SafeGraph and COVID-19 School Data Hub Comparison Project Whitepaper

October 20, 2022

1 Introduction

Over the course of the 2020-21 school year, schools and districts responded differently to the threat of COVID-19. Faced with uncertainty about the role of schools in the spread of the virus, school districts used a wide range of schooling modes (sometimes called "learning models") throughout the pandemic, including "school closures with virtual learning options, full-time in-person instruction, and various combinations of these approaches denoted as 'hybrid' schooling modes" (Halloran et al., 2021; Kaufman and Diliberti, 2021; NFES, 2021).

The COVID-19 School Data Hub (CSDH) has collected and centralized schooling mode data directly from state agencies from most U.S. states in order to improve researchers' understanding of how the pandemic and corresponding pandemic policies affected a variety of student outcomes. However, this data is largely at the month level (and is missing for some states). An alternate source of learning model information is available through cell phone data, which can provide more real-time, granular data on schooling mode (Parolin and Lee, 2021). These data are available nationwide and at a frequent time interval. However, cell phone data is only an indirect measure of schooling mode.

In this paper, we compare CSDH's schooling mode data through the pandemic with Parolin and Lee's U.S. School Closure and Distance Learning Database, sourced from data aggregation company SafeGraph's cell phone data. By comparing school cell phone traffic data with standardized, state-reported schooling mode data, we hope to better understand the reliability of cell phone traffic data as a proxy for schools' schooling modes. This exercise also informs and contextualizes the growing body of work that uses cell phone data to measure schooling mode directly (Fuchs-Schündeln et al., 2021; Kurmann and Lalé, 2021).

We provide descriptive statistics for both datasets to compare how they track schooling mode during the COVID-19 pandemic. We then use two sets of methods for comparing CSDH schooling mode data with the cell phone traffic data. We first conduct a simple concordance analysis in which we ask how often SafeGraph-inferred schooling mode would match the CSDH schooling mode. Second, we use a rich set of information from the SafeGraph data *and* district demographics from the National Center for Education Statistics to predict CSDH schooling mode using a multinomial logit model and a neural network. The latter analysis determines how effective we expect cell phone information would be in filling in missing schooling mode data, given other district characteristics. Our machine learning approach produces higher accuracy rates than the logit regression: in predicting the full CSDH data, the methods yields a maximum of 78 percent accuracy, with up to a 88 percent accuracy on certain subsamples.

2 Background and Data

Our analyses use three sources of data: (1) district-month level schooling mode and COVID-19 case count data from the 2020-21 school year; (2) district-month level cell phone traffic data from 2020 to 2021; and (3) district-level demographic, economic, housing, and geographical data. We explain these data sources below.

2.1 COVID-19 School Data Hub

Schooling mode and COVID-19 case count data are drawn from the COVID-19 School Data Hub (CSDH, 2022). The CSDH team submitted data requests to state education agencies (SEAs) for their record of schooling modes used by schools and districts during the 2020-21 school year. The team requested the data at either the school or district level, as available by the state, and at the most frequent reporting intervals possible. At each reporting interval, every school or district is classified as *in-person* if most students had access to traditional, daily in-person instruction; as *virtual* if all or most students received instruction online on a daily basis; or as *hybrid* if its schooling mode does not match either of these approaches. We note that for in-person school-months and district-months, students may still have had the opportunity to attend virtually. Currently, the CSDH has school-level data for 28 states. For 9 states, only district-level data is available. We begin with the most granular data available and collapse the data to the district-month level. Additionally, we obtain COVID-19 case counts and case rates from states (aggregated by CSDH), which were obtained and structured in a similar fashion: cases were reported at the school and district-levels in weekly, bi-weekly, and monthly intervals. Case numbers were then collapsed to the district-month level and merged with CSDH schooling mode data.

2.2 U.S. School Closure and Distance Learning Database (SafeGraph Data)

We use district-month level cell phone traffic data from the U.S. School Closure and Distance Learning Database (henceforth, "SafeGraph data"), which has tracked "in-person visits across more than 100,000 schools" monthly from August 2020 to May 2022. In this analysis, we focus on the period from August 2020 to June 2021. (Parolin and Lee, 2021). Parolin and Lee construct the database using aggregated cell phone data from SafeGraph, a company which uses application GPS data from around 40 million mobile devices to track mobility patterns. For each school-month, they calculate the change in traffic as a measure of schooling mode as a share of traffic during the same school-month in 2019. This data is aggregated to the district-month level, covering nearly 13,000 districts (about 94 percent of districts nationwide).

The final dataset contains four measures of visitorship for each district-month:

(1) Share of schools with at least 25 percent decline in visitors

(2) Share of schools with at least 50 percent decline in visitors

(3) Share of schools with at least 75 percent decline in visitors

(4) Mean change in in-person visits for all schools

Using the above measures, we produce simple predictions of schooling mode which we use in Section 3, on concordance matching. There, we classify district-months as virtual or non-virtual. If for a single district-month, measure (1) is greater than 0.5 (meaning that more than half of its schools experienced a 25 percent decline in visitors), that district-month is classified as virtual at the *25 percent threshold*. Each district month is classified analogously at the *50 percent threshold* and *75 percent threshold*.

2.3 National Center for Education Statistics

District-level demographic and socioeconomic data come from the National Center for Education Statistics (NCES) and the 2020-21 Common Core of Data (CCD) (NCES, 2022a). NCES constructs district-level statistics for children enrolled in each district using the most recent American Community Survey *Education Tabulation* (ACS-ED) five-year estimates. Specifically, we use select variables from the Select Economic Characteristics as well as the Demographic and Housing Characteristics tables, which includes data on districts' student-body sex, age, race, and family income. District-level locale and other socioeconomic data come from the 2020-21 NCES Common Core of Data (NCES, 2022b).[1]

2.4 Descriptive Statistics

Table 1 displays descriptive statistics for each of the samples used in this paper. Panels A and B give descriptive statistics for the 67,747 district-months observed in both the CSDH data and the SafeGraph data. Panel A gives the distribution of schooling modes across this sample. Just under half of all district-months had an in-person schooling mode (48 percent), while 33 percent and 19 percent of district-months were hybrid and virtual, respectively.[2]

[1]NCES classifies all school districts into one of four groups: *rural*, *town*, *suburb*, or *city*. Each district is further classified within one of three subgroups "based on population size or proximity to populated areas," with *small*, *midsize*, or *large* classifications for cities and suburbs based on population and *fringe*, *distant*, or *remote* classifications for towns and rural areas based on proximity to cities. For more information on data processing, see Appendix section 8.1

[2]Figure 1 displays geographic variation in schooling modes for fall 2020. To create this figure, we aggregate monthly schooling mode data for each district to calculate the modal schooling mode for the entirety of fall 2020. Figure 2 displays geographic variation in schooling modes analogously for spring 2021.

Panel B of Table 1 gives descriptive statistics for variables that come from the SafeGraph data, in the CSDH and SafeGraph merged sample. In the average district-month, just eight percent of schools experienced a 75 percent or greater drop in visitorship. This is consistent with the fact that, according to the CSDH, most districts avoided fully virtual learning during the 2020-21 school year.[3] However, over half of schools (56 percent) in the average district-month experienced a 25 percent or greater drop in visitorship. In other words, while virtual learning was uncommon, visitorship was often below pre-pandemic levels, as the average district-month saw a 23 percent decline in cellphone traffic.

Panels C and D of Table 1 present demographic and socioeconomic characteristics of the mean district in the CSDH, SafeGraph, and NCES merged data.[4] The sample size for panels C and D is lower than for Panels A and B because this information is invariant across time within the dataset, so entries are recorded at the district rather than district-month level. Of interest in Panel D is that there is sizable variation in each demographic characteristic across districts. Figure 4 visualizes districts' shares of white students, as an example of a district-level demographic characteristic that varies considerably.

3 Concordance Matching Approaches

The first approach we take is to determine how well a simple classification of the SafeGraph data performs in matching CSDH schooling mode. School-months in the CSDH data are classified as operating in one of three modes: virtual, hybrid or in-person. We explore concordance matching to determine whether a school is fully virtual or non-virtual (having at least some in-person learning).

In this method, we re-classify all district-months that are either hybrid or in-person in the CSDH data as *non-virtual*. Combining these classifications with the first set of predictions detailed in Subsection 2.2, district-months can be sorted into four groups:

(1) District-months classified as non-virtual by both the SafeGraph and CSDH data

(2) District-months classified as non-virtual by the SafeGraph data and virtual by the CSDH data

(3) District-months classified as virtual by the SafeGraph data and non-virtual by the CSDH data

(4) District-months classified as virtual by both the SafeGraph and CSDH data

Table 2 aggregates the results of this analysis, giving the percentage of district-months in our sample which fall into each group. Panels A, B, and C display these results across three different thresholds for classifying schools as virtual or non-virtual with SafeGraph data: the 25, 50, and 75 percent SafeGraph thresholds.[5] In each panel, the

[3] Data from the COVID-19 School Data Hub indicates that 19 percent of district-months were classified as virtual learning during the 2020-21 school year.

[4] Figure 3 visualizes data from Panel C geographically, mapping districts we observe according to their locale codes.

[5] See Section 2.2 for a more detailed description of these thresholds and how we use them to predict schooling mode.

sum of the top left and and bottom right cells give the percentage of district-months where SafeGraph and CSDH classifications align. We interpret this value as the *accuracy* of our first baseline method.

At the 25, 50, and 75 percent SafeGraph thresholds, we obtain accuracies of approximately 60, 80, and 83 percent, respectively. In other words, districts are correctly classified as virtual or non-virtual 60 percent of the time with a 25 percent drop in traffic as the closure threshold. We obtain 80 percent and 83 percent accuracy using 50 and 75 percent as the closure threshold, respectively. We also calculate accuracy for different subsets of our data. In Figure 5, we display the accuracy of our first baseline method across states using a 50 percent drop in traffic as a threshold for classifying schools as virtual, as an example. We note significant heterogeneity in accuracy across states, which could be influenced by a variety of state-level differences in characteristics such as locale and income.

We then constructed a concordance matching approach that allowed for classification into three, rather than just two, schooling modes. We constructed four classifications based on SafeGraph data, and explored their match with the full set three CSDH classifications (in-person, hybrid, and virtual). [6] Table 3 illustrates the share of district-months that fall into each classification. We then calculate the agreement score analogously to the simpler concordance method. [7] Compared to the simpler method, the agreement score for the full range of three CSDH classifications are lower at 0.51 percent. This is may be expected, as our tasks goes from classifying one of two possible outcomes to one of three, and a significant portion of this drop in agreement score was driven by district-months being classified as open in the CSDH data, but experiencing between a 25 and 50 percent drop in cell phone traffic.

4 Predictive Modeling

In this section, we outline and evaluate two supervised learning models which attempt to predict CSDH schooling modes at the district-month level using SafeGraph data. As noted in Section 1, districts' schooling modes are correlated with locale type, district demographics, and other variables. In Section 3, we find large differences in the accuracy rates of a simpler matching method across states. Our models take advantage of this insight, combining district information such as demographic and socioeconomic covariates with SafeGraph data to produce more accurate predictions of districts' schooling modes.

Both models use the same input data, pre-processing methods, and training and testing pipelines. Following the literature's best practices, numerical variables are normalized and categorical variables are converted to dummies (Ekenel and Stiefelhagen, 2006; Hancock and Khoshgoftaar, 2020). We train both models on the same training set, which we construct by randomly selecting 80 percent of all observations. Both models are evaluated on the same

[6] The four classifications for district-months are based on combinations of SafeGraph classifications: one, a shallow hybrid, saw the cell phone activity in a district-month drop by between 25 and 50 percent. Another, a deep hybrid, saw the cell phone activity in a district month drop by between 50 and 75 percent. With the addition for the possibilities of a less than 25 percent drop in traffic and a more than 75 percent drop in traffic, this accounted for all the possible changes in traffic a district-month could take on.

[7] See Table 3 caption for more details on calculations of agreement scores.

testing set consisting of the remaining observations, with accuracy scores reported separately across subsets of the testing set. We describe our pipeline in detail in the appendix.

4.1 Multinomial Logistic Regression Model

Our first model is a multinomial logistic regression model. This model is a generalization of conventional logistic models, which are used in prediction problems where there are *two* possible output classes, allowing for more than two output classes (Park and Kerr, 1990). In this case, we attempt to predict a schooling mode outcome which can take on *three* values (in-person, hybrid, or virtual). Our baseline model attempted to make predictions about CSDH schooling mode using very simple prediction algorithms based on SafeGraph data. By training a multinomial logistic regression model on SafeGraph data and NCES/CCD variables, we produce a more complex predictive algorithm which takes advantage of the relationship between schooling mode and NCES variables such as race, income, and locale.

The multinomial logit model predicts schooling mode with 70 percent accuracy across the entire testing set. There is significant heterogeneity in accuracy across subsets of the testing set. [8] In particular, the model performs best when evaluated on in-person district-months, rural districts, and city districts (as compared to districts that NCES classifies as suburbs or towns). Better accuracy in evaluating in-person district-months could be attributable the fact that the share of in-person district-months is larger than the shares of hybrid or virtual district-months.

One possible explanation for the model's strong performance in rural district-months and urban district-months relative to district-months of other locale types is that measures of cell phone traffic are less noisy in extremely low-density and extremely high-density areas. In school buildings located in rural districts, there may be very little ex ante cell phone traffic, so changes in cell phone traffic are easier to discern. In urban districts, schools may be so large in terms of enrollment that any change in schooling mode is associated with a significant absolute change in traffic. Suburban and town districts may not exhibit these characteristics, reducing predictive accuracy in moderate-density districts.

4.2 Neural Network Model

Our multinomial logit model imposes an assumption of linearity between predictors and outcomes. However, it is possible that the true relationship between district-month's schooling modes and SafeGraph or NCES variables is non-linear, potentially limiting our predictive accuracy. Neural networks with at least one hidden layer are universal approximators and should be able to better capture these non-linearities, if they exist (Scarselli and Chung Tsoi, 1998). For our second model, we thus implement a feed-forward neural network with three dense layers of sizes 60, 30, and three respectively. The ReLU activation function is applied between each layer.

[8] See Table 4 for multinomial logit accuracy scores.

Our neural network predicts schooling mode with generally higher accuracy than our multinomial logit model, correctly predicting 78 percent of district-months' schooling modes across the entire testing set. It performs best on different subsets of the data compared to the logit model. Specifically, its performance on hybrid district-months is slightly higher than its performance on in-person district months. [9]

We do not in this section display prediction accuracy across different neural network architectures. Because this prediction problem is relatively simple—there are only three possible output classes and 82 input variables—accuracy scores are not likely to differ substantially across architectures.

5 Conclusion

In this paper, we explored the extent to which SafeGraph cell phone traffic data is informative of CSDH schooling mode at the district-month level. We first attempt to answer this question using two simple baseline methods. Then, we develop two machine learning models which attempt to predict schooling mode using cell phone traffic data.

We find that baseline levels of match between the two datasets result in approximately 83 percent accuracy on a simplified form of the data in predicting virtual or non-virtual district-months. However, the accuracy of this baseline method varies considerably by state, and moving to a model which predicts exact learning mode results in reduced accuracy, so we implement a set of machine learning methods which additionally take this information, districts demographics, and COVID-19 case counts into account: using supervised learning methods results in significant improvements, with 78 percent accuracy in predicting three schooling modes. The machine learning methods also predict schooling modes of city and rural districts best, compared to districts in suburbs or towns (as classified by NCES). The methods are also more accurate in classifying in-person district-months and hybrid district-months relative to virtual district-months. These heterogeneities potentially make predictions in some communities more accurate than others – researchers should take note of this variable accuracy in interpreting cell phone traffic data.

There is no perfect substitute for the precision of state-reported schooling mode data. However, when such data is unavailable, cell phone traffic data could provide a useful alternative. It has many desirable qualities: it can be reviewed in near real-time, at a geographically granular level, and covers a near-unlimited range of locales. As such, better understanding how it compares to existing data sources is critical to its use. We find that with socioeconomic and demographic context, this data could be a powerful tool for the study of schooling mode and mobility more broadly.

[9]See Table 5 for neural network accuracy scores.

References

COVID-19 School Data CSDH. All School and District Learning Model Database, 2022. URL https://www.covidschooldatahub.com/for$_{r}esearchers.publisher : CSDH$.

H.K. Ekenel and R. Stiefelhagen. Analysis of Local Appearance-Based Face Recognition: Effects of Feature Selection and Feature Normalization. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 34–34, New York, NY, USA, 2006. IEEE. ISBN 9780769526461. doi: 10.1109/CVPRW.2006.29. URL http://ieeexplore.ieee.org/document/1640474/.

Nicola Fuchs-Schündeln, Dirk Krueger, André Kurmann, Etienne Lalé, Alexander Ludwig, and Irina Popova. The Fiscal and Welfare Effects of Policy Responses to the Covid-19 School Closures. Technical Report w29398, National Bureau of Economic Research, Cambridge, MA, October 2021. URL http://www.nber.org/papers/w29398.pdf.

Clare Halloran, Rebecca Jack, James Okun, and Emily Oster. Pandemic Schooling Mode and Student Test Scores: Evidence from US States. Technical Report w29497, National Bureau of Economic Research, Cambridge, MA, November 2021. URL http://www.nber.org/papers/w29497.pdf.

John T. Hancock and Taghi M. Khoshgoftaar. Survey on categorical data for neural networks. *Journal of Big Data*, 7(1):28, December 2020. ISSN 2196-1115. doi: 10.1186/s40537-020-00305-w. URL https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00305-w.

Julia Kaufman and Melissa Diliberti. Divergent and Inequitable Teaching and Learning Pathways During (and Perhaps Beyond) the Pandemic: Key Findings from the American Educator Panels Spring 2021 COVID-19 Surveys. Technical report, 2021. URL https://www.rand.org/pubs/research$_{r}eports/RRA168-6.html$.

Andre Kurmann and Etienne Lalé. School Closures and Effective In-Person Learning during COVID-19: When, Where and for Whom. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3963981. URL https://www.ssrn.com/abstract=3963981.

NCES. American Community Survey – Education Tabulation (ACS-ED), 2022a. URL https://nces.ed.gov/programs/edge/demographic/acs. publisher: National Center for Education Statistics.

NCES. Locale Classifications, 2022b. URL https://nces.ed.gov/programs/edge/geographic/schoollocations. publisher: National Center for Education Statistics.

NFES. Forum Guide to Attendance, Participation, and Engagement Data in Virtual and Hybrid Learning Models, June 2021. URL https://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NFES2021058. publisher: U.S. Department of Education.

Kang H. Park and Peter M. Kerr. Determinants of Academic Performance: A Multinomial Logit Approach. *The Journal of Economic Education*, 21(2):101, 1990. ISSN 00220485. doi: 10.2307/1181978. URL https://www.jstor.org/stable/1181978?origin=crossref.

Zachary Parolin and Emma K. Lee. Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature Human Behaviour*, 5(4):522–528, April 2021. ISSN 2397-3374. doi: 10.1038/s41562-021-01087-8. URL `http://www.nature.com/articles/s41562-021-01087-8`.

Franco Scarselli and Ah Chung Tsoi. Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results. *Neural Networks*, 11(1):15–37, January 1998. ISSN 08936080. doi: 10.1016/S0893-6080(97)00097-X. URL `https://linkinghub.elsevier.com/retrieve/pii/S089360809700097X`.

6 Tables

Table 1: Descriptive Statistics

	Mean	S.D.	N
Panel A: CSDH Variables (District-Month Level)			
Modal learning model: Hybrid	0.33	0.47	67747
Modal learning model: In-person	0.48	0.50	67747
Modal learning model: Virtual	0.19	0.39	67747
COVID-19 positivity rate	0.07	0.04	67747
Panel B: Safegraph Variables (District-Month Level)			
Total number of schools	6	9.87	67747
Share of schools with at least 25 pct. decline in visitors	0.56	0.37	67747
Share of schools with at least 50 pct. decline in visitors	0.30	0.34	67747
Share of schools with at least 75 pct. decline in visitors	0.08	0.19	67747
Mean change in in-person visits for all schools	-0.23	0.35	67747
Total number of students	2950	6741.04	67747
Panel C: NCES Locale Variables (District Level)			
Large city	0.03	0.16	7135
Midsize city	0.01	0.11	7135
Small city	0.03	0.18	7135
Large suburb	0.24	0.43	7135
Midsize suburb	0.03	0.16	7135
Small suburb	0.02	0.14	7135
Fringe town	0.05	0.21	7135
Distant town	0.09	0.29	7135
Remote town	0.05	0.22	7135
Fringe rural	0.12	0.32	7135
Distant rural	0.20	0.40	7135
Remote rural	0.13	0.34	7135
Panel D: NCES Demographic Variables (District Level)			
Pct. male, total population	0.51	0.06	6539
Pct. Native American/Alaska Native	0.02	0.09	7117
Pct. Hispanic/Latino	0.15	0.20	7117
Pct. Black	0.08	0.17	7117
Pct. white	0.69	0.28	7117
Pct. Asian	0.03	0.06	7117

This table contains descriptive statistics for each sample in the paper. Panels A and B provide descriptive statistics for district-months observed in both the COVID-19 School Data Hub and SafeGraph data. Panel A provides statistics for the CSDH variables. Panel B provides statistics for the SafeGraph variables. Panels C and D provide descriptive statistics for the CSDH, SafeGraph, and NCES merged sample. Panel C provides statistics for the NCES locale variables, and Panel D provides statistics for the NCES demographic variables.

Table 2: Baseline Analysis with 2-by-2 Tables

	Non-virtual from CSDH	Virtual from CSDH
Panel A *Threshold of 25 pct. decrease in visitorship*		
Non-virtual from SG	0.45	0.04
Virtual from SG	0.37	0.14
Panel B *Threshold of 50 pct. decrease in visitorship*		
Non-virtual from SG	0.70	0.08
Virtual from SG	0.12	0.10
Panel C *Threshold of 75 pct. decrease in visitorship*		
Non-virtual from SG	0.81	0.16
Virtual from SG	0.01	0.02

This table provides 2x2 tables for each of the SafeGraph closure thresholds of 25, 50, and 75 percent drops in traffic in Panels A, B, and C respectively. Each district-month in the COVID-19 School Data Hub and SafeGraph merged sample is classified by its CSDH status and SafeGraph status. Entries along the main diagonal of each Panel are considered to 'agree' between the two datasets.

	Open 25	Closed 25, Open 50	Closed 50, Open 75	Closed 75
Open Hub	.2654878	.1543907	.0652955	.0148399
Hybrid Hub	.0785748	.1229412	.1013433	.0185743
Closed Hub	.0268233	.0349748	.0788117	.0379428

Table 3: This table provides the 3x4 concordance matching approach for all district-months in the sample. Each district-month in the COVID-19 School Data Hub and SafeGraph merged sample is classified by its CSDH status and combinations of SafeGraph statuses. Entries in positions [1,1], [2,2], [2,3], and [3,4] are considered to "agree" between the two datasets.

	Accuracy Score (w/o district dummies)
Entire Sample	0.70
Hybrid District-Months	0.60
In-person	0.81
Virtual	0.59
City Districts	0.71
Suburban Districts	0.67
Town Districts	0.69
Rural Districts	0.73

Table 4: This table displays the prediction accuracy scores for the multinomial logit model across the entire testing set and within subsets of the testing set.

	Accuracy Score (w/o district dummies)
Entire Sample	0.77
Hybrid District-Months	0.80
In-person	0.79
Virtual	0.69
City Districts	0.80
Suburban Districts	0.74
Town Districts	0.77
Rural Districts	0.79

Table 5: This table displays the prediction accuracy scores for the neural network across the entire testing set and within subsets of the testing set.

7 Figures

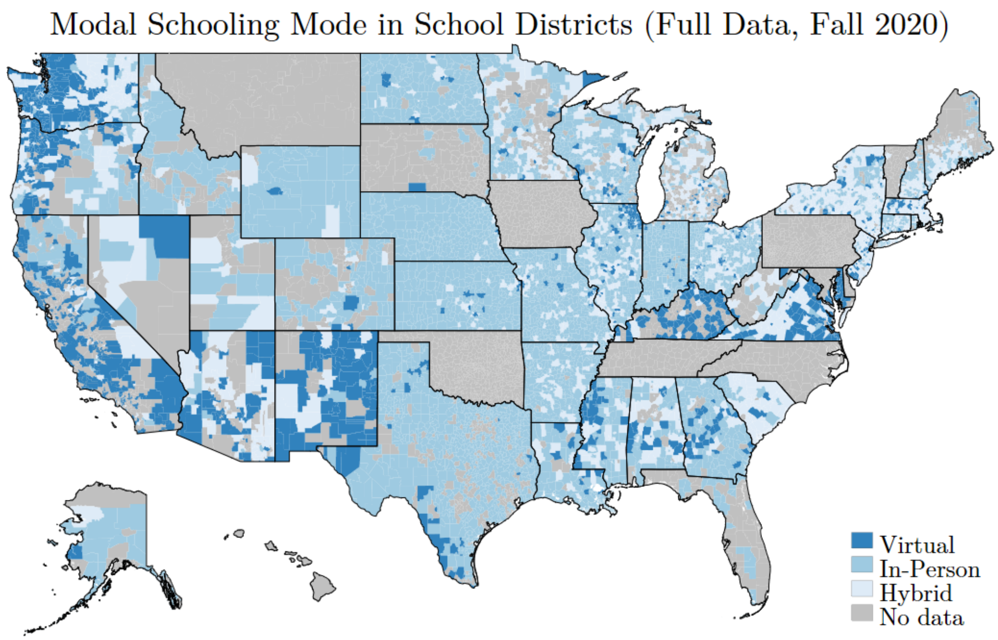

Figure 1: This figure depicts the most common schooling mode at the district level for each district in the CSDH and SafeGraph merged sample for the months from August to December 2020.

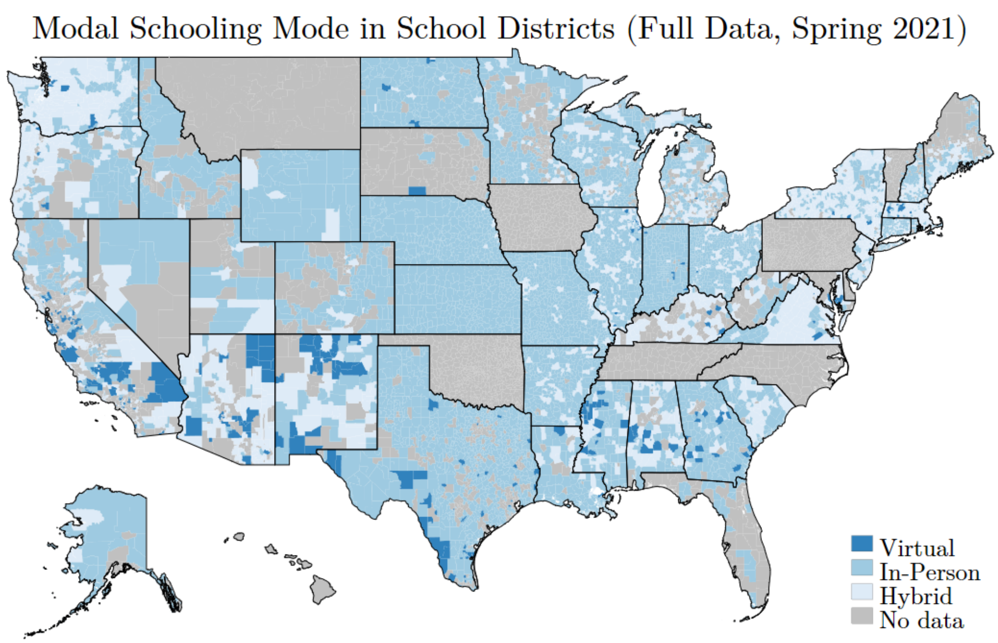

Figure 2: This figure depicts the most common schooling mode at the district level for each district that appears at least once in the CSDH and SafeGraph merged sample from January to June 2021.

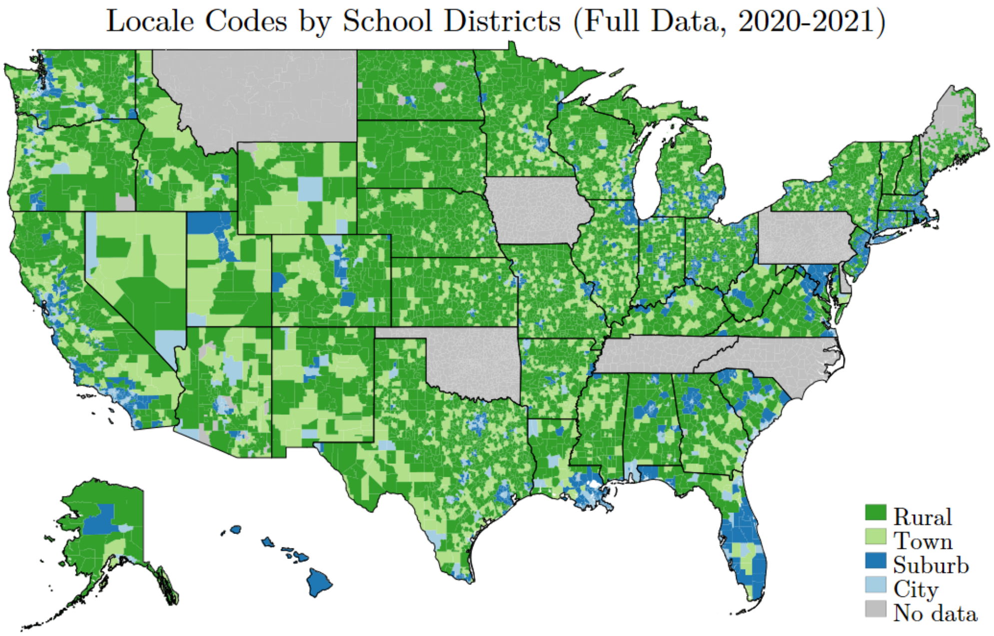

Figure 3: This figure depicts the locale code at the district level for each district that appears at least once in the CSDH, SafeGraph, NCES merged sample. Each district is classified as rural, town, suburban, or city.

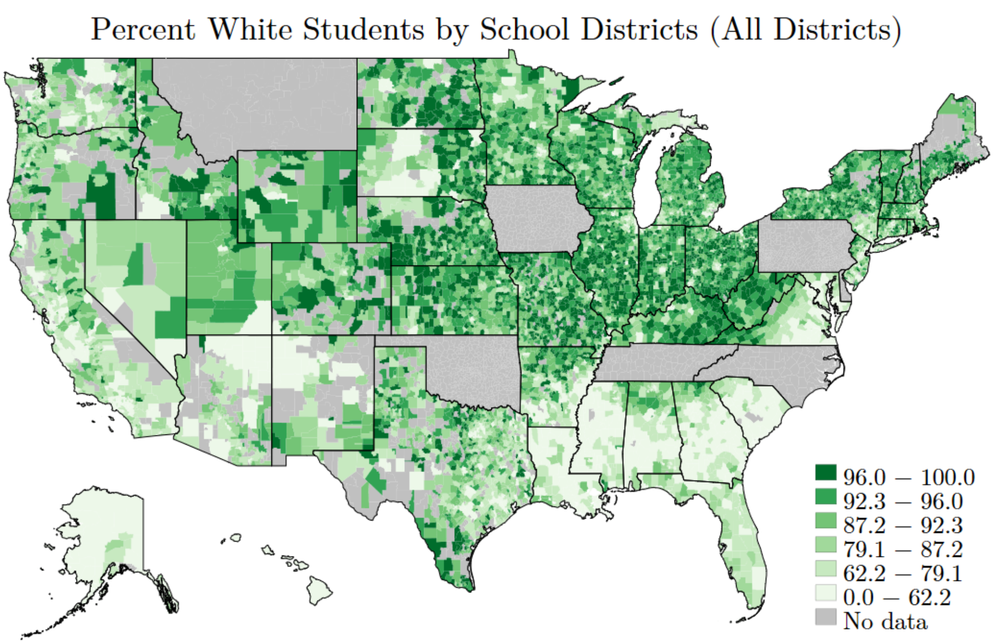

Figure 4: This figure depicts the share of relevant students in each school district that appears at least once in the CSDH, SafeGraph, NCES merged sample who are white. Districts with the highest shares of white students are located in New England and the Midwest. Districts with the lowest shares of white students are located in the Southeast and Southwest.

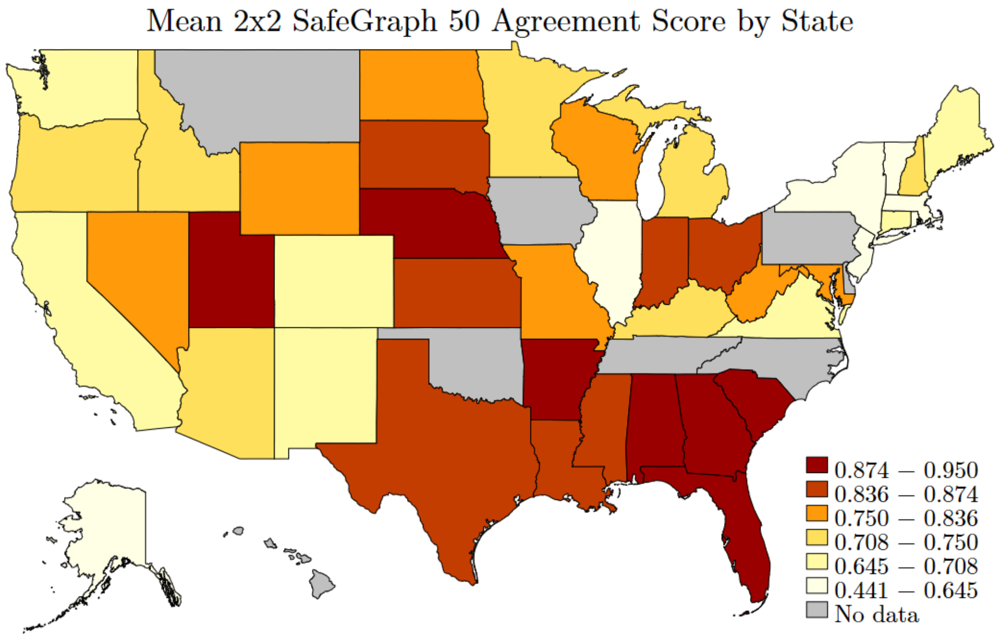

Figure 5: This figure displays the mean 2x2 agreement score at the SafeGraph 50 threshold for districts in each state. The highest agreement scores are found in states in the Southeast, as well as Nebraska and Utah. The lowest agreement scores were found in states in the Northeast, in addition to Alaska and Illinois.

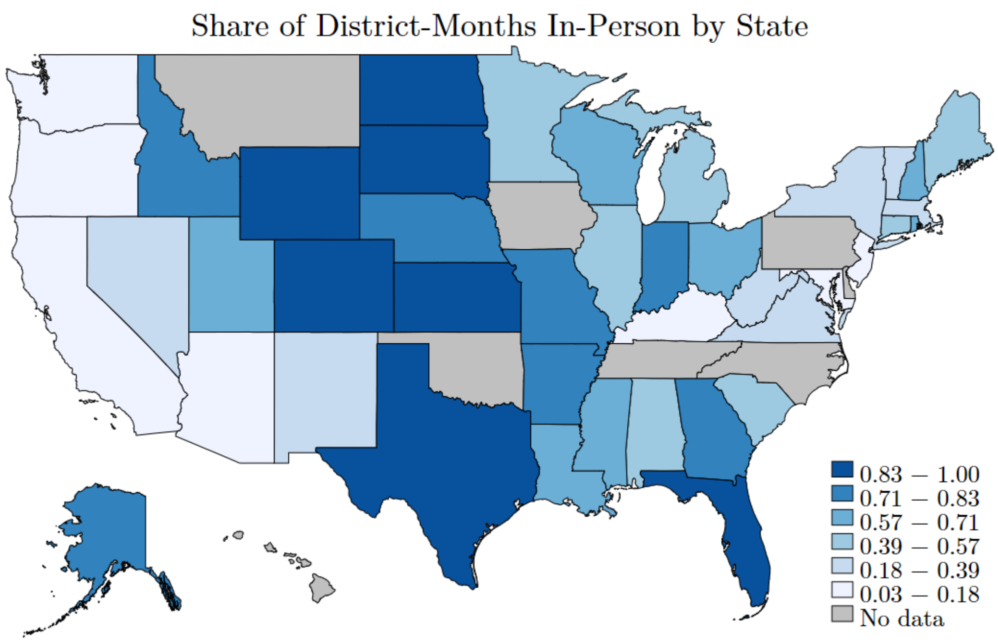

Figure 6: This figure displays the share of district-months spent in in-person schooling in each state. States with the most district-months in-person were Florida, Texas, North Dakota, South Dakota, Wyoming, Colorado, and Kansas. States with the least district-months in-person were on the West Coast, as well as Arizona, Kentucky, Maryland, and New Jersey.

8 Appendix

8.1 Methodology: Collapse of CSDH schooling mode data

For 28 states, the CSDH team has access to school-level data. For nine states, only district-level data is available. We begin with the most granular data available. In this initial dataset, each observation can represent either a school or a district at the monthly, weekly, bi-weekly, or biweekly level. We collapse this data to the district-month level using the following steps:

(1) We assign a month and year to each school-reporting interval and district-reporting interval, calculated as the month and year of the reporting interval's middle day.

(2) We consider rows which do not have a single schooling mode—for instance, within some district-months, schooling modes varied across grades. In such cases, the CSDH team records schooling mode separately for elementary, middle, and high school students. We impute schooling modes for these rows as follows:

   (a) When two of these three groups of students have the same schooling mode within a single row, we assign this schooling mode as the row's overall schooling mode.

   (b) We drop a small portion (less than six percent) of the sample containing district-months for which we cannot impute an overall schooling mode following this method

(3) For the purposes of this analysis, we re-classify the small portion (less than five percent) of district-months containing *closed* schools as *virtual.*

(4) We group observations by district-month and calculate the modal schooling mode within each group. We break ties following three rules:

   (a) Groups containing ties between *hybrid* and *in-person* are classified as *hybrid*

   (b) Groups containing ties between *hybrid* and *virtual* are classified as *hybrid*

   (c) Groups containing ties between *virtual* and *in-person* are classified as *virtual.*

Before merging CSDH schooling mode data with other datasets, we are left with a dataset at the district-month level; 11,433 school districts are observed over 113,066 district-months.

For district-months where data is missing in one or both the ACS-ED Tabulation and Common Core of Data, we impute these values as -1, then include a "missing" indicator to ensure that the observation is included, but that it does not affect the coefficients on other variables.

8.2 Methodology: Supervised Learning Pipeline

We begin with a restricted sample of 62,842 observations for which all variables listed in Table 1 are non-missing. Then, we take the following steps:

(1) Following best practices in the supervised learning literature, we normalize the values of each numeric variable (for example, *total students*) (Ekenel and Stiefelhagen, 2006).

(2) Also in keeping with best practices, we one-hot encode all categorical variables (for example, *locale type*) (Hancock and Khoshgoftaar, 2020). One-hot encoding a categorical variable is equivalent for generating a dummy variable for each of its values. One of these dummies is dropped before estimating the multinomial logit models.

   (a) We note that even though *NCES District ID* is not listed as a variable in Table 1, we also one-hot encode this categorical variable.

(3) We randomly select 80 percent of the observations in the resulting dataset and train both models on this dataset. The remaining 20 percent of observations constitute our testing set. Randomization is stratified so that each set contains the same distribution of *schooling mode*.

(4) For each model, we generate predictions of *schooling mode* for each observation in the testing set.

(5) For each model, we calculate prediction accuracy as the portion of predictions which match the true value of *schooling mode*.