".

Citation: COVID-19 School Data Hub [CSDH]. 2021. Community Case Rate Data by District. Accessed 1 October 2021. https://www.covidschooldatahub.com/for researchers.

NCES Demographic Data

Both district- and school-level demographic data from the National Center for Education Statistics Common Core of Data (https://nces.ed.gov/ccd/) were accessed through Urban Institute's Education Data Portal using Stata 16 (https://educationdata.urban.org/documentation/). Data documentation can be found at the website above. We use three different NCES files with data from 2015-2020 in our analysis which are included in the replication archive:

- School-level CCD directory data: Data/Raw/nces_school_2015_2020.dta
- District-level CCD directory data: Data/Raw/nces_district_2015_2020.dta
- District-grade-level CCD enrollment data: Data/Raw/nces_district_grade_2015_2020.dta

Citation: U.S. Department of Education (ED). 2022. "Common Core of Data [dataset]". Accessed via Urban Institute Education Data Portal (Version 0.15.0) on April 4, 2022, https://educationdata. urban.org/documentation/.

Commuting Zone Data

We accessed 2000 commuting zone data (commuting_zones.xls) from the U.S. Department of Agriculture's Economic Reporting Service (https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/). Data documentation can be found at the website above. These data are included in the replication kit in the file "commuting_zones.xls".

Citation: U.S. Department of Agriculture (USDA), Economic Research Service. 2012. 2000 Commuting Zone Data. Washington, DC: U.S. Department of Agriculture. https://www.ers.usda.gov/dataproducts/ commuting-zones-and-labor-market-areas/.

U.S. Bureau of Labor Statistics Unemployment Data

We accessed monthly, county-level unemployment data (county_unemployment.txt) through the U.S. Bureau of Labor Statistics for 2016-2021 (see "la.data.64.County"

at https://download.bls.gov/pub/time.series/la/). Data documentation can be found at the website above. These data are included in the replication kit in the file "county_unemployment.txt".

Citation: U.S. Bureau of Labor Statistics. 2021. County-level Unemployment Data (la.data.64.County).Washington, DC: U.S. Bureau of Labor Statistics. https://download.bls.gov/pub/time.series/la/.

Republican Vote Share by County in the 2020 Election

County-level election results from 2020 (Vote_share_2020_data.csv) were accessed through the "United States General Election Presidential Results by County from 2008 to 2020" Github Repository (https://github.com/tonmcg/US_County_Level_Election_Results_08-20#readme). Data documentation can be found at the website above. Data for 2020 were collected from Fox News, Politico, and the New York Times. These data are included in the replication kit in the file "Vote_share_2020.csv".

Citation: McGovern, Tony. 2021. United States General Election Presidential Results by County from 2008 to 2020. Github Repository. https://github.com/tonmcg/US County Level Election Results 08- 20.

Clean Data File

The final data file used for all analyses is also included in the replication archive in the "Data/Clean" folder (state_score_data.dta). All variables are labeled.

Dataset list

See "Dataset_list.xlsx" included in the replication kit.

Computational requirements

Software Requirements

- Stata (code was last run with version 16)
  - `labmask` (as of April 21, 2022)
  - `tabstatmat` (as of April 21, 2022)
  - `estout` (as of April 21, 2022)
  - `coefplot` (as of April 21, 2022)
  - the program "`package_install_check`" in "`analysis.do`"will install all dependencies locally, and should be run once.

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:

☐ <10 minutes

☐ 10-60 minutes
☒ 1-8 hours
☐ 8-24 hours
☐ 1-3 days
☐ 3-14 days
☐ > 14 days
☐ Not feasible to run on a desktop machine, as described below.

Details The code was last run on an **8-core Apple M1-based laptop with MacOS version 12.0.1**.

Description of programs/code

The “Code" folder contains three DO files to reproduce all figures and analysis in the paper. Both clean.do and analysis.do should be run from the "Code" folder as the working directory.

1. master.do: runs all of the coded needed to replicate the results for this paper.
2. code_to_combine_learning_model.do: This file is used to append raw state learning model data files from the COVID-19 School Data Hub into one data file to be used in the analysis. The resulting output (schooling_mode_data.dta) of this file is also included in the replication archive.
3. clean.do: This file cleans and combines all raw data files listed above into the final data file for analysis (state_score_data.dta).
4. analysis.do: This file uses the final data file (state_score_data.dta) and produces Figure 1 and Tables 1–4 in the paper, as well as the three additional figures in Appendix Figure A1 and an additional robustness check that excludes Connecticut from the regressions due to a significant difference in students who took the test remotely.[1]

Instructions to Replicators

- Run master.do which runs all code to reproduce results.

List of tables and programs

The provided code reproduces:

☒ All numbers provided in text in the paper
☒ All tables and figures in the paper

[1] Connecticut State Department of Education. 2021. *Statewide Summative Assessment Report 2020-21.* Hartford, Connecticut: CSDE. Available at: https://portal.ct.gov/-/media/SDE/Student-Assessment/Main-Assessment/Connecticut_Assessment_Summary_Spring2021.pdf

☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table #	Program	Line Number	Output file	Note
Table 1	Code/analysis.do		summary_stats_panela.tex	
Table 2	Code/analysis.do		demographics.csv	
Table 3	Code/analysis.do		main_regressions.tex	
Table 4	Code/analysis.do		robustness.tex	
Figure 1	Code/analysis.do		pass_rate_comparisons.pdf	
Figure A1	Code/analysis.do		box_hybrid.pdf, box_inperson.pdf, box_virtual.pdf	

References

References are included in the paper and are included above for each data source.