COVID-19 School Data Hub: Introduction and Data Availability Overview

COVID-19 School Data Hub Team

Introduction

In March of 2020, virtually all schools in the United States closed for in-person learning due to the COVID-19 pandemic. The subsequent 2020-21 school year was characterized by inconsistent student access to in-person learning. Some schools and districts opened consistently for in-person learning, while some transitioned to a hybrid model and others remained predominantly virtual for much of the school year. Still others transitioned between learning models depending on a variety of factors, including local case rates and holidays.

Early research has begun to explore the ways in which the COVID-19 pandemic has affected the U.S. school system and its students. For example, some research has suggested that factors related to the pandemic contributed to the loss of over 1 million students in the U.S. public school system (Dee, Huffaker, Phillips, & Sagara, 2021). Other research has found that while students demonstrated gains in academic achievement during the 2020-21 school year, these gains were smaller than pre-pandemic trends in math and reading (Lewis, Kuhfeld, Ruzek, & McEachin, 2021).

There remain many open questions about both short and long term consequences of school closures and remote learning. Detailed analysis of many questions has been hampered by limited data on which schools and districts operated under which learning model (e.g., in-person, hybrid, virtual) over the school year, and when. For example, the Common Core of Data (CCD) within the National Center for Education Statistics (NCES), which serves as the federal database of the nation's public schools and districts, does not currently include this information (U.S. Department of Education, 2021).

The COVID-19 School Data Hub (CSDH) aims to provide comprehensive data to facilitate analyses of these questions. The CSDH is a database that draws from state-produced raw data on schooling mode in U.S. K-12 public schools and districts throughout the 2020-21 school year. In addition, the database includes COVID-19 case data from schools and districts during the 2020-21 school year, and data on masking policies in place by school districts for their staff and students. The data can be accessed at the [CSDH Home Page](#).

In this white paper, we briefly describe the data and data collection efforts. We also provide context on geographic variation in data availability.

Data and Data Collection Methods

The CSDH consists of three primary datasets: 1) Learning Model Data, 2) COVID-19 Case Data, and 3) Masking Data. We discuss our methods for collecting each of these components below.

Learning Model Data
The CSDH includes learning model data for K-12 public schools; it does not include private schools, schools serving only as pre-kindergarten/early childhood centers, student services spanning multiple school districts, programs designed for youth ages 18-21, or adults.

To create the CSDH, we submitted data requests to state education agencies (SEAs) for their raw data on learning models used by schools and districts during the 2020-21 school year. Our data requests consisted of the following components: 1) Learning models used by schools or districts throughout the school year, 2) Student enrollment *by learning model*, 3) NCES IDs for the data provided, 4) State-assigned IDs for the data provided, and 5) Staff employment counts, and 6) the number of staff working in-person at either the school- or district-level.

We requested the data at the school level if available, or the district level if not. In addition, we requested all data at the most frequent time periods available. Thus, while our ideal was school-by-week data, we accepted data at the district level, and at lower frequencies as well (e.g., monthly or quarterly). In some cases, states had only learning model data available, but not student enrollment by model. We accepted data in all formats and received a variety of single or multiple spreadsheets, PDFs, and links to publicly available datasets online. In addition to the data requests, we accessed total school and district student enrollment for the 2020-21 school year via SEA websites, and limited enrollment data to K-12, where possible.

We organized all data into a consistent structure across states, including learning model information by time period and NCES identifiers such as IDs and school/district type. When questions arose regarding data inconsistencies, missing information, time intervals, learning model definitions or other components, we communicated with the state for clarification. Data files were structured to include all schools or districts per state, across all time periods throughout the school year, even if we did not have learning model information for those cases. This approach aims to more clearly identify where data were missing or not defined.

Each state has an accompanying Data Details document, where we have provided detailed information about data sources, how the learning model information were coded by state and the corresponding CSDH labels, enrollment information, the data level included in the file (i.e., school or district), and information on the staff counts provided. The Data Details document also includes any information about missing data and decision rules followed by CSDH when organizing the data, when applicable. As of the CSDH launch in mid-September, we have clean data from 30 states and the District of Columbia, and all data are available in Excel and CSV form for download.

Learning Model Data Availability by State

There is significant variation in the availability of learning model data across U.S. states. At the most basic level, state SEAs varied in whether they collected any information or not. However, they also varied in whether they collected data at the district, school, grade band, or grade level; in the frequency with which they collected this information; and, importantly, in whether or not they made this information available via public dashboards or other means.

Much of the variation in data access reflects differences in the degree of centralized data collection. Some SEAs collected centralized data before the pandemic, making pandemic data collection easier, while in other cases learning model information was collected but was not able to be made public yet. In Table 1, we provide a summary of states' current level of data availability through the CSDH.

Table 1. Level of Learning Model Data Availability, by State

Description	States	*n*
Data Available Through the COVID-19 School Data Hub (as of 9/15/21)		**31**
Learning Model Data + Student Enrollment by Model Data Available at Least Monthly, October 2020-May 2021		**11**
a. School-level data	KY, NE, NY, ND, OR, RI, TX, VT	8
b. District-level data	MA, MI, MD	3
Learning Model Data Available at Least Monthly, October 2020-May 2021 (Inconsistent or Lack of Enrollment Data)		**12**
a. School-level data	IN, ME, MN, UT, WI	5
b. District-level data	CO, CT, MO, NJ, OH, VA, WY	7
Learning Model Data + Student Enrollment by Model Data Available Less Frequently Than Monthly, or For Only Part of School Year		**5**
a. School-level data	AR, FL	2
b. District-level data	AL, DC, WA	3
Learning Model Data Available Less Frequently Than Monthly, or For Only Part of School Year (Lack of Enrollment Data)		**3**
a. School-level data	-	0
b. District-level data	LA, NM, SC	3
Data Not Available Through the COVID-19 School Data Hub (as of 9/15/21)		**20**
Data Pending Integration into CSDH		**6**
a. Data received; pending review and/or response to questions.	CA, ID, KS, NH	4
b. Data expected to be received in Fall 2021	AK, IL	2
No Data Currently Available		**14**
a. The SEA reported that data of interest were not sufficiently collected.	AZ, GA, MS, MT, PA	5
b. The SEA reported that it is not providing public access to learning model data.	DE, NV, NC, TN	4
c. The SEA has not responded to our data request	HI, IA, OK, SD, WV	5
TOTAL		**51**

COVID-19 Case Data

In addition to comprehensive learning model information, CSDH also includes data on school and district COVID-19 cases. These are not contact tracing data and we do not provide data for where infections occurred. In addition, these data do not represent COVID-19 cases that states have aggregated based on school district boundary, as these cases are not specific to K-12 staff and students. These data only report positive cases in school and district populations as collected by each state. These data were collected throughout the 2020-21 school year via each state's public COVID-19 dashboards or reports from each state's Department of Health. Where data were not posted publicly, we submitted data requests to state departments of health for COVID-19 case counts for staff and students (separately), by school and/or district, as available, along with any information about relevant time periods.

Similar to the learning model data, raw COVID-19 case data were available in various formats, including single or multiple spreadsheets, PDFs, and publicly-available datasets. The data reflected several inconsistencies across states, with some states reporting new cases per week, others reporting active cases within the past two weeks or one month, and others reporting cumulative cases. Moreover, some states separated staff and student counts, while others aggregated this information or masked case counts. We created a consistent data structure across states, including time intervals at which the state reported the data, NCES identifiers such as IDs and school/district type, and all available COVID-19 case data.

In each state's Data Details document, we have provided information about how the case counts were reported, the reporting frequency, and the data source. At the time of the CSDH launch in mid-September, we have cleaned data from 30 states and the District of Columbia, and all data are available in Excel and CSV form for download (note, however, that these data do not represent the same states for which we have available learning model data). In Table 2, we report the COVID-19 case data available by state through CSDH.

Table 2. Level of COVID-19 Case Data Availability, by State

Description	States	*n*
Data Available Through COVID-19 School Data Hub as of September 15, 2021		**31**
a. Both school- and district-level data	IA	1
b. School-level data	CO, CT, FL, ID, IL, IN, KS, ME, MD, MS, MT, NV, NH, NY, NC, OR, RI, TX, VT, VA, WV	21
b. District-level data	AR, DC, DE, HI, KY, LA, MA, OH, UT	9
Data Not Available Through COVID-19 School Data Hub as of September 15, 2021		**20**
a. No COVID-19 Data Available	AL, AK, AZ, CA, GA, MI, MN, MO, NE, NJ, NM, ND, OK, PA, SC, SD, TN, WA, WI, WY	20
TOTAL		**51**

Masking Data

The third set of data available from CSDH includes masking policy data by school districts across the U.S. throughout the 2020-21 school year. These data document whether and when school districts had masking policies in place for staff and students (separately). In all, 31 states and the District of Columbia had state-level mask mandates in place through at least May 14, 2021; for the purposes of this project, we considered these to be full-year mask mandates for all districts in these states.

The remaining 19 states had either a state-level mask mandate that ended at some point during the 2020-21 school year (prior to May 1), or no mandate in place at any time. For these states, the CSDH administered a brief survey to district leaders (commonly, superintendents or principals), asking them to indicate if the district had a mask policy in place at any time for staff and students, and if so, to indicate when. Given the demands on education leaders' time, the survey did not delve into the nuance of the the varied policies in place in school districts across the U.S., such as policies on school buses, in the hallways, outside of the school building, and policies within classrooms that distinguished between, for example, instruction time and group work. Surveys were administered to nearly 6,000 education leaders (*n*=5,962) between June 23-July 31, 2021, with an average response rate of 28% by state (*n*=1,586).[1]

The Masking Data files include a single row of data for all school districts within the state, even if there are no available masking data. All files reflect a consistent data structure, with NCES district IDs, state IDs, and additional district-level information.

In each state's Data Details document, we have provided information about the source of the data and information about the state's mask mandate, if applicable. At the time of the CSDH launch in mid-September, we have data files with at least partial data from all 50 states and the District of Columbia, and all data are available in Excel and CSV form for download. In Table 3, we report CSDH's data availability of masking data by state.

Table 3. Masking Data Availability, by State

Description	*n*
Mask Mandate Data Available Through COVID-19 School Data Hub as of September 15, 2021	**32**
CA, CO, CT, DE, DC, HI, ID, IL, IN, KY, LA, ME, MD, MA, MI, MN, MS	
NV, NJ, NM, NY, NC, OH, OR, PA, RI, UT, VT, VA, WA, WV, WY	
Masking Survey Data Available Through COVID-19 School Data Hub as of September 15, 2021	**19**
AL, AK, AZ, AR, FL, GA, IA, KS, MO, MT, NE, NH, ND, OK, SC, SD, TN, TX, WI	
TOTAL	**51**

---

[1] All survey invitations were administered via email. CSDH team staff identified all contact information via state education agencies' directory information, if available, or via individual district websites. After the initial invitation was sent to school leaders, CSDH team staff worked to correct any email addresses that failed to send, and re-sent the survey invitation. At least two reminders were sent to all individuals who had not responded to the survey, and in some cases three reminders were sent.

Additional Data and Resources

In addition to these three primary datasets, we also provide additional data resources that we believe could be of use to researchers. This includes community COVID-19 case rates by NCES district ID, demographic information from the NCES Common Core of Data's 2019-20 school and district files, and an NCES School District and Zip Code Crosswalk (which provides a crosswalk of school district identifiers and location information, such as district name, state name, NCES district ID, state FIPS, county FIPS, county name, and zip code).

Finally, we host two other sources that could be used in conjunction with CSDH datasets to explore questions related to variation in student outcomes by learning model. This includes: 1) Burbio data, which tracked K-12 learning plans across school districts in 2020-21, representing 3,000 US counties in the US. These data are available for free but require researchers to complete an agreement with Burbio, and 2) SafeGraph data, which measure in-person visits to U.S. schools (from January 2019 - December 2020) using anonymized cell phone data. Both data sources are identified with NCES codes, allowing them to be easily merged with other sources.

Summary

The CSDH provides a publicly available database on K-12 school and district learning model data, COVID-19 cases, and masking policies. The project's aim is to allow researchers to address important questions about school reopening patterns over the 2020-21 school year and the consequences of these transitions - if any - on academic and health outcomes for students and adults.

References

Dee, T., Huffaker, E., Phillips, C., & Sagara, E. (2021). *The revealed preferences for school reopening: Evidence from public-school disenrollment.* (NBER Working Paper No. 29156). National Bureau of Economic Research. https://www.nber.org/papers/w29156

Lewis, K., Kuhfeld, M., Ruzek, E., McEachin, A. (2021). Learning during COVID-19: Reading and math achievement in the 2020-21 school year. Center for School and Student Progress, NWEA Research. https://www.nwea.org/content/uploads/2021/07/Learning-during-COVID-19-Reading-and-math-achievement-in-the-2020-2021-school-year.research-brief-1.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data. (2021). *2020-21 Common Core of Data (CCD) Preliminary Files*. Available at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021122rev.