Kurmann, André; Lalé, Etienne

Working Paper

School closures and effective in-person learning during COVID-19: When, where, and for whom

Document de travail, No. 2022-01

Provided in Cooperation with:
Department of Economics, School of Management Sciences (ESG UQAM), University of Quebec in Montreal

Suggested Citation: Kurmann, André; Lalé, Etienne (2022) : School closures and effective in-person learning during COVID-19: When, where, and for whom, Document de travail, No. 2022-01, Université du Québec à Montréal, École des sciences de la gestion (ESG UQAM), Département des sciences économiques, Montréal

This Version is available at:
https://hdl.handle.net/10419/264153

ESG
RECHERCHE

DOCUMENT DE TRAVAIL / WORKING PAPER

No. 2022-01

School Closures and Effective In-Person Learning during COVID-19: When, Where, and for Whom

André Kurmann et Etienne Lalé

Janvier 2022

ESG UQÀM
Département des sciences économiques
École des sciences de la gestion
Université du Québec à Montréal

School Closures and Effective In-Person Learning during COVID-19: When, Where, and for Whom

André Kurmann, Drexel University

Etienne Lalé, Université du Québec à Montréal

Document de travail No. 2022-01

Janvier 2022

Département des Sciences Économiques
Université du Québec à Montréal
Case postale 8888,
Succ. Centre-Ville
Montréal, (Québec), H3C 3P8, Canada
Courriel : frechette.karine@uqam.ca
Site web : http://economie.esg.uqam.ca

School Closures and Effective In-Person Learning during COVID-19: When, Where, and for Whom*

André Kurmann
Drexel University

Etienne Lalé
Université du Québec
à Montréal

December 21, 2021

Abstract

We match cell phone data to administrative school records and combine it with information on school learning modes to study effective in-person learning (EIPL) in the U.S. during the pandemic. We find large differences in EIPL for the 2020-21 school year. Public schools averaged less EIPL than private schools. Schools in more affluent localities and schools with a larger share of non-white students provided lower EIPL. Higher school spending and federal emergency funding is associated with lower EIPL. These results are explained in large part by regional differences, reflecting political preferences, vaccination rates, teacher unionization rates, and local labor conditions.

JEL Classification: E24, I24
Keywords: COVID-19; School closures and reopenings; Effective in-person learning; Inequality

*Contact: Kurmann: Drexel University, LeBow College of Business, School of Economics, 3220 Market Street, Philadelphia, PA 19104 (email: kurmann.andre@drexel.edu); Lalé: Université du Québec à Montréal, Department of Economics, C.P. 8888, Succ. centre ville, Montréal (QC), H3C 3P8 Canada (email: lale.etienne@uqam.ca). Aseni Ariyaratne provided excellent research assistance. We thank Dennis Roche from Burbio, Nat Malkus from Return to Learn, and Safegraph for generously sharing their data. We also thank seminar participants at the 2nd joint IZA & Jacobs Center Workshop on "Consequences of COVID-19 for child and youth development" for their comments. All errors are our own.

1 Introduction

The COVID-19 pandemic led many schools in the U.S. to suspend or substantially reduce in-person learning. While available studies report conflicting results on the extent to which school closures helped prevent the spread of the virus,[1] evidence is emerging that remote instruction led to substantial learning losses and social-emotional harm with possibly large adverse long-term effects, especially for students from disadvantaged backgrounds.[2] A key input for analyzing these consequences is the availability of consistent estimates of effective in-person learning (EIPL) across schools. The goal of this paper is to provide such estimates and analyze the extent to which differences in EIPL reflect school characteristics as opposed to systematic geographic factors.

We match anonymized cell phone data from Safegraph on school visits to the universe of U.S. public and private schools from the National Center for Education Statistics (NCES). We then map changes in school visits during the pandemic to information on school learning mode by Burbio and Return2Learn (R2L) to estimate EIPL for a sample of almost 70,000 schools that is highly representative. We organize our analysis of this dataset into three parts.

First, we provide new evidence on the disparities in EIPL over time and across regions of the U.S. While EIPL dropped to less than 20% of its pre-pandemic level in most places during Spring of 2020, EIPL increased to over 50% on average during the 2020-21 school year but with large differences across regions. For instance, in cities in Florida such as Jacksonville, Orlando or Tampa, EIPL averaged more than 75% from September 2020 to May 2021 whereas in cities in California, Oregon and Washington such as Los Angeles, Portland or Seattle, EIPL averaged 20% or less.

At the same time, we find that even within counties, there are sizable differences in EIPL across schools. This suggests that some schools returned to in-person learning more quickly than others not just because of systematic regional differences, but also because of school and local characteristics that apply similarly across the country.

In the second part, we investigate the extent to which pre-pandemic school and local characteristics predict EIPL during the 2020-21 school year. Naturally, these correlations should not be interpreted as causal, but they provide us with a set of stylized facts to understand the factors behind school closings and which segments of the student population were most affected. We find the following main results:

1. Public schools provided substantially less EIPL than private schools, with public charter schools ranking below public non-charter schools and private religious schools ranking above private non-religious schools.

2. For both public and private schools, EIPL was lower in more affluent and more educated localities with a larger share of dual-headed households, and for schools with a larger share of non-white students.

3. For public schools, EIPL is negatively related to pre-pandemic school test scores, school size, and school spending per student as well as district Elementary and Secondary School Emergency Relief (ESSER) funding per student.

[1] See Bravata et al. [2021], Chernozhukov et al. [2021a, 2021b], Ertem et al. [2021], and the references therein.

[2] See Dorn et al. [2021], Halloran et al. [2021], Kogan and Lavertu [2021], Lewis et al. [2021] as well as Agostinelli et al. [2020], Fuchs-Schündeln et al. [2021] or Jang and Yum [2020] among others.

In the third part, we ask to what extent these results are driven by systematic regional differences unrelated to schools. We find that most of the inverse relation of EIPL with local affluence, education, parental structure, and school funding is explained by whether a school is located in a state with a Republican governor and the margin by which the county voted Republican in the 2020 presidential election. COVID vaccination rates also predict higher EIPL while a state's teacher unionization rate and local labor conditions for teachers predict lower EIPL. In contrast, local COVID case and death rates and whether a school is located in a city, suburb, or town/rural area do not have significant predictive power. The results suggest that school reopenings were driven in large part by policies that align with local political preferences.

At the same time, we find that even after controlling for different school characteristics and regional factors, there remains a sizable negative relation of EIPL with a school's share of non-white students and, to lesser extent, district ESSER funding. This is important both because schools with a larger share of non-white students generally perform worse in terms of learning outcomes, and because ESSER was advertised in Congress primarily as support for schools to reopen to in-person learning.

The paper is part of a growing literature that attempts to measure the extent and consequences of school closures during the pandemic. The papers closest to ours are Bravata et al. [2021], Chernozhukov et al. [2021a], and Ertem et al. [2021] who use Safegraph, respectively Burbio to predict the effect of school reopenings on COVID infections and deaths, Dee et al. [2021] who use Burbio to study the effect of school closings on school enrollment, and Parolin and Lee [2021] who use Safegraph to estimate school closures. The contribution of our paper relative to this literature is two-fold.

First, the EIPL measure we construct has clear advantages over estimates of school closing based solely on mobility data such as Safegraph or learning mode trackers such as Burbio and R2L. While Safegraph covers a large and representative sample of all U.S. schools, it is not clear what a given decline in visits to a school during the pandemic represents in terms of lost in-person learning.[3] Furthermore, attributing cell phones to a particular location is challenging, thus raising questions about the reliability of the information. The learning mode trackers from Burbio and R2L, in turn, cover only public schools and provide only averages at the county-, respectively the district level. This limits the analysis of how EIPL relates to school characteristics and local conditions surrounding the school. Even more importantly, we find large variations across the two trackers in the percentages spent in traditional, hybrid, and virtual learning modes that vary by region and time period. These variations are due in part to differences in how the three learning modes are defined and in part to differences in coverage. By mapping the Safegraph data separately to both Burbio and R2L at various geographic levels and retaining the estimate with the highest predictive power, we attempt to address these limitations while performing important quality checks.

Second, we relate our estimate of EIPL to a host of indicators measuring not just the socio-economic characteristics of a school's student body, but also local conditions and school test scores, size, and funding. We document striking associations between these variables and EIPL that, to our knowledge, are new. Furthermore, we show that many but not all of these associations are driven by systematic regional differences that are correlated with political preferences, vaccination rates, and local labor market conditions. The results raise critical questions about the extent of school closings in different parts of

[3] See the appendix for further discussion.

the country. The patterns of EIPL that we uncover are also important to understand the impact of in-person learning loss during the pandemic on future educational attainment and income inequality (see Fuchs-Schündeln et al. [2021] for an application). At the same time, we emphasize that our analysis does not contribute to the ongoing debate of whether and under what circumstances school closings helped slow down the spread of COVID.

2 Data sources, matching, and sample restrictions

This section describes the different datasets used to construct our EIPL measure. Further details and extensive analysis of the data are provided in the online appendix.

2.1 Safegraph school visit data

The primary source of information for measuring in-person learning comes from Safegraph, which provides data on over 7 million Places of Interest (POIs) for the U.S., including visits derived from over 40 million anonymized cell phones. We retain all POIs with North American Industry Classification System (NAICS) code 611110 ("Elementary and Secondary Schools") that have consistent weekly visit data. We match these POIs by school geolocation or name and address to the universe of public and private schools from the NCES's Common Core of Data (CCD) and the Private School Universe Survey (PSS), which results in about 102,500 high-quality matches. Relative to the universe of schools, we lose about 22,000 schools. But the matched sample remains highly representative in terms of demographic and geographic makeup.

2.2 Constructing changes in school visits

The weekly visit count for each school is organized in seven dwell time intervals, ranging from less than 5 minutes to more than 240 minutes. Visits decline during major holidays and summer break; and following the declaration of the public health emergency on March 13, 2020, visit drop precipitously and remain substantially lower on average thereafter. At the same time, due to the increase in cell phones covered by Safegraph, visits generally trend upward prior to the pandemic; and visits to individual schools can be subject to substantial variation, both from one week to another and across dwell time intervals. While some of these variations reflect school characteristics and idiosyncratic events, others are due to the inherent difficulty of attributing cell phones to a particular location.

To address these measurement issues, we proceed in three steps. First, we construct a dwell-time weighted average of weekly visits for each school that is normalized by the weekly count of cell phones covered by Safegraph at the county level

$$\widetilde{v}_{j,t} = \frac{1}{n_{c(j),t}} \sum_{d=1}^{7} \omega_j(d)\, v_{j,t}(d), \tag{1}$$

where $v_{j,t}(d)$ denotes raw visits of dwell time d for school j in week t; $\omega_j(d) = \frac{\sum_{t=t_{-1}}^{t_0} v_{j,t}(d)}{\sum_{t=t_{-1}}^{t_0} v_{j,t}}$ measures the importance of visits of dwell time d for school j during reference period $t = t_{-1}, \ldots, t_0$ beginning in November 2019 through the end of February 2020 (excluding the weeks of Thanksgiving, Christmas and

New Year); and $n_{c(j),t}$ is the normalization by the count of devices during week t in county $c(j)$ in which school j is located.

Second, we drop about 33,000 schools with sparse or noisy visit data. For the approximately 70,000 schools that remain, we construct sample weights that we use in the regressions below to keep the results representative of the universe of schools.

Third, we construct the change in school visits as the percent difference in dwell-time weighted, normalized visits relative to the average $\widetilde{v}_{j,0} = \frac{1}{t_0 - t_{-1} + 1} \sum_{t=t_{-1}}^{t_0} \widetilde{v}_{j,t}$ over the reference period

$$\Delta \widetilde{v}_{j,t} = 100 \times \frac{\widetilde{v}_{j,t} - \widetilde{v}_{j,0}}{\widetilde{v}_{j,0}}. \tag{2}$$

2.3 Burbio and Return2Learn learning mode data

Burbio publishes a weekly School Opening Tracker for almost all U.S. counties based on information from 1,200 public school districts representing 47% of U.S. K-12 student enrollment in over 35,000 schools. The tracker starts in August 2020 and reports the share of public school students engaged in either "Traditional", "Hybrid", or "Virtual" learning mode in a county, where Hybrid means that students attend school 2-3 days per week in-person.[4]

Return2Learn (R2L) is another learning mode tracker constructed by the American Enterprise Institute and Davidson College. Like Burbio, the R2L data consists of weekly indicators from August 2020 onward of the share of public school students engaged in one of the three learning modes, although their definitions differ somewhat.[5] The data is available at the school district level, covering about 8,000 districts in over 3,000 counties that account for 90% of U.S. K-12 student enrollment.

Given the differences in definition and coverage, it is important to compare the two trackers. As shown in the appendix, while the two trackers generally show a similar evolution over time, there are also sizable differences, especially with regards to hybrid versus traditional learning modes. These differences are more pronounced for some regions and time periods than others. Over the entire sample, the correlation between traditional modes across the two trackers at the county level averages 0.78 across counties, with a 5-95 percentile range of 1.42 (i.e. for a non-negligible share of counties, the correlation is negative). Furthermore, since these trackers only cover public schools and are only available at the county, respectively the district level, they cannot be used to document differences in school closures by school characteristics.

3 From school visits to effective in-person learning

To construct a measure of EIPL, we map changes in school visits from Safegraph to changes in learning mode from Burbio and R2L. This is the main methodological contribution and proceeds in two steps. First, we aggregate changes in weekly visits to the county level for Burbio, respectively the district level for R2L, and estimate the mapping. Second, we use the estimates to predict EIPL at the individual school level.

[4] See https://about.burbio.com/methodology/ for further details.
[5] See the appendix and https://www.returntolearntracker.net/about/ for details.

For expositional purposes, we focus on the county-level aggregation; the steps for the district-level aggregation are analogous. Denote by $\Delta\widetilde{v}_{c,t} = \sum_{j \in c} \kappa_j \Delta\widetilde{v}_{j,t}$ the average change in school visits in county c in week t relative to the reference period, where κ_j is the share of county c's students enrolled in school j. Next, define the fraction of total school time that is EIPL by students in county c during week t as

$$EIPL_{c,t} = T_{c,t} + \gamma H_{c,t}, \tag{3}$$

where $T_{c,t}$ is the share of students in traditional in-person learning mode, $H_{c,t}$ is the share of students in hybrid learning mode, and γ defines the fraction of hybrid learning spent in person.

Since both $\Delta\widetilde{v}_{c,t}$ and $EIPL_{c,t}$ measure percent deviations from the pre-pandemic baseline, we can formulate the relationship between the two variables as $EIPL_{c,t} = \alpha + \beta\Delta\widetilde{v}_{c,t} + \varepsilon_{c,t}$, or equivalently

$$T_{c,t} = \alpha + \beta\Delta\widetilde{v}_{c,t} - \gamma H_{c,t} + \varepsilon_{c,t}. \tag{4}$$

The regression tells us not only how a given change in school visits maps into EIPL, but also the proportion γ of hybrid learning spent in person.

While (4) can be estimated as a panel across all counties, respectively districts, one important concern is that the mapping between changes in school visits and learning modes may not always be the same; e.g. because of differences in hybrid learning arrangements across school districts. A second concern is that the quality of the Burbio and the R2L data differs across regions and time periods.

To address these concerns, we estimate (4) separately for the Burbio and the R2L data at both the state level and the Core-Based Statistical Areas (CBSA) level. For each regression, we restrict $\alpha = 100$ as implied by the pre-pandemic baseline when schools were fully in person (i.e. $T_{c,0} = 100$, $\Delta\widetilde{v}_{c,0} = 0$, and $H_{c,0} = 0$), and choose the sample period between August 2020 and May 2021 with the highest R-squared. This provides us with two sets of Burbio estimates $\left\{\hat{\beta}_{\text{Burbio}}, \hat{\gamma}_{\text{Burbio}}\right\}$ and two sets of R2L estimates $\left\{\hat{\beta}_{\text{R2L}}, \hat{\gamma}_{\text{R2L}}\right\}$ for every school j. To predict $EIPL_{j,t}$, we then use the regression coefficients associated with the highest R-squared (either at the state or CBSA level) subject to the restrictions that the proportion of hybrid learning spent in person is within its theoretical bounds, $0 < \hat{\gamma} < 1$, and the highest regression R-squared is at least 0.25.[6]

The algorithm effectively trades off geographic variation in regression coefficients with regression fit and does so by using either the Burbio or the R2L data that is of higher quality. We could apply the same algorithm at the county level. In many cases, however, this would result in R-squared that are lower than at the CBSA level. More generally, we have experimented with several modifications of the algorithm, and the results reported below remain very robust.

Table 1 reports summary statistics for the estimated coefficients retained to predict $EIPL_{j,t}$. As shown in panel (a), across the approximately 70,000 schools, the retained estimates are evenly distributed between Burbio and R2L and between the state and CBSA level, indicating that both Burbio and R2L data are useful and that allowing for finer geographic variation would in about half the cases produce a worse fit.[7]

[6] The restrictions on γ applies only in a few cases. R-squared is lower than 0.25 for all four regressions in Arkansas and Maine, where both the Burbio and R2L data appear to be of low quality. For these states, we use regression coefficients from neighboring states. See the appendix for details.

[7] One may be concerned that the Safegraph data is more noisy in some parts of the country, but the pre-pandemic variance

The second panel shows the school-weighted distribution of the retained estimates and R-squared across schools. The regressions are generally tightly estimated with a median R-squared of 0.86 and the estimated coefficients are almost all highly significant. The estimated mapping between changes in school visits and EIPL, $\hat{\beta}$, ranges from about 0.8 to 1.5 while the estimated proportion of hybrid learning spent in person, $\hat{\gamma}$, ranges from about 0.05 to 0.7. As confirmed in further analysis, these distributions reflect large regional variations in the mapping from school visits to EIPL. Focusing on means, a one percentage point decline in school visits predicts an average reduction in EIPL by 1.2 percentage points, and the predicted fraction of hybrid learning spent in person is 0.3 or the equivalent of 1.5 days out of a 5 day school week.

4 The "when" and "where" of in-person learning

We start our empirical analysis by documenting differences in EIPL over time and across regions. To do so, we compute average student-weighted EIPL for each county for which we have at least three schools with reliable data.

As documented in more detail in the appendix, from March to May 2020, EIPL dropped to between 0% and 20% for almost all counties, as many schools switched to virtual learning. From September 2020 to May 2021, in contrast, we observe large disparities in EIPL, reflecting broad differences in school reopening policies. This is shown in panel (a) of Figure 1. While EIPL recovered to 60% or higher in many counties in the South and Central North, EIPL in counties in the North and Mid-Atlantic and the West remained stuck in the 0% to 35% range.[8]

Panel (b) further illustrates the regional disparities in EIPL by reporting the top 10 and bottom 10 cities in terms of average EIPL from September 2020 through May 2021 among the 50 biggest U.S. cities by population. In cities in Florida such as Jacksonville, Tampa or Orlando, EIPL averaged over 75%, whereas in cities in California, Oregon and Washington such as Los Angeles, Portland or Seattle, EIPL averaged 20% or less.

While these regional disparities in EIPL are striking, there are also large differences in EIPL within counties. For instance, the within-county interquartile range of EIPL averages 16% for the 2020-21 school year, and the extent of this dispersion is similar across counties with different levels of average EIPL. This suggests that disparities in EIPL for the 2020-21 school year not only reflect systematic regional differences, but may also be driven by school and local characteristics that apply similarly across the country.

5 What accounts for the disparities in EIPL?

We begin by analyzing how various observable school and local characteristics surrounding the school correlate with EIPL. Then we return to geography and examine the extent to which these correlations are driven by systematic regional differences. Naturally, these correlations should not be interpreted as causal. The objective is merely to document a set of stylized facts that can help us understand the "for

of visit changes is evenly distributed across CBSAs.

[8]The appendix provides additional evidence on the temporal disparities across U.S. Census divisions.

whom" in order to quantify the consequences of pandemic-induced learning losses for different segments of the student population and formulate appropriate policies going forward.

To save on space, we refer the interested reader to the appendix for details on the data used in this section and robustness checks.

5.1 School type and grade

Perhaps the most obvious observable school characteristics are type (public non-charter, public charter, private non-religious, or private religious school) and grade (elementary school, middle school, high school, or a combination thereof) as designated by the NCES. We use this information to group individual schools accordingly and compute average student-weighted EIPL for each group.

From March to May 2020, all school type and grade combinations averaged less than 20% of EIPL. For the 2020-21 school year, however, there are substantial differences. EIPL is lowest for public charter schools (averaging 36%), followed by public non-charter schools (averaging 44%), private non-religious schools (averaging 51%), and private religious schools (averaging 57%). In turn, EIPL is lower for middle and high-schools (averaging 39%) than for elementary schools (averaging 56%), and these differences are more pronounced for public schools than for private schools.

The EIPL ranking by school type, which is robust to controlling for other school characteristics, may come as a surprise for two reasons. First, public charter schools are typically independent and not unionized whereas public non-charter schools belong to school districts that, for some urban areas, are comprised of several hundred schools and often unionized. One could have expected that being independent and non-unionized would have made it easier for charter schools to reopen to in-person learning. Second, according to Hanson [2021], tuition for non-religious private schools is on average more than twice as high as tuition for religious private schools. The additional resources and resulting smaller class sizes could have made it easier for non-religious private schools to reopen to in-person learning. Yet, in both cases, exactly the opposite occurred.

5.2 Local affluence, education, family structure, and student race

Next we consider EIPL by local affluence, education and family structure, which are prominent inputs for models of human capital accumulation (e.g. Cunha et al., 2010), as well as student race. We proxy local affluence by average household income, education by the share of households with a college degree or higher, and family structure by the share double-headed households with children, all measured at the zip-code level of the school and based on 2016-2019 estimates from the American Community Survey (ACS). For race, we use the school's share of non-white students as provided by the NCES. Results are robust to using the variables at the census block group or tract of the school, or at the school district level.

We have also considered several alternative socio-economic indicators describing the neighborhood of the school, including Chetty et al. [2020]'s indicators of Income Mobility and many of the other variables available from their Opportunity Atlas database. While these variables are correlated with EIPL, they are also highly correlated with the above measures of local affluence and education and do not add significant explanatory power to the below regressions.

Figure 2 reports nonparametric binned scatterplots of the unconditional relationship between average school EIPL from September 2020 to May 2021 and the different variables. As shown in panel (a), there is a striking inverse relationship between EIPL and household income: schools in zip-codes with *high* household income provided on average *lower* EIPL. Panel (b) shows a similar result for education: schools in zip-codes with a *high* share of college-educated people provided on average *lower* EIPL. Consistent with the above results, EIPL is on average about 10% higher for private schools than for public schools but the relationship of EIPL with household income and education is otherwise very similar.

Panel (c) shows that there is positive relationship between EIPL and local share of double-headed households for public schools but no systematic relationship for private schools. As we will see below, the positive relationship for public schools changes once we control for other observable school characteristics.

Finally, as shown in panel (d), there is also a striking inverse relationship between EIPL and the share of non-white students. For schools with close to 0% of non-white students, EIPL averaged over 60%, independent of whether the school is public or private. For schools with close to 100% of non-white students, in contrast, EIPL averaged only about 35% for public schools and just below 50% for private schools. Interestingly, the gap between public and private schools with respect to race widens noticeably only for schools with more than one third of non-white students.

Given the general association of poverty with race, the inverse relationship of EIPL with both affluence and race may come as a surprise. But as shown in the appendix, the share of a school's non-white students is essentially uncorrelated with local affluence, education, and parental structure. To illustrate this result, we estimate OLS regressions of average school EIPL between September 2020 and May 2021 on the different measures. Here and below, all estimates are weighted by the school-specific sampling weights to ensure representativeness; standard errors are clustered at the county level; and the coefficients are scaled so that they show the implied change in EIPL of going from the 25th to the 75th percentile of the distribution of a variable.[9]

Figure 3 reports the results. To save on space, we show only results for public schools, but the results for private schools are very similar. The brown round-shaped plots show the point estimates and 95% confidence intervals from regressing EIPL separately on zip-level household income, education, and share of double-headed households together with the share of non-white students and controls for school type and school grade.[10] The yellow and red plots are discussed further below.

All of the estimates are statistically significant and quantitatively important. Most remarkably, even after conditioning on local income, education, and parental structure, EIPL for a school with a student body at the 75th percentile of the non-white distribution was on average 15-22% lower during the 2020-21 school year than for a school at the 25th percentile of the distribution. Further analysis shows that this negative relationship is in large part driven by the share of hispanic students and less by the share of black students. In turn, EIPL for a school located in a zip-code at the 75th percentile of the income and education distribution was on average 4-6%, respectively 6-7% lower than for a school at the 25th percentile. Finally, EIPL is also significantly negatively related to the share of double-headed households with children, due to the fact that schools in zip codes with a larger share double-headed households have

[9] More precisely, all right-hand side variables are expressed as deviations from the mean, normalized by the interquartile range. The appendix provides summary statistics.

[10] The appendix provides the full regression results for all variables. Since local income and education are highly correlated with each other, regressing EIPL on both of them jointly would not noticeably increase the explanatory power but would make it more difficult to interpret the estimates.

on average a smaller share of non-white students. This highlights the importance of analyzing the relation between EIPL and different school characteristics in a multivariate setting.

5.3 Public school tests scores, school size, and school funding

We extend the analysis by considering the relation between EIPL and indicators of school quality, school size, and school funding. Following Chetty et al. [2014], we proxy school quality with pre-pandemic average tests scores, based on data from Fahle et al. [2021].[11] For size, we use student enrollment of the school, available from NCES. For school funding, we consider both pre-pandemic school spending per student obtained from EdunomicsLab [2021] as well as district-level ESSER funding by student compiled by Malkus [2021b]

As above, we estimate OLS regressions of EIPL on local income, education, parental structure, and student race and then add the different variables.[12] As shown by the yellow diamond-shaped coefficient plots in Figure 3, EIPL is estimated to be higher for schools in districts with higher pre-pandemic test scores, but lower for larger schools. Interestingly, EIPL is also inversely related to school spending and ESSER funding per student, although not significantly so in the latter case. Note that these estimates control for local income, education, parental structure, and race, suggesting that school quality, school size, and school funding are independent predictors of EIPL.

The negative association between EIPL and school spending may have been expected given the positive correlation between school spending and local affluence. The absence of a positive association of EIPL with ESSER funding per student, in turn, is remarkable because ESSER, which was appropriated by Congress in three waves totaling $190 billion or almost five times the annual federal K-12 spending prior to the pandemic, was advertised primarily as support for schools to reopen to in-person learning. We return to this point below.

Also note that controlling for district test scores, school size, and school funding barely changes the negative relation of EIPL with local affluence, education, and parental structure, but reduces the effect of share of non-white students by roughly one third. This is because the share of non-white students is negatively correlated with district test scores and positively correlated with school size. Even so, the share of non-white students remains a strong predictor of lower EIPL.

5.4 Geography

For the last part of the analysis, we return to geography and ask how much of the relation of EIPL with school and local characteristics is driven by systematic regional differences that are not directly related to the school. From Figure 1 we know that schools with higher EIPL are generally located in the central and southern parts of the U.S. – regions that in general were more favorable towards reopening the economy despite potential health risks and at the same time have seen lower COVID vaccination rates. To assess the relative importance of these factors for EIPL, we re-estimate the above regressions and add party affiliation of the state governor and voter preferences in the 2020 presidential election as proxies for the

[11] We use district-level average test scores for 2018-19, which are available for almost all districts. We also used school-level test scores, which are available for only a subset of schools, and the results are very similar.

[12] As above, we include local income, education, and parental structure one-by-one together with race and the other variables. Since the estimates for race and the other variables barely change across regressions, we report these estimates controlling for local income, education, and parental structure jointly.

general stance towards reopening the economy as well as two-week lagged vaccination rates. Furthermore, we add state-level teacher unionization rates, a comparable wage index for PK-12 educators in the local labor market, controls for a county's COVID health situation as measured by pre-pandemic ICU bed capacity, two-weeks lagged COVID case and death rates and various non-pharmaceutical interventions (NPIs), as well maximum weekly temperature and indicators of urban density.

The red round-shaped coefficient plots in Figure 3 report the results. First, both the presence of a Republican governor and a large share of Republican votes in the 2020 presidential election are strong predictors of higher EIPL during the 2020-21 school year, but so are higher vaccination rates. Schools in states with higher teacher unionization rates and a higher comparable wage index, in turn, provided on average lower EIPL.[13] These results are all the more remarkable because they arise in a multivariate context; i.e. while controlling for all the other variables.

Second, the addition of the different regional attributes – in particular voter preferences – increases the predictive power of the regression substantially while reducing the relationship of EIPL with local affluence, education, and parental structure to almost zero. In other words, the negative predictive effect of local affluence, education, and parental structure on EIPL in the above regressions arose primarily because these variables proxied for voter preferences and to a lesser extent vaccination rates, teacher unionization rates, and the local wage comparison index for teachers.

Third, the inverse relation between EIPL and the share of non-white students is also reduced. Yet, even after controlling for all the state and county-specific attributes, a school with a student body at the 75th percentile of the non-white distribution is predicted to average 3-7% lower EIPL during the 2020-21 school year than a school at the 25th percentile of the distribution.

Fourth, the negative coefficient estimate on district test scores and school size remain largely unaffected, indicating that even after controlling all the other school characteristics and regional attributes, school quality and size played an important role for EIPL. By contrast, the regional variables completely absorb the negative association of EIPL with pre-pandemic school spending per student, indicating that school spending did not play a decisive role for EIPL overall but instead picked up systematic differences across regions.[14] Finally, the inverse relation between EIPL and ESSER funding per student becomes somewhat more negative. Hence, even within counties, schools in districts with more ESSER funding per student provided on average not more but slightly less EIPL.

6 Conclusion

Using a new empirical approach that combines comprehensive data on school visits with two school learning mode trackers, we construct a measure of effective in-person learning (EIPL) during the pandemic for both public and private schools. We document large differences in EIPL over time, across regions, as well as with respect to school and local characteristics.

While the results from our regression analysis should not be interpreted as causal, they nevertheless reveal interesting patterns that raise critical questions on why EIPL in some areas and for some students

[13] In contrast, the variables measureing the local COVID health situation and NPIs are not systematically associated with EIPL. Similarly, the predictive content of the urban density indicators is small.

[14] The main reason why school spending per student no longer plays an important role is the addition of the two voter preference variables and not the addition of the local wage comparison index for teachers.

was so much lower during the 2020-21 school year. We highlight three questions in particular:

1. Why did schools in more affluent and more educated areas with higher funding per student provide less EIPL? The simple answer is that this inverse relationship is in large part about political preferences. But why would more democratic-leaning areas have been more reluctant to let students return to in-person learning? One potential explanation is that independent of political preferences, more affluent and educated parents were on average more likely to be able to work from home and therefore considered the cost of supervising students' virtual learning from home (either in person or by hiring help) more manageable. Another potential explanation is that more affluent and educated parents had a different perception of the risk of sending students back to in-person school, for instance due to different news and social-media exposure. Both of these explanations contrast, however, with the observation that even within counties, private schools (which generally attract students from more affluent and educated parents) provided more EIPL than public schools. No matter the explanation, it remains that students in more affluent and more educated areas of the U.S. received on average less EIPL.

2. Why did schools with a higher share of non-white students provide less EIPL, even within a given county and controlling for neighborhood poverty and other school characteristics? This striking result defies a simple explanation and yet seems key given the large and persistent educational achievement gaps between students of different races that existed already before the pandemic.

3. Why did schools in districts with more ESSER funding per student not provide more EIPL? One possible reaction is that without ESSER funding, schools would have been closed for even longer. Yet, the absence of a positive relationship arises even within counties and despite controlling for many other school characteristics, which makes this an unlikely explanation. Another potential explanation is that Congress imposed few constraints on how ESSER funding could be used, and according to estimates by Malkus [2021a], less than 20% had been spent by August 2021. If these funds were spent primarily to improve students' remote learning capacities (e.g. providing students with computers and wireless connections) instead of upgrades to the school buildings and personal protection equipment, then ESSER funding would have primarily facilitated remote learning instead of a return to in-person learning; i.e. its main advertised purpose.[15]

Exploring these questions goes beyond the scope of the paper but they are clearly important to understand the causes and consequences of school closings during the pandemic.

[15]The remaining unspent ESSER funds could still be put to good use, for instance by providing summer learning programs for students, especially since ESSER funds were allocated to school districts according to pre-pandemic Title I spending and thus benefited disproportionally schools with students from poorer backgrounds.

References

Francesco Agostinelli, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *NBER Working Paper 28264*, 2020.

Dena Bravata, Jonathan H Cantor, Neeraj Sood, and Christopher M Whaley. Back to school: The effect of school visits during COVID-19 on COVID-19 transmission. *NBER Working Paper 28645*, 2021.

Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf. The association of opening K-12 schools and colleges with the spread of COVID-19 in the United States: County-level panel data analysis. *Proceedings of the National Academy of Sciences*, 118(42), 2021a.

Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *Journal of Econometrics*, 220(1):23–62, 2021b.

Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, 129(4):1553–1662, 2014.

Raj Chetty, John N Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya S. Porter. The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. *NBER Working Paper 25147*, 2020.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931, 2010.

MIT Election Data and Science Lab. County Presidential Election Returns 2000-2020, 2018. URL https://doi.org/10.7910/DVN/VOQCHQ.

Thomas Dee, Elizabeth Huffaker, Cheryl Phillips, and Eric Sagara. The Revealed Preferences for School Reopening: Evidence from Public-School Disenrollment. *NBER Working Paper 29156*, August 2021.

Emma Dorn, Bryan Hancock, Jimmy Sarakatsannis, and Ellen Viruleg. Covid-19 and education: The lingering effects of unfinished learning. Technical report, McKinsey & Company, 2021.

EdunomicsLab. National education resource database on schools, November 2021. URL www.edunomics.org/nerds.

Zeynep Ertem, Elissa M Schechter-Perkins, Emily Oster, Polly van den Berg, Isabella Epshtein, Nathorn Chaiyakunapruk, Fernando A Wilson, Eli Perencevich, Warren BP Pettey, Westyn Branch-Elliman, et al. The impact of school opening model on SARS-CoV-2 community incidence and mortality. *Nature Medicine*, pages 1–7, 2021.

Erin M Fahle, Benjamin R Shear, Demetra Kalogrides, Sean F Reardon, Richard DiSalvo, and Andrew D Ho. Stanford Education Data Archive (Version 4.1), 2021. URL http://purl.stanford.edu/db586ns4974.

Nicola Fuchs-Schündeln, Dirk Krueger, Andre Kurmann, Etienne Lale, Alexander Ludwig, and Irina Popova. The fiscal and welfare effects of policy responses to the COVID-19 school closures. *NBER Working Paper 29398*, October 2021.

Clare Halloran, Rebecca Jack, James Okun, and Emily Oster. Pandemic Schooling Mode and Student Test Scores: Evidence from US States. *NBER Working Paper 29497*, November 2021.

Melanie Hanson. Average Cost of Private School. Technical report, EducationData.org, 2021. URL https://educationdata.org/average-cost-of-private-school.

Youngsoo Jang and Minchul Yum. Aggregate and intergenerational implications of school closures: A quantitative assessment. *Covid Economics, Vetted and Real-Time Papers*, 57:46–93, 2020.

Vladmir Kogan and Stéphane Lavertu. How the COVID-19 pandemic affected student learning in Ohio: Analysis of Spring 2021 Ohio state tests. Technical report, 2021.

Karyn Lewis, Megan Kuhfeld, Erik Ruzek, and Andrew McEachin. Learning during COVID-19: Reading and math achievement in the 2020-21 school year. Technical report, NWEA Center for School and Student Progress, 2021.

Nat Malkus. How much of federal COVID-19 relief funding for schools will go to COVID-19 relief? Technical report, American Enterprise Institute, 2021a.

Nat Malkus. Federal COVID Elementary and Secondary School Emergency Relief funding district-level data compilation. Technical report, Return to Learn Tracker, American Enterprise Institute, 2021b. URL http://www.returntolearntracker.net/esser/.

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, 2021.

Figures and Tables

Table 1: Mapping school visits to Effective In-Person Learning

(a) Source of regression coefficients to map school visits to EIPL

	Burbio (CBSA)	R2L (CBSA)	Burbio (State)	R2L (State)
Number of schools	19,798	16,486	15,715	17,337
Percent of schools	28.6	23.8	22.7	25.0

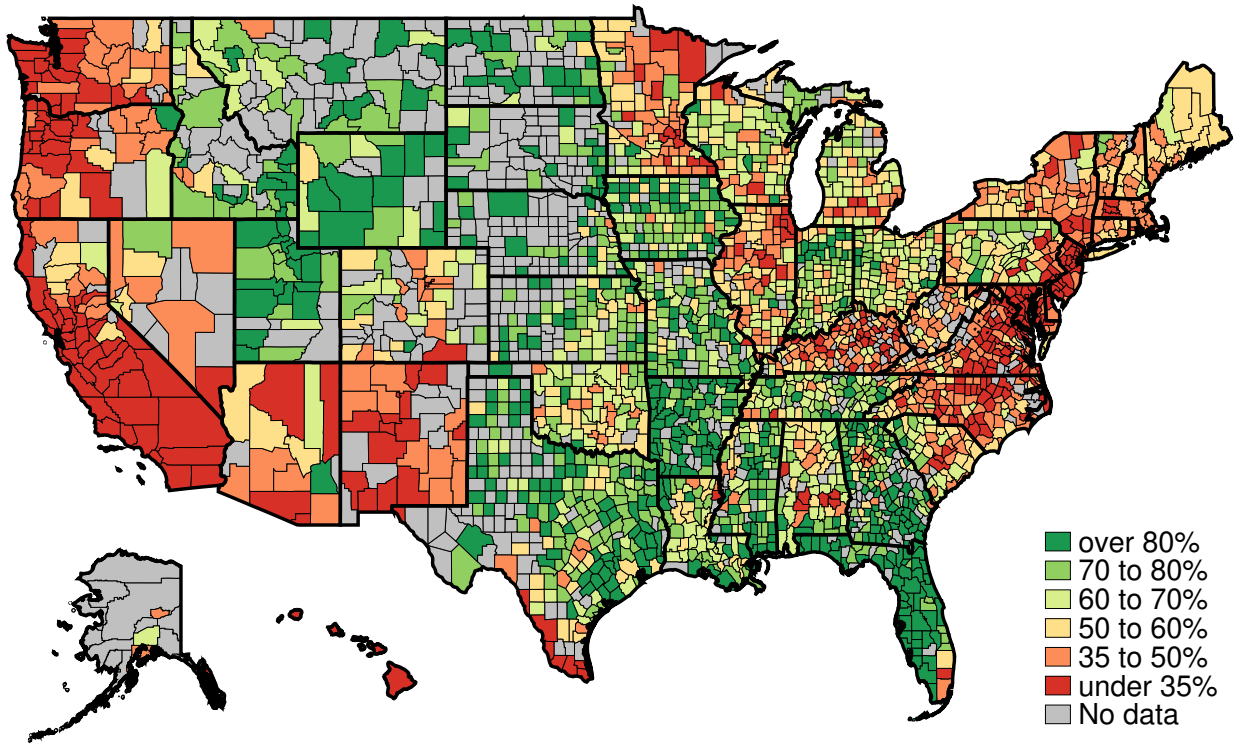
(b) Distribution of regression coefficients to map school visits to EIPL

	Mean	Percentile				
		5th	25th	50th	75th	95th
$\hat{\beta}$	1.22	0.79	1.16	1.21	1.31	1.47
$\hat{\gamma}$	0.31	0.04	0.20	0.30	0.40	0.71
R squared	0.81	0.51	0.69	0.86	0.96	0.99

Notes: Panel (a) shows the distribution of schools by type of regression coefficient retained for the OLS estimation of (4). Panel (b) shows the distribution of retained regression coefficients and R-squared, weighted by the different school weights.

Figure 1: Regional Disparities in Effective In-Person Learning

(a) Effective in-person learning across U.S. counties during the 2020-21 school year

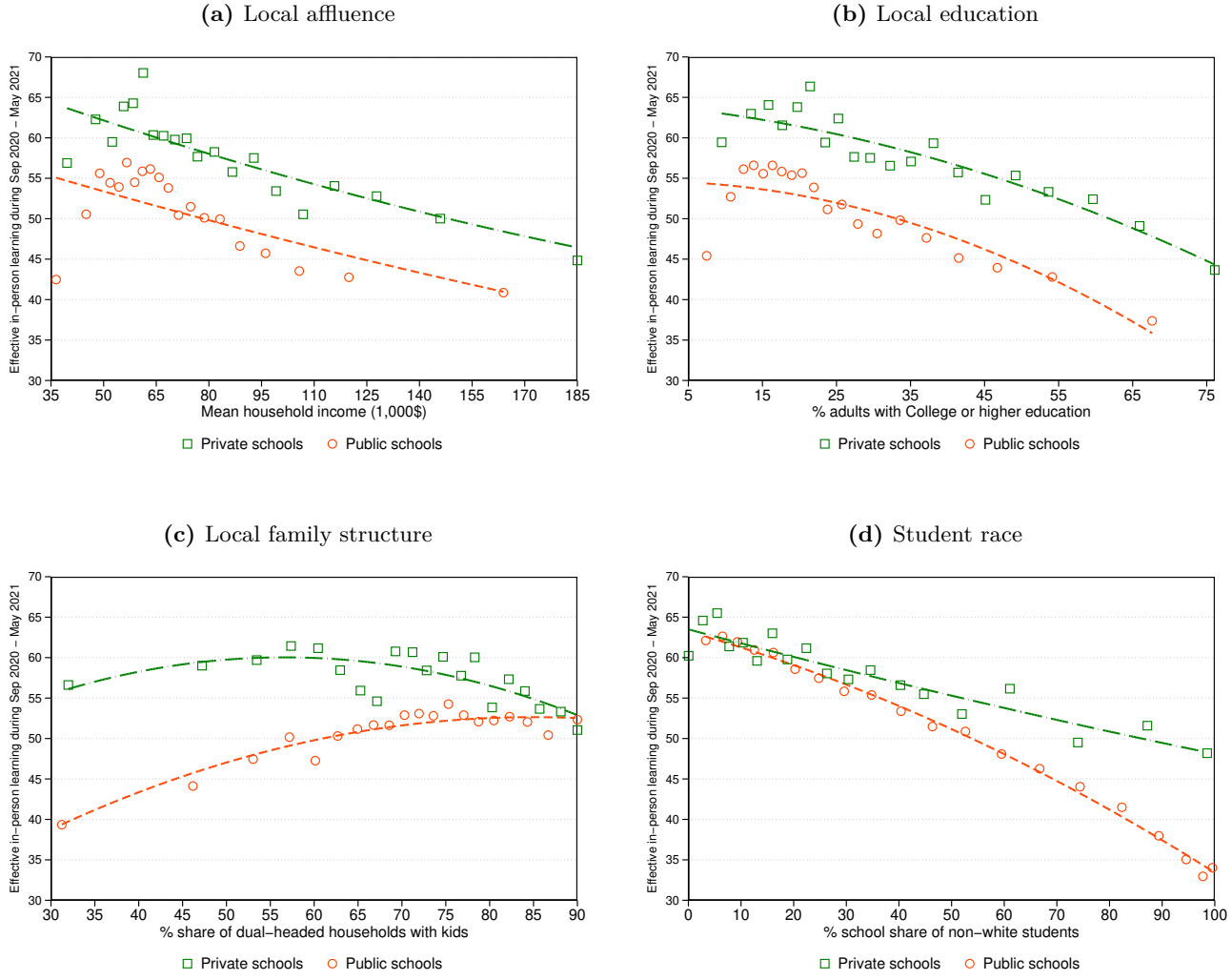


(b) The top 10 and bottom 10 U.S. cities in terms of effective in-person learning

Rank	CBSA name	EIPL	Rank	CBSA name	EIPL
1	Jacksonville, FL	87.6%	41	Sacramento-Arden-Arcade-Roseville, CA	22.9%
2	Tampa-St. Petersburg-Clearwater, FL	81.3%	42	Washington-Arlington-Alexandria DC-VA	22.9%
3	Orlando, FL	77.1%	43	Baltimore-Towson, MD	22.4%
4	Houston-Baytown-Sugar Land, TX	61.9%	44	Seattle-Bellevue-Everett, WA	20.3%
5	Fort Worth-Arlington, TX	61.8%	45	Portland-Vancouver-Beaverton, OR-WA	20.2%
6	Cincinnati-Middletown, OH-KY-IN	61.3%	46	San Jose-Sunnyvale-Santa Clara, CA	17.0%
7	Dallas-Plano-Irving, TX	60.7%	47	Las Vegas-Paradise, NV	16.5%
8	Detroit-Livonia-Dearborn, MI	58.3%	48	Los Angeles-Long Beach-Santa Ana, CA	16.4%
9	Nassau-Suffolk, NY	57.6%	49	Oakland-Fremont-Hayward, CA	15.9%
10	Nashville-Davidson--Murfreesboro, TN	57.4%	50	Riverside-San Bernardino-Ontario, CA	14.5%

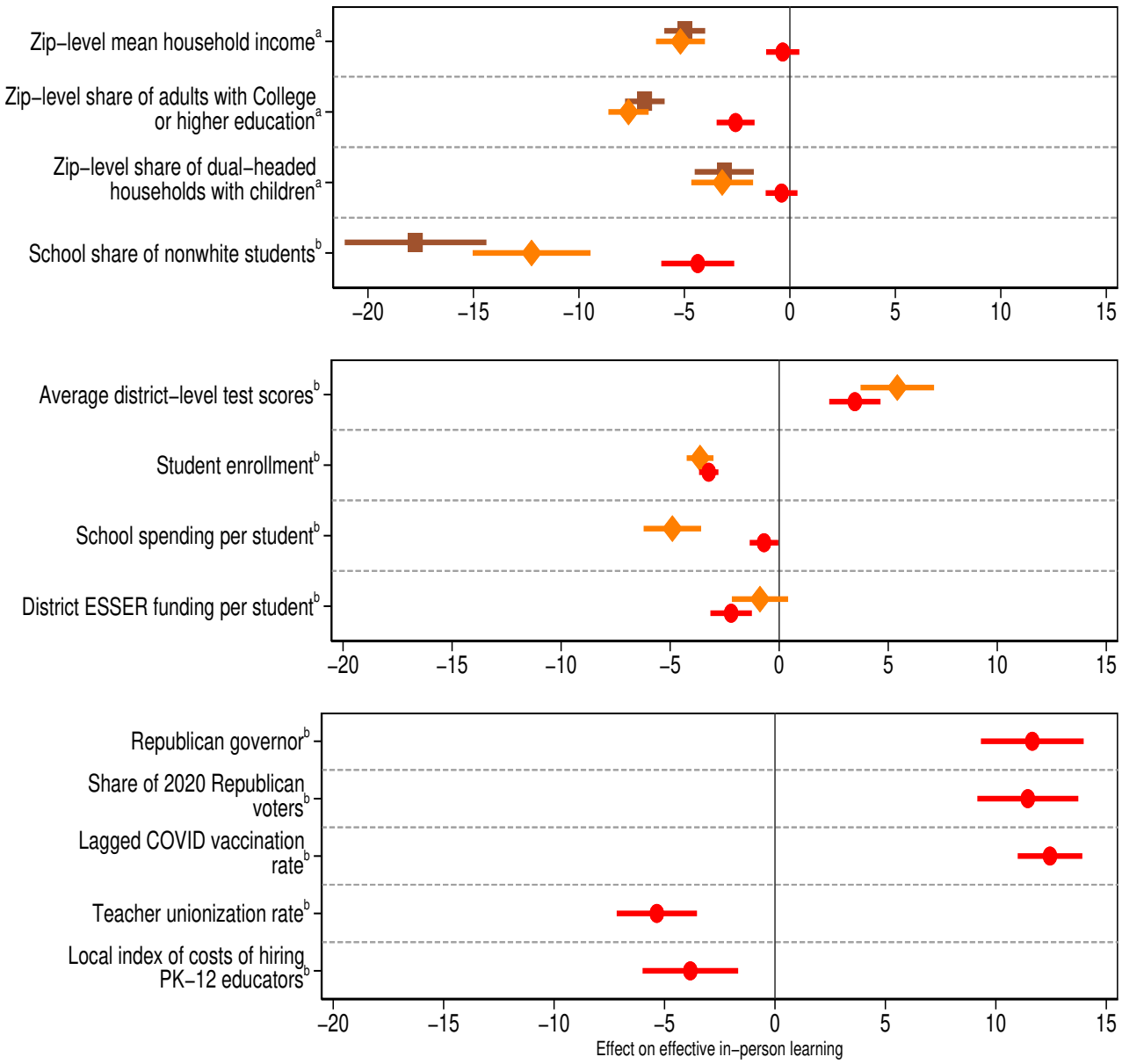
Notes: The top panel shows the student-weighted average county EIPL from September 2020 to May 2021 by different percentile ranges for all counties for which we have reliable data on at least three schools. The bottom panel shows the top-10 and bottom-10 Core-Based Statistical Areas (CBSAs) in terms of average EIPL among the 50 largest CBSAs by population. EIPL for each CBSA is computed as the student-weighted average across schools with reliable data.

Figure 2: Effective In-Person Learning by Local Affluence, Education, Family Structure, and Race



Notes: The figures show binned scatterplots of average EIPL from September 2020 to May 2021 for public schools and private schools, respectively, by (a) zip-code average household income, (b) zip-code average share of household with a college degree or higher, (c) zip-code share of dual-headed households, and (d) school share of non-white students. Observations are weighted by the school-specific sampling weight described in the appendix.

Figure 3: The Relationship of Effective In-Person Learning with School and Local Characteristics



Notes: The figure shows the estimated effects on EIPL from weighted OLS regressions with standard errors clustered at the county level in parentheses and school weights calculated as explained in the appendix. The sample consists of approximately 55,000 public schools. The regressions are estimated for weekly school EIPL from September 2020 to May 2021. All estimates except for the "Republican governor" dummy are scaled so that they show the implied change in EIPL of going from the 25th percentile to the 75th percentile of the distribution of a variable. The estimates for the first three variables, denoted by [a], are the result of separate regressions for each of the three variables in combination with the other variables listed below. The coefficient estimates for the other variables denoted by [b] are the result of regressions where all the variables are included jointly. The brown square-shaped estimates show the effects of regressing EIPL on the variables in the top box only. The yellow diamond-shaped estimates show the effects of regressing EIPL on the variables in the top and middle box. The red round-shaped estimates show the effects of regressing EIPL on the variables in all three boxes. All regressions control for school type (charter vs. non-charter school) and school grade (elementary vs. middle vs. high. vs. combined school). In addition, the regressions for the red round-shaped estimates control for pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, dummies for various non-pharmaceutical interventions, maximum weekly temperature in the county, county population density, and a dummies for rural-urban continuum codes. See the appendix for details.

Online Appendix for: School Closures and Effective In-Person Learning during COVID-19: When, Where, and for Whom*

André Kurmann
Drexel University

Etienne Lalé
Université du Québec à Montréal

December 21, 2021

Contents

*Contact: Kurmann: Drexel University, LeBow College of Business, School of Economics, 3220 Market Street, Philadelphia, PA 19104 (email: kurmann.andre@drexel.edu); Lalé: Université du Québec à Montréal, Department of Economics, C.P. 8888, Succ. centre ville, Montréal (QC), H3C 3P8 Canada (email: lale.etienne@uqam.ca). Aseni Ariyaratne provided excellent research assistance. We thank Dennis Roche from Burbio, Nat Malkus from Return to Learn, and Safegraph for generously sharing their data. We also thank seminar participants at the 2nd joint IZA & Jacobs Center Workshop on "Consequences of COVID-19 for child and youth development" for their comments. All errors are our own.

A School data appendix

This section presents additional information about the NCES data (Subsection A.1); the Safegraph school visits data, our algorithm to match it to NCES datasets, and to construct school weights for the matched data (Subsection A.2); additional information about the Burbio and Return2Learn learning mode data (Subsection A.3); details on the construction of the EIPL measure (Subsection A.4); and a description of the other data used in the regressions (Subsection A.5).

A.1 NCES data

The U.S. Department of Education's NCES is the primary federal entity for collecting and analyzing data related to education. The NCES regularly publishes statistics on both public and private schools and also makes available different datasets on individual schools. We mainly make use of two NCES dataset.

The first one is the Common Core of Data (CCD; see https://nces.ed.gov/ccd/). CCD is a comprehensive annual database of all public elementary and secondary schools and school districts (including public charter schools). The CCD consists of five surveys completed annually by state education departments from their administrative records. The information includes a general description of schools and school districts, including name, address, and phone number; number of students and staff, demographics (including the gender and racial makeup of the schools students); and fiscal data, including revenues and current expenditures. We use the 2019-2020 CCD school data files released in March 2021.

The second dataset is the Private School Universe Survey (PSS; see https://nces.ed.gov/surveys/pss/), which is a biennial survey that collects data on private schools and serves as a sampling frame for other NCES surveys of private schools. The PSS data include a general description of schools, teachers, and students (including the gender and racial makeup of the schools students) in the survey universe. The schools surveyed in the PSS come with a survey weight. We use the 2017-2018 data files released in August 2019 (there is no more recent version of these data as of this writing).

We complement the NCES datasets with information from the Education Demographic and Geographic Estimates (EDGE; see https://nces.ed.gov/programs/edge/). EDGE is a program run by the NCES to create and assign address geocodes (estimated latitude/longitude values) and other geographic indicators to public schools, public local education agencies, private schools, and post-secondary schools, and create area-type indicators (City, Suburban, Town, and Rural). We use the 2019-2020 geocodes to improve the reliability of the match between the CCD/PSS files and Safegraph data (see Subsection A.2), and the area-type indicators to assess the relation between EIPL and local school characteristics (see Figure C5).

A.2 Safegraph data, matching algorithm, school weights

A.2.1 Safegraph data Safegraph is a data company that aggregates anonymized location data from cell phone applications in order to provide insights about foot traffic (visits) to physical places, called Places of Interest (POI). Each POI in Safegraph's data is identified by a unique persistent `safegraph_place_id`. Details of the spatial hierarchy of POIs are important to understand visit attribution.[1] A POI is a polygon, and some of the polygons are encompassed into larger polygons. When it so happens, the "child" polygon receives a `parent_safegraph_place_id` equal to the `safegraph_place_id` of the encompassing "parent" POI. Except for a handful of POIs (about 1% of the universe of Safegraph's POIs), each `safegraph_place_id` comes with a 6-digit industry NAICS code.[2] About 80-85% of Safegraph's POIs come with information on visits. In our analysis of POIs with NAICS 611110 ("Elementary and Secondary Schools"), about 5% have a `parent_safegraph_place_id`, which is almost always shared with

[1] See Safegraph's Places Manual as well as this blog post for additional details.

[2] See https://docs.safegraph.com/docs/core-places#section-naics-code-top-category-sub-category for information on Safegraph's algorithm for attributing NAICS codes to the POIs covered by the Core places dataset.

a POI that is classified as NAICS 624410 ("Child and Youth Services") or NAICS 813110 ("Religious organizations"). To reduce noise in visits data, we aggregate up these visits and attribute them to the school that is paired to these non-611110 NAICS POIs.

A.2.2 Matching of Safegraph POIs with NCES school records Our algorithm to match the Safegraph dataset of elementary and secondary schools to the NCES's CCD and PSS files works as follows:

1. Prior to matching schools data to Safegraph, we deduplicate and pre-treat the Safegraph data by cleaning POIs' names and addresses. For names, we convert the capital letters to lower case and remove all the "%", "&", etc., numbers (if any), and spaces from the raw Safegraph location names. More importantly, we replace abbreviated school information in the Safegraph names by a complete descriptor using the following rules:[3]

Portion of the raw Safegraph name:	**Recoded as:**
elemsch	elementaryschool
highsch	highschool
kindergsch	kindergarten
middlesch	middleschool
primarysch	primarychool
schoolthe	school

   Last, we clean schools' addresses by using Stata's `stnd_address` command to standardize street address names.

2. In the CCD files, we have information on school names and addresses that describe the physical location of schools (street address and postal code). Firstly, we clean school names by converting the capital letters to lower case and removing all the "%", "&", etc., numbers (if any), and spaces, and standardize street addresses in a format similar to that applied to Safegraph data. Then, we match files by attempting in the following order: (i) a direct merge on name/address/zip-code, (ii) a direct merge on name/zip-code, then within each 5-digit zip codes: (iii) a fuzzy name match on name/address, (iv) a fuzzy name match on name, (v) a fuzzy name match on address. For the fuzzy name matching, we use Stata's `reclink2` command and retain only those with a matching score higher than 0.85. We manually compare a random sample of the matched schools obtained through steps (iii) to (v) of the algorithm to confirm that a threshold of 0.85 provides us with good enough matches.[4]

3. In the PSS files, we have information on school names and GPS coordinates. Firstly, we clean school names by converting the capital letters to lower case and removing all the "%", "&", etc., numbers (if any), and spaces. We then pool together all PSS and Safegraph schools that belong to the same geographic area (defined by GPS coordinates rounded to the first decimal place), and within the area we match each PSS school to the closest Safegraph school by measuring distance as (i) the geographic distance based on GPS coordinates and (ii) the Levenshtein distance between school names (normalized by the length of the longest string of school name). We retain only those matches where the geographic distance is less than 250 meters or the string distance is under 0.250.

[3] As an example, consider the Safegraph location called "Big Spring Lake Kinderg Sch". After removing the spaces and converting the capital letters to lower case, we obtain "bigspringlakekindergsch". We then rename it as: "bigspringlakekindergarten". This enables us to increase the quality of the match to NCES data where typically the word "Kindergarten" is not abbreviated.

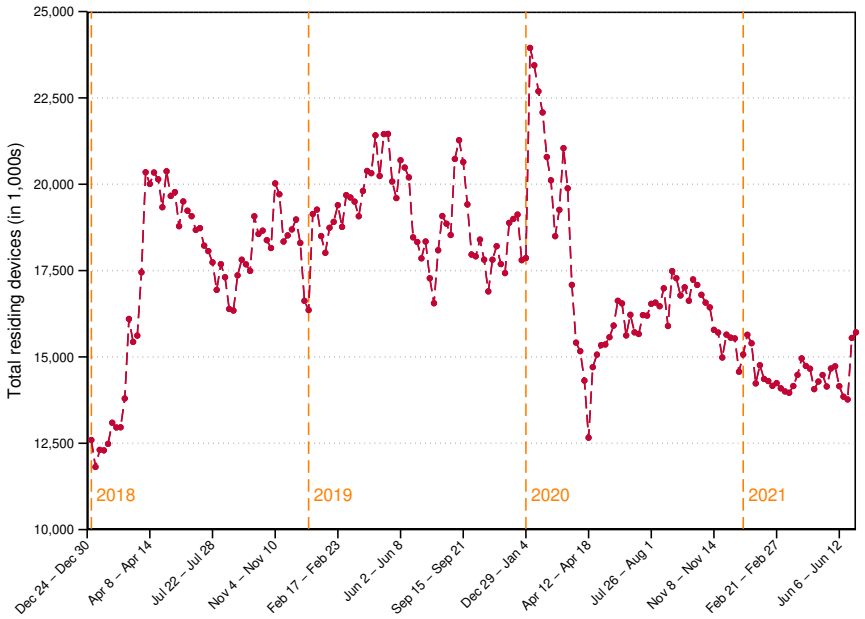
[4] Consider Safegraph's "Big Spring Lake Kinderg Sch" described in Footnote 3. The name of this school in the CCD file is "Albertville Kindergarten and PreK". Through our algorithm, we obtain a fuzzy match at the name/address level (within the same 5-digit zip code) because the street addresses in Safegraph and in the CCD file turn out to be exactly the same. This, together with our update of the Safegraph's school name, yields a matching score of 0.92 according to `reclink2` standard score metric.

We manually compare a random sample of the matched schools to confirm that these thresholds provide us with good enough matches.

Through this algorithm, we obtain high-quality matches for 102,774 schools (85,446 public schools and 17,328 schools).

A.2.3 Normalization and sample selection An important concern when working with Safegraph's data is that changes in visits counts over time can be driven by changes in the sample of cell phone devices that Safegraph uses. Figure A1, which plots the total number of residing Safegraph devices (in counties that contain POIs with NAICS code 611110) over the 2018-2021 period, illustrates the magnitude of these changes.[5] As can be seen, following large variations in the first two quarters of 2018, the sample

Figure A1: Safegraph: Number of residing cell phone devices



Notes: The figures show the sum of Safegraph cell phone devices across all counties that contain POIs with NAICS code 611110 ("Elementary and secondary schools").

size expands until mid-2019, then drops during the second half of 2019 and expands again in January of 2020. More importantly, the sample sizes drops substantially at the beginning of the pandemic and never recovers afterwards; in 2021 the sample size actually decreases relative to the second half of 2020.

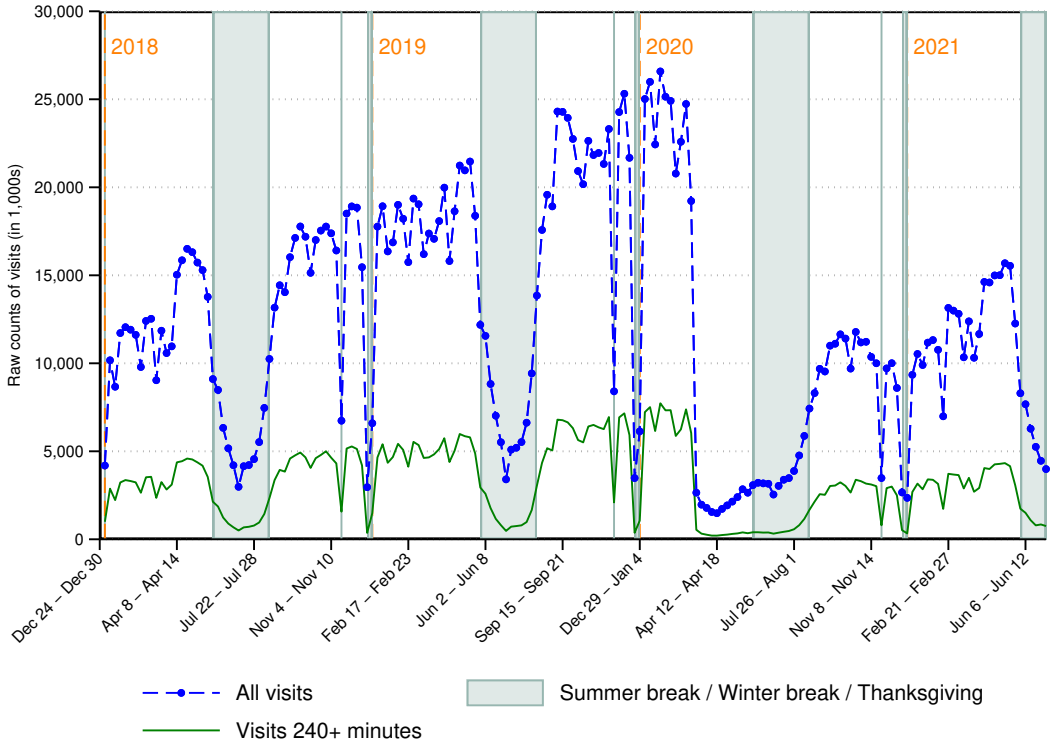
Figure A2 illustrates the impact of these variations on counts of visits to all Safegraph's POIs with NAICS code 611110. In the upper panel, there is a clear upward trend in raw visits throughout 2018, 2019, and early 2020, as well as an incomplete recovery of visits in 2021 relative to pre-pandemic levels of visits. The bottom panel shows that normalizing by county-level counts of cell phone devices removes the trend in 2018 and 2020, while inducing visits at the end of 2019 and at the beginning of 2020 to be higher than before the Summer of 2019. The effects of normalization is also important for the recovery in 2021: normalized school visits return to their pre-pandemic levels, whereas in the not normalized data they remain about 25% lower. Motivated by these observations, throughout our analysis we normalize school visits with the weekly county-level counts of Safegraph cell phone devices.

In an effort to reduce measurement error, we implement the following sample restrictions. First, we drop schools where the raw visits count on average during the base period is less than 10, and schools

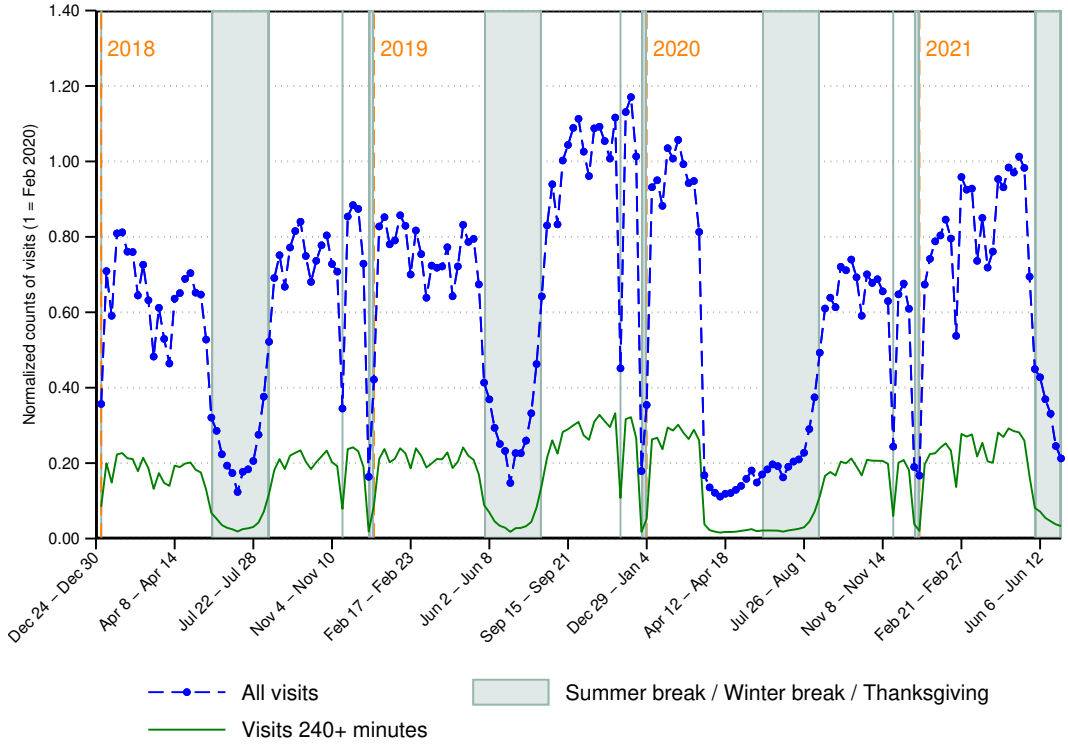
[5] Residing devices are cell phones with a primary nighttime location. Counts of residing devices are released at the CBG level. These counts are typically used for the purpose of normalizing visits data at a granular level.

Figure A2: Safegraph: Aggregate time series of school visits

(a) Raw counts of visits



(b) Normalized counts of visits



Notes: The figures show the raw (upper panel) and normalized (lower panel) counts of total weekly visits and counts of visits longer than 240 minutes to all Safegraph POI with NAICS code 611110 ("Elementary and secondary schools").

where $\Delta\widetilde{v}_{j,t}$ is larger than 50 more than once during the based period. The goal of these first two restrictions is to ensure that the measurement of school visits for the base period are reliable enough to compare them with school visits in any other period. Together these restrictions reduce the sample size by 20%. Then, we drop schools where $\Delta\widetilde{v}_{j,t}$ is larger than 75 more than once, either during the period from beginning of September 2019 to November 2019 or the period from beginning of September 2020 to the end of the sample period. This procedure intends to purge the data from extreme values that affect the average of changes in visits in any given period. We use a larger threshold (75 instead of 50) to trim the data because it is to expected that the visits time series for each school are more volatile outside of the November 2019 to February 2020 period. This sample restriction reduces the sample size by an additional 10%. The resulting "in-scope" dataset contains 69,910 schools or about 70% of all schools that we successfully match to the CCD-PSS file.

A.2.4 School weights As explained in Section 2 of the main text and Subsections A.2.2 and A.2.3 of this appendix, the dataset of our analysis includes about 60% of the schools from the pooled CCD/PSS file after filtering out schools with sparse or noisy visit data. We augment the dataset with school-level weights to alleviate concerns about its representativeness. We estimate a Probit model where the left-hand side variable is an indicator y_j that takes the value of 1 if school j is included in the dataset of our analysis and is 0 otherwise. The regressors of the Probit model are: a polynomial of county population, population density, county-level shares of High School and College workers, county-level shares of married adults, dummy variables for local area types (i.e., city, suburban, town or rural area) and dummy variables for the nine U.S. Census divisions. Then, we weight each public school by the inverse of the predicted probability $\hat{\Pr}\{y_j = 1\}$, and each private school by its PSS sampling weight times the inverse of the predicted probability $\hat{\Pr}\{y_j = 1\}$.[6] We check the quality of this adjustment by comparing the weighted counts of students, teachers, and schools in the data to the same counts based on the pooled CCD/PSS file (i.e. those reported in the second column of Table A1).

A.2.5 Comparison of selected sample to the NCES universe of schools Table A1 compares aggregates from the CCD and PSS to the NCES's digest of education's statistics (see https://nces.ed.gov/programs/digest/). The CCD files we are using were released only recently and have not yet been used by the NCES to produce official statistics, but the close similarity between all counts (number of educational institutions, number of students, number of teachers) suggests that the CCD and PSS files put together cover the universe of elementary and secondary schools.

Through direct merges, matching and fuzzy-name matching (see Appendix A), we combine the CCD-PSS file to Safegraph data. Out of 124,583 schools in the CCD-PSS file, we obtain 102,774 matches to the Safegraph data for which the quality of the match is high enough to be considered as reliable. We call this subset of schools the "matched" sample. Out of those schools, we retain 69,336 schools in our analysis and discard the remaining ones that suffer from excessively sparse or noisy $\Delta\widetilde{v}_{j,t}$ (see Subsection A.2.3 for details). The remaining schools constitute the "in scope" dataset. Table A2 compares different observables in the full CCD-PSS file (columns 1 and 4) with the schools that we match to Safegraph data (columns 2 and 5) and to the subset of these schools that we retain in our analysis of school visits (columns 3 and 6). The characteristics of the full CCD/PSS file and our matched dataset are very similar, suggesting that the algorithm does not selectively pick certain schools while discarding others. In columns 3 and 6 all the entries (except sample counts) are computed using the school weights described in Subsection A.2.4, which by construction make the in-scope dataset representative of the universe of CCD-PSS schools.

[6] Since the CCD contains the universe of public schools, the sampling weight of public schools is 1 and therefore the adjusted weight is 1 divided by the probability of selection into the "in scope" dataset. Across all schools, the final weights that we obtain range from 1.24 to 138.6 with an average of 1.73 and a median of 1.56. For public schools, the weights range from 1.24 to 13.7 with an average of 1.64 and a median of 1.52. The larger weights of the "in scope" dataset are for private schools, but the large values come from the PSS sampling weights (which can go all the way up to a value of 85), as opposed to reflecting very small values of $\hat{\Pr}\{y_j = 1\}$.

Table A1: Comparison to the NCES digest of education's statistics

Number of educational institutions		
	NCES table 105.50	CCD & PSS
	(1)	**(2)**
Public Schools	**98,469**	**101,688**
Elementary	67,408	68,953
Secondary	23,882	21,434
Combined	6,278	6,678
Other[a]	901	4,623
Private Schools	**32,461**	**27,641**
Elementary	20,090	17,378
Secondary	2,845	2,301
Combined	9,526	7,962
All	**130,930**	**129,329**
Number of students (in 1,000s)		
	NCES table 105.20	CCD & PSS
	(1)	**(2)**
Public Schools[b]	**50,686**	**50,834**
Prekindergarten to grade 8	35,496	33,415
Grades 9 to 12	15,190	17,419
Private Schools	**5,720**	**4,090**
Prekindergarten to grade 8	4,252	3,450
Grades 9 to 12	1,468	0.639
All	**56,406**	**54,924**
Number of teachers (in 1,000s, full-time equivalents)		
	NCES table 105.40	CCD & PSS
	(1)	**(2)**
Public Schools	3,170	2,911
Private Schools	482	401
All	**3,652**	**3,312**

Notes: NCES numbers refer to the year 2017-2018. Public schools classified as "Other", denoted by [a], includes special education, alternative, and other public schools not classified by grade span. NCES enrollment numbers in public schools, denoted by [b], include imputations for public school prekindergarten enrollment in California and Oregon.

A.2.6 A closer look at the school visits data As a final step towards preparing our combined Safegraph-NCES "in scope" dataset, we adjust changes in school visits $\Delta\widetilde{v}_{j,t}$ in the following way. First, we top-code $\Delta\widetilde{v}_{j,t}$ at 100%. Second, if in any week t outside of the reference period $\Delta\widetilde{v}_{j,t} > 25\%$ while $\Delta\widetilde{v}_{j,t-1} \leq 25\%$ and $\Delta\widetilde{v}_{j,t+1} \leq 25\%$, we replace $\Delta\widetilde{v}_{j,t}$ by the average of $\Delta\widetilde{v}_{j,t-1}$ and $\Delta\widetilde{v}_{j,t+1}$. This adjustment implements the assumption that during the school year 2020-21, schools did not reopen for only one week at a time.

Figure A3 plots the distribution of average changes in visits $\Delta\widetilde{v}_{j,t}$ in September and October of 2019 (that is, before the base period) and in January and February of 2020 (during the base period) for schools that we retain in our analysis. Despite the various adjustments (Section 2 of main text and Subsection A.2), we see a substantial variation in school visits: each panel in Figure A3 uses 4 weeks of data for each school j, and yet a non-trivial share of changes in visits fall outside of the [-20%, +20%] interval. This said, some of this dispersion may capture variations in school activity across months. For instance, some schools may not reopen right in the beginning of September 2019, which would explain why the

Table A2: Comparison between all schools and schools from the in-scope dataset

	Public schools			Private schools		
	All	Matched[a]	In scope[b]	All	Matched[a]	In scope[b]
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Sample count	101,688	85,276	57,730	22,895	17,498	11,606
Student-teacher ratio	15.7	15.5	15.5	10.6	10.6	10.9
% Male	52.2	52.1	51.8	52.1	52.2	51.8
% Indian	1.84	1.68	1.20	0.68	0.66	0.63
% Asian	3.87	3.88	4.22	5.51	5.64	6.12
% Pacific	0.40	0.34	0.36	0.49	0.50	0.59
% Hispanic	25.2	24.6	25.0	12.1	12.4	13.7
% White	49.9	51.4	51.5	65.6	64.9	62.6
% Black	14.6	13.8	13.5	11.6	11.9	11.8
% Other	4.29	4.34	4.33	3.99	4.08	4.49
% Free lunch[c]	44.2	43.8	43.6	n.a.	n.a.	n.a.
% Reduced-price lunch[c]	5.07	5.13	5.33	n.a.	n.a.	n.a.
City	27.6	26.1	27.8	32.1	33.0	38.8
Suburban	31.4	31.9	30.8	36.7	37.7	40.1
Town	13.2	13.7	13.8	8.90	9.06	8.42
Rural	27.8	28.4	27.7	22.4	20.3	12.6

Notes: Schools marked as "Matched", denoted by [a], refer to schools matched to Safegraph data. Schools marked as "In scope", denoted by [b], refer to schools matched to Safegraph data and with visits data that is neither too sparse or too noisy (see Subsection A.2.3). Except for the sample count, all the statistics for the "In scope" data are based on the school weights computed in Subsection A.2.4. % Free lunch and % Reduced-price lunch, denoted by [c], refer to the school shares of students who are eligible for free and reduced-price lunches, respectively.

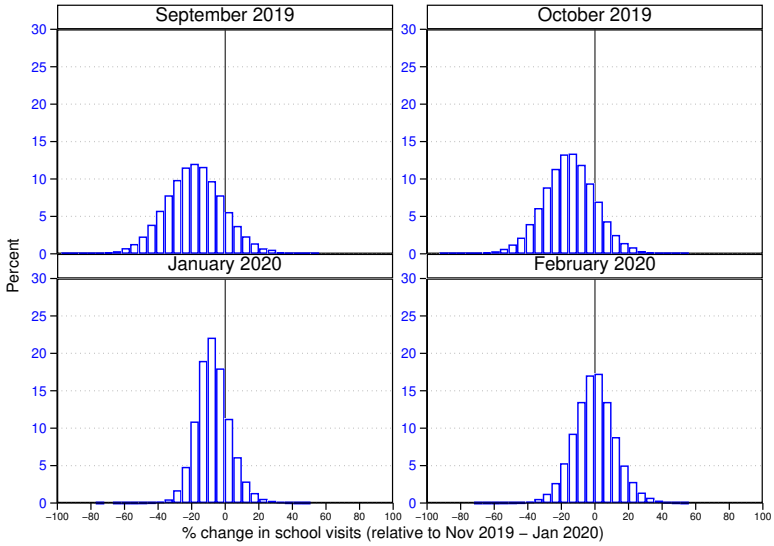
distribution is shifted to the left. Similarly, in January 2020, some schools might start later than others and therefore, the distribution is also shifted to the left.

Figure A4 plots the distribution of average $\Delta\widetilde{v}_{j,t}$ during different months of the COVID-19 pandemic. The upper panel focuses on the first 4 months of the pandemic. The top left plot in this panel shows that $\Delta\widetilde{v}_{j,t}$ does well in capturing week-to-week variations: most schools were open during at least the first two weeks of March 2020 before being shut down, and as a result the change in school visits averaged over the 4 weeks of this month is -46 on average. In the other plots of the upper panel, the shift closer to -100% is obviously indicative of school closures.[7]

The middle and lower panels of Figure A4 show the distribution of average $\Delta\widetilde{v}_{j,t}$ during the Fall of 2020 and Spring of 2021. Note that the scale on the vertical axes of the plots is the same in the two

[7]In April, May and June in the upper panel of Figure A4, we observe some schools with changes in visits not lower than -60% or -80%. To understand how this relates to the upper map of county-level loss of EIPL (Figure 1, main text), recall that: 1) these changes in visits are translated into EIPL by being multiplied by a coefficient that can be greater than 1 as shown in Table 1 of the main text, and 2) week-to-week variations in visits at the individual school level imply that a school might have $\Delta\widetilde{v}_{j,t}$ between, say, -60% and -80% in May and between -80% and -100% in April and June. The latter source of variation is not present in Figure 1 since the data is averaged over longer periods of time.

Figure A3: Distribution of changes in school visits before and during the base period



Notes: The figures show the distribution of the average change in school visits during 4 months prior to the pandemic.

panels. In panel (b), we see a recovery of $\Delta\widetilde{v}_{j,t}$ relative to the first few months of the pandemic, which is likely indicative of school reopenings in some regions. Then, we observe a slight reversal in November and December relative to September-October 2020, which is possibly linked to a tightening of health restrictions but also due to the fact that both months include one week of vacation (Thanksgiving in November 2020, Christmas in December 2020). The lower panel of Figure A4 shows a clearer recovery in school visits, though with substantial mass around 0 or higher. In fact, the distributions plotted in the lower panel are better thought of mixtures of two distributions: one for the still closed schools similar to the distribution of April 2020, and another one for re-opened schools that is thus similar to the distributions in Figure A3.

A.3 Burbio and Return2Learn data

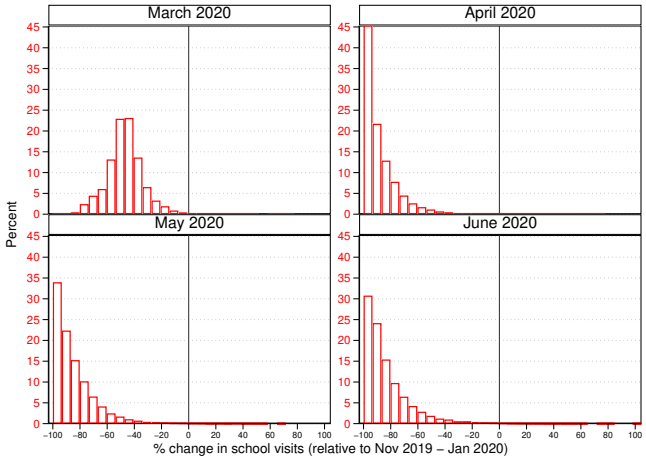
Burbio and Return2Learn (R2L) are learning mode trackers that consist of weekly indicators of the share of public school students engaged in either "Traditional", "Hybrid", or "Virtual" learning mode. R2L data is available at the school district level, while Burbio aggregates the information gathered from school districts to the county level.

Figure A5 plots time series of the percentage share of each learning mode according to either of the school trackers, aggregated across counties or school districts using student enrollment. The figure shows a similar evolution over time, in particular with regards to the decline in virtual learning in 2021. It also evidences important differences in levels in what concerns the share of hybrid versus traditional in-person learning. For instance, in May 2021, Burbio shows 70% of traditional learning and 30% of hybrid learning, whereas Return2Learn shows 55% of traditional learning and 45% of hybrid learning. Across all counties and weeks of the school year 2020-21, the correlation (weighted by student enrollment at the county level) between Burbio's and R2L's shares of traditional learning is 0.59, and the correlation for hybrid learning is 'only' 0.45. On the other hand the correlation for virtual learning is 0.74.

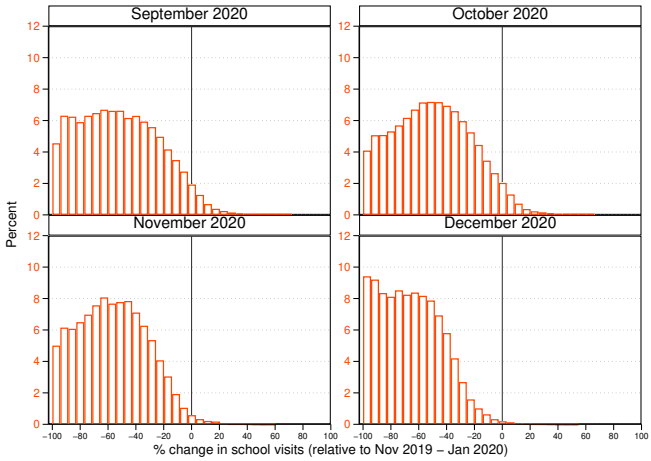
The differences between Burbio's and R2L are even more pronounced when we focus on certain regions and/or during certain periods of the school year 2020-21. To illustrate the former, we computed for each learning mode the correlation between Burbio and R2L at the county level across all weeks of 2020-21. For traditional and hybrid learning modes, the correlation between Burbio and R2L is negative for about 20 to

Figure A4: Distribution of changes in school visits during the pandemic

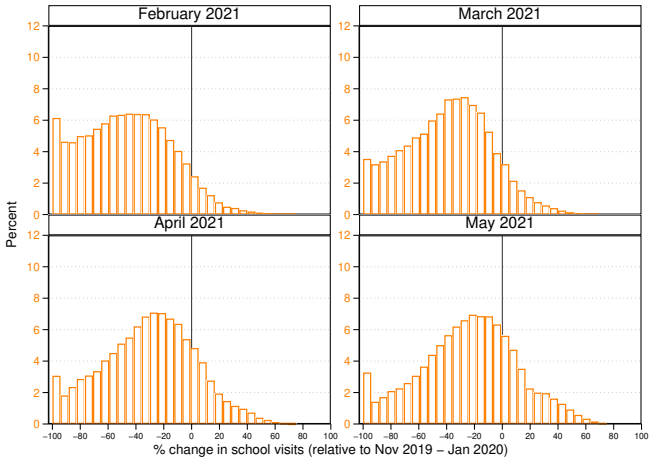
(a) March 2020 to May 2020



(b) September 2020 to December 2020

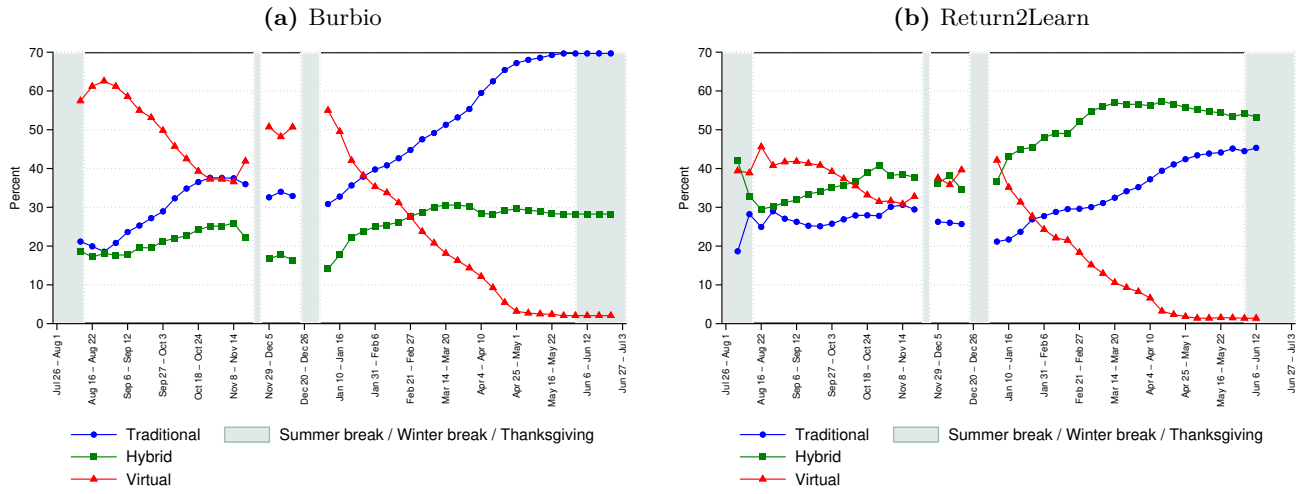


(c) January 2021 to May 2021



Notes: The figures show the distribution of the average change in school visits at points in time during the pandemic.

Figure A5: Evolution of learning mode trackers over time



Notes: The figures show the weekly percentage share of each learning mode according to Burbio and Return2Learn, aggregated using public school student enrollment at the county (Burbio) or district (Return2Learn) level.

25% of all counties (again, weighted by student enrollment). The interquartile ranges of the correlations are 0.78 for traditional and 0.66 for hybrid learning, and the 5-95 percentile ranges are respectively at 1.42 and 1.25. The dispersion is less pronounced in what concerns virtual learning mode, with an interquartile range of 0.41 and a 5-95 percentile range of 1.02 (indicating that the correlation between Burbio and R2L is negative for about 5% of all counties). Through a similar exercise, it can be shown that the differences between Burbio and R2L are more pronounced during the Fall of 2020 than during the Spring of 2021, especially for virtual learning.

A.4 Details on construction of EIPL measure

Our main methodological contribution is to construct a measure of EIPL that exploits the information contained in school visits from Safegraph and learning modes from Burbio and Return2Learn. As explained in Section 2, the analysis revolves around the estimation of Equation (4), which is repeated here for reference:

$$T_{c,t} = \alpha + \beta \Delta \widetilde{v}_{c,t} - \gamma H_{c,t} + \varepsilon_{c,t}. \tag{A.1}$$

$T_{c,t}$ and $H_{c,t}$ are shares of traditional and hybrid learning modes in county c during week t, $\Delta \widetilde{v}_{c,t}$ are changes in school visits aggregated to the county-level using student enrollment in public schools, α, β, γ are the regression coefficients, and $\varepsilon_{c,t}$ is the regression residual. When working with R2L data, instead of the county level we work with data at the school-district level. Essentially, Equation (A.1) configures a panel regression that can be estimated at different levels of geographic aggregation and/or over different sample periods.[8] We focus on the Core-Based Statistical Areas (CBSA) and state levels and proceed through the following steps:[9]

[8] When the sample period used to estimate Equation (A.1) includes the weeks of August and early September, the panel based on R2L data is unbalanced in some regions because R2L's earliest tracking date differs across school districts. The panel is always balanced when working with the Burbio data.

[9] Obviously Equation (A.1) can also be run directly at the county or school district levels. The downside of this approach is that some counties or school districts either have no time-variation in $T_{c,t}$ or $H_{c,t}$, or $T_{c,t}$ and $H_{c,t}$ almost perfectly negatively related to each other. The latter occurs when $V_{c,t} = 0$ for most t, making traditional learning mechanically related to hybrid learning; i.e. $T_{c,t} \approx 100 - H_{c,t}$. In a regression context, this implies $\gamma \to 1$ and $\beta \to 0$ since $\Delta \widetilde{v}_{c,t}$ is subject to idiosyncratic variation. As a result, steps 2 and 3 of our approach do not work well at these finer geographical levels. We

1. Restrict the data to counties or school districts for which our Safegraph-NCES dataset includes at least 5 schools.

2. For each geographical level (CBSA, state) and each school tracker (Burbio, R2L), find the sample period to estimate Equation (A.1) that yields the highest R-squared. In practice, we find that the best fitting sample periods are mostly during the Fall semester, in the sense that less than 25% of all regressions have a sample period that extends beyond mid-February 2021.

3. For each CBSA and each school tracker (Burbio, R2L), use the best fitting sample period to estimate Equation (A.1) and obtain two sets of estimates, $\left\{\hat{\beta}_{\text{Burbio}}, \hat{\gamma}_{\text{Burbio}}\right\}$ and $\left\{\hat{\beta}_{\text{R2L}}, \hat{\gamma}_{\text{R2L}}\right\}$. Similarly, obtain sets of estimates from each school tracker at the state level. Store the R-squared that are associated with each set of estimates.

   (a) Some CBSAs have empty estimates for either Burbio or R2L. This occurs when we have insufficient variation in $T_{c,t}$ or $H_{c,t}$, or when $T_{c,t}$ or $H_{c,t}$ are perfectly negatively related to each other; see Footnote 9.

4. At this point we have (at best) four sets of estimates to choose from to determine $\hat{\beta}$ and $\hat{\gamma}$ for each school j. Rank the estimates based on the associated R-squared, i.e. denote by $\hat{\beta}_{(k)}$ and $\hat{\gamma}_{(k)}$ the ranked estimates, where $\hat{\beta}_{(4)}$ and $\hat{\gamma}_{(4)}$ refer to the highest and $\hat{\beta}_{(1)}$ and $\hat{\gamma}_{(1)}$ to the lowest R-squared estimates. Set $\hat{\beta} = \hat{\beta}_{(k)}$ where k is the largest index such that $0 < \hat{\gamma}_{(k)} < 1$.

The fact that our procedure can be applied in almost all instances is due to a combination of factors. First, our Safegraph-NCES dataset has extensive coverage of schools for which $\Delta\widetilde{v}_{j,t}$ is neither too noisy or sparse. Second, by exploiting information from both Burbio or R2L, we increase the number of sets of estimates from which to choose from in the last step of our procedure. Not only is $0 < \hat{\gamma}_{(k)} < 1$ always satisfied for at least one set of estimates, but in many instances we find that $0 < \hat{\gamma}_{(4)} < 1$, which enables us to use the estimates with the highest R-squared.

There are, however, a few exceptions. For all four set of estimates (after finding the best fitting sample periods), the R-squared turn out to be no higher than 0.25 for most areas in Arkansas and Maine. This reflects low quality of the Burbio and R2L data in those areas. For instance, in Arkansas the regressions with R2L data at both the CBSA and state levels and the CBSA-level regressions with Burbio data almost always suffer from insufficient variation. As a result, we are left with no option but to use the state-level Burbio estimates, which suffer from a R-squared close to 0. In Maine, the best fitting R-squared is 0.23, except in one of the CBSA. To address these shortcomings, we replace the $\hat{\beta}$ and $\hat{\gamma}$ for schools in those areas of Arkansas and Maine with Burbio estimates from neighboring states (states of the West South Central division for Arkansas, states of the New England division for Maine). Meanwhile, in both states, we do have a few MSA-level estimates that we do not override given that they have a R-squared higher than 0.25.[10] Although not a validation of our approach, these CBSA-level $\hat{\beta}$'s turn out to be not too different from the estimates that we construct for Arkansas and Maine using neighboring states.

A.5 Details on data used in regressions

ACS data The sociodemographic and income variables (mean household income, share of individuals with some College or higher educational attainment, share of dual-headed households with children) are based on the American Community Survey (ACS) 5-year estimates for the release years 2016-2019. The estimates are computed at the Census Block Group (CBG) level. To aggregate data to the 5-digit zip-code level, we use the ZIP-TRACT crosswalk provided by the U.S. Housing and Urban Development

sometimes run into a similar issue at the CBSA level.

[10] In Maine, we have CBSA-level estimates for the Portland-South Portland-Biddeford metropolitan area. In Arkansas, we have CBSA-level estimates for Fayetteville-Springdale-Rogers and Memphis metropolitan areas, but note that both are multi-state metropolitan areas.

(HUD)'s Office of Policy Development and Research (see https://www.huduser.gov/portal/datasets/usps_crosswalk.html). We aggregate data using the so-called 'total ZIP ratio', which is the ratio of all addresses for each Census tract (the first eleven digits of the CBG code) associated with each zip code. To measure population density, we combine the ACS population estimates with land area data from the U.S. Department of Agriculture (USDA; see https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/). From the USDA, we also use the Rural-Urban continuum codes (see https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx).

EDGE data Data on the local labor costs of hiring PK-12 educators come from the Education Demographic and Geographic Estimates (EDGE; see https://nces.ed.gov/programs/edge/). The index proxies the outside options of PK-12 educators by using local information (obtained from restricted-use data from the ACS) on the wages and salaries of comparable workers, while excluding from the estimation sample anyone who has a teaching or educational administration occupation or who is employed in the elementary and secondary education industry. "Comparable" means that the index controls for a host of sociodemographic and employment characteristics on individuals who are college graduates; see https://nces.ed.gov/programs/edge/Economic/TeacherWage for methodological details. We use additional data from EDGE in robustness checks presented in Appendix C. We use the school neighborhood poverty estimates, constructed by EDGE based on data from the Census Bureau's American Community Survey which allow to compute local income-to-poverty ratios (IPR). IPRs measure the percentage of family income that is above or below the federal poverty threshold set for the family's size and structure. IPRs are then aggregated to the levels of the school neighborhood as identified by EDGE.

CPS data Data on teachers' unionization rates are computed from the Current Population Survey (CPS; see https://www.census.gov/programs-surveys/cps.html) using information from the outgoing rotation group samples of the survey. We pool together data from 2018 through 2020 for teachers and instructors in elementary and secondary schools (PEIO1OCD 2310, 2320, 2330, 2340) and define the unionization rate as the share of teachers who are either members of a labor union or covered by a union. We aggregate unionization rates to the CBSA level or to the state level for missing CBSAs. We validate the state-level unionization rates against official tabulations of state-level unionization rates of public school teachers published by the NCES (see https://nces.ed.gov/surveys/sass/tables/sass0708_043_t1s.asp). The NCES numbers, which measure the percentage share of teachers in a union or employees' association, are based on the public school teacher data files of the Schools and Staffing Survey (SASS), whose microdata is not publicly available.

SEDA data Data on school-level and district-level test scores come from the Stanford Education Data Archive (SEDA; see http://purl.stanford.edu/db586ns4974). SEDA is a data initiative launched in 2016 with the purpose of providing nationally comparable, publicly available test score data for U.S. public schools and public school districts. In our main analysis, we use test scores at the levels of school districts due to wider data coverage. These test scores are pooled across all school grades and across years 2013 through 2018; they are normalized either through a cutscore standardized to the nationally averaged reference cohort within subject, grade, and year (the CS scale), or through a grade-cohort standardized score (GCS scale); they are available separately for mathematics and RLA. We use the CS scale and take the mean of the mathematics and RLA test scores. At the school level, test scores are not available by subject and they cover a smaller portion of our dataset (see Appendix C for additional information). The data we use are from version 4.1 of the SEDA data, which is the most recent version as of this writing.

NERD$ data Data on spending per student at the school level come from NERD$, the National Education Resource Database on Schools. NERD$ is a data initiative of the Edunomics lab and the Massive Data Institute at Georgetown University (see https://edunomicslab.org/). It builds on the

federal Every Student Succeeds Act (ESSA) passed in December 2015 which, among other provisions, stipulates that states must report for every public school (and local educational agency) the total per-pupil spending of federal, state and local money disaggregated by source of funds for the preceding fiscal year. In practice, the school spending data tends to be scattered across different states' website, but NERD$ gathers these data together. The data we use are from the latest update of NERD$ dated from October 8th, 2021 and contain school-by-school actual spending amounts for the year of 2018-2019. The data matches 94% of the public schools of our dataset.

ESSER data Data on ESSER funding come from the compilation put together by Return2Learn and available on R2L's website (see https://www.returntolearntracker.net/esser/). The raw data covers all three waves of ESSER, that is to say the funds from the Coronavirus Aid, Relief, and Economic Security (CARES) act, the Coronavirus Response and Relief Supplemental Appropriations (CRRSA) act and the American Rescue Plan (ARP). Data are available at the level of school districts, and the R2L database comes with the NCES identifier for school districts. When matched to our own, it covers about 91% of school districts that include 95% of the public schools in our dataset.

COVID data Data for COVID cases and COVID deaths at the county level are based on the daily count and rates from the New York Times, the Johns Hopkins Coronavirus Resource Center, and the Centers for Disease Control and Prevention (CDC). Data on COVID vaccinations at the county level are daily rates from the CDC. We download these data from the COVID from the Opportunity Insights Economic Tracker repository (see https://github.com/OpportunityInsights/EconomicTracker). To aggregate each variable to the weekly level, we take the mean of the daily values for each variable. County-level counts of ICU beds come a report from Kaiser Health News accessed through a compilation of county-level health data available at: https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/tree/master/data.

Election data County-level results for presidential elections are downloaded from the MIT election Data and Science Lab (see https://electionlab.mit.edu/data). We use results for the 2020 presidential elections in our main analysis and results for the 2016 presidential elections in robustness analyses. State-level data on voter turnout rates are taken from the U.S. elections project (see http://www.electproject.org/). We use turnout rates for the 2018 and 2020 general elections.

NPIs data Data at the county-week level on Non-Pharmaceutical Interventions (NPIs) are downloaded from the repository of the Centers for Disease Control and Prevention (CDC; see https://data.cdc.gov/). We use information on the following NPIs: 1) Stay-at-home orders, which can be advisory/recommendation, mandatory only for individuals in certain areas of the jurisdiction, mandatory for at-risk individuals, or mandatory for all individuals, 2) Gathering bans, which can be bans on gatherings of more than 100 persons, more than 50 persons, more than 25 persons, more than 10 persons, or all social/public gatherings, 3) Mask mandates, which is an indicator that takes the value of 1 when a mask is required in public and is 0 otherwise. All county-week time series are from the September 10, 2021 update of the CDC data.

OA data In robustness checks presented in Appendix C, we use data from the Opportunity Atlas (OA; see https://www.opportunityatlas.org/). OA provides access to social mobility data assembled by researchers from the Census Bureau, Harvard University, and Brown University. The data we use is at the level of Census tracts, which we can link to our dataset through the Census Block Group (obtained through Safegraph) of each school. The variables from OA that we use are: (i) the mean household income rank for children whose parents were at the 25th percentile of the national income distribution, where incomes for children are measured as mean earnings in 2014-2015 when they were between the ages

31-37; (ii) the fraction of children born in 1978-1983 birth cohorts with parents at the 25th percentile of the national income distribution who were incarcerated in 2010; (iii) the average rent for two-bedroom apartments in 2015. For (i) and (ii), we use data pooled across genders and races/ethnicities.

B Comparison with other Safegraph-based indicators

Inferring in-person learning or school closures solely from Safegraph school visits data is difficult because a given decline in visits to a school buildings can mean different things. For instance, suppose cell phone usage is concentrated among school staff and parents. If school staff returned to school more quickly than students (e.g. to prepare the return of students or to teach only some students in person while others remained in remote-learning mode), or if students get dropped off and picked up by parents instead of using buses, then changes in Safegraph visits alone would overestimate in-person learning. Vice versa, consider a school that contains a playground or sports fields that are usually open to the public or used for games. If due to the pandemic, access to this playground or sports field is restricted even though the school has reopened, then changes in Safegraph visits alone would underestimate in-person learning. Our methodology of combining changes in Safegraph visits data with information from learning mode trackers at a relatively detailed level addresses these issues at least at the county, respectively district level.

This said, it is useful to review other studies' approaches to measuring school activity during the pandemic using visit data from Safegraph. The first of these approaches is from Bravata et al. [2021] who indirectly measure school closures using Safegraph data through:

$$\ln \triangle \text{school}_{ct} = \ln \left(\text{school}_{ct,2020} \right) - \ln \left(\text{school}_{ct,2019} \right) \tag{B.1}$$

(Equation (1) in their paper). In Equation (B.1), school_{ct} denotes the sum of visits to all POIs with NAICS 611110 within county c in week t, and $\ln \triangle \text{school}_{ct}$ is then the log-transformed county-level difference in school visits between each week t in 2020 and the same week in 2019. There are obvious important differences with our approach, such as the level of aggregation (county- vs. school-level) and the focus on year-over-year variation (as opposed to variation relative to a fixed, base period). The year-over-year variation is useful to address seasonal variations, but note that in our analysis we omit data that correspond to the Summer and Winter breaks and data for the week of Thanksgiving. Another difference is that the metric of the indicator in Equation (B.1) is not interpretable in terms of EIPL, but this issue is irrelevant to Bravata et al. [2021]'s analysis because they use $\ln \triangle \text{school}_{ct}$ as a right-hand side variable to proxy for school closures in a predictive regression of COVID transmission rates.

Parolin and Lee [2021] measure school closures with Safegraph data through:

$$\text{school closures}_{j,m} = \mathbb{1} \left\{ \frac{v_{j,m}}{v_{j,m-12}} - 1 < -\vartheta \right\}, \tag{B.2}$$

where $v_{j,m}$ denotes visits to school j during month m, ϑ denotes some threshold value, and $\mathbb{1}\{.\}$ is the indicator function. Equation (B.2) flags school j as being closed in 2020 if its year-over-year growth rate in visits drops by at least $\vartheta\%$. In their paper (and in the Internet repository providing their data) Parolin and Lee [2021] use a threshold value ϑ of 50%.[11] Their data cover public schools, and according to their data description they do not control for changes in Safegraph's sample size (i.e. changes in the number of cell phone devices) nor purge the data from potential outliers.[12] More importantly, this

[11] In data posted in their Internet repository, they also report the mean year-over-year growth rate of visits as well as the indicators from Equation (B.2) based on $\vartheta = 25\%$ and $\vartheta = 75\%$.

[12] When reporting year-over-year growth rates in visits, i.e. $\frac{v_{j,m}}{v_{j,m-12}} - 1$, the median rather than the mean value might be a more robust statistics because these growth rates unavoidably jump to extremely large values due to the denominator being closed to 0 in some months. These outlier values have a non-negligible impact on the growth rates averaged across schools (even in a large sample like the Safegraph sample of schools).

Table B1: Assessment of Equation (B.2) using data from Return to Learn

		(a) Burbio		(b) Return2Learn	
	Learning mode tracker:	Virtual learning	Not virtual	Virtual learning	Not virtual
	Equation (B.2):	Not closed	Closed	Not closed	Closed
		(1)	(2)	(3)	(4)
All public schools		32.1	39.5	15.7	56.7
Schools in:	Cities	32.4	37.5	29.7	48.3
	Suburbs	29.1	38.9	20.9	51.4
	Towns and rural areas	33.6	42.5	7.9	72.5
School size:	Small	38.7	43.5	10.2	75.4
	Medium	32.9	37.8	17.8	54.3
	Large	17.9	41.4	14.7	52.8

Notes: The table reports the share of schools× weeks in which schools not flagged as closed according to Equation (B.2) but are operating in "Virtual" learning mode according to Burbio or R2L (Columns (1) and (3)); and share of schools× weeks in which schools flagged as closed according to Equation (B.2) but not operating in "Virtual" learning mode according to Burbio or R2L (Columns (2) and (4)). Small: schools with fewer than 250 students; Medium: schools with between 250 and 750 students; Large: schools with more than 750 students. Data are public schools during the weeks of the Fall term of 2020.

threshold-based approach can be problematic when working with cell-phone based foot traffic data: its high volatility implies that the growth rates repeatedly cross the threshold ϑ and thereby generate many false positives and false negatives. To illustrate this problem, we apply Equation (B.2) to the Safegraph data and combine it with information from Burbio and R2L.[13,14] Our calculations, reported in Table B1, indicate substantial disagreement between the different indicators of school closures: 32.1% of schools are not closed according to Equation (B.2) but are located in counties that are operating in "Virtual" learning mode. Part of this could be explained by differences between school districts within counties. However, the R2L's estimates show that the rate of disagreement remains high (15.7%) when we identify schools in "Virtual" learning mode based on school district information. The middle and lower panels of the table further show that false positives and false negatives vary by geographic locations and school size. This suggests that the extent of disagreement between Equation (B.2) and school tracker information depends at least partly on the underlying quality of the Safegraph data.

C Additional tables and figures

This section presents complementary analyzes of the distribution of EIPL across regions and time (Subsection C.1); of the relation between EIPL and school types, grades and schools' locality (Subsection C.2); additional regression results behind the results presented in Section 5 (Subsection C.3); and descriptive statistics for the main regression variables (Subsection C.4).

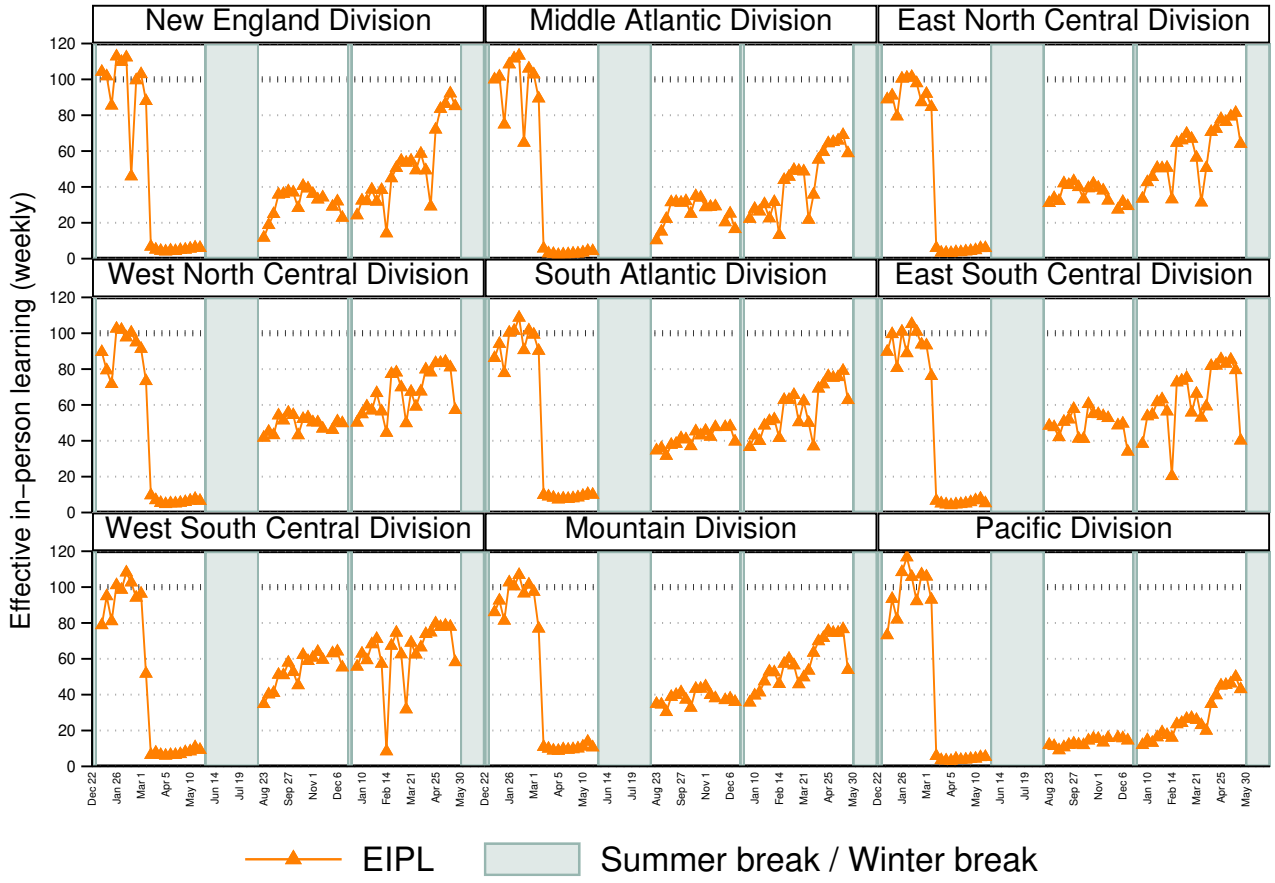
C.1 Regional disparities in EIPL over time

Figure C1 summarizes the temporal and geographic variation in EIPL by averaging weekly student-weighted EIPL for each of the nine U.S. Census Divisions. While EIPL drops to near zero for all divisions between March and May 2020, we see large differences during the 2020-21 school year. EIPL in states

[13] We use $\vartheta = 50\%$, but adjusting the threshold ϑ changes little the conclusions of our analysis.

[14] We restrict the analysis to the public schools of our dataset (to match the sample of Parolin and Lee [2021]'s analysis) during the weeks of the Fall term of 2020. The data contains 76,077 public schools and 16 weeks.

Figure C1: Weekly effective in-person learning, by Census divisions



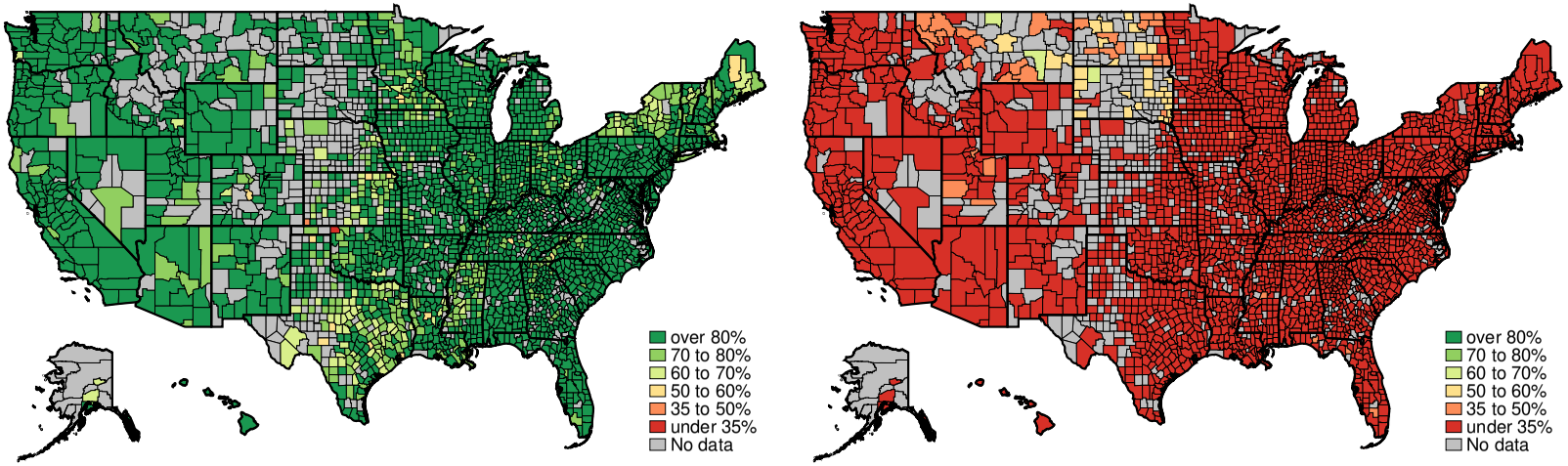
Notes: The figure shows student-weighted, weekly effective in-person learning for the different U.S. Census Divisions. U.S. Census Divisions are: New England (CT, MA, ME, NH, RI, VT), Middle Atlantic (NY, NJ, PA), East North Central (IL, IN, MI, OH, WI), West North Central (IA, KS, MN, MO, NE, ND, SD), South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV), East South Central (AL, KY, MS, TN), West South Central (AR, LA, OK, TX), Mountain (AZ, CO, ID, MT, NM, NV, UT, WY), Pacific (CA, OR, WA).

in the West North Central, East South Central and West South Central division quickly increase to 60% from September 2020 through December 2020 and climb to over 80% from January through May 2021. In contrast, EIPL in states in the New England, the Middle Atlantic and especially the Pacific division remains below 50% for most of the 2020-21 school year. Some of the regional disparities appear right after the end of the Summer break, while others built up later during the school year.

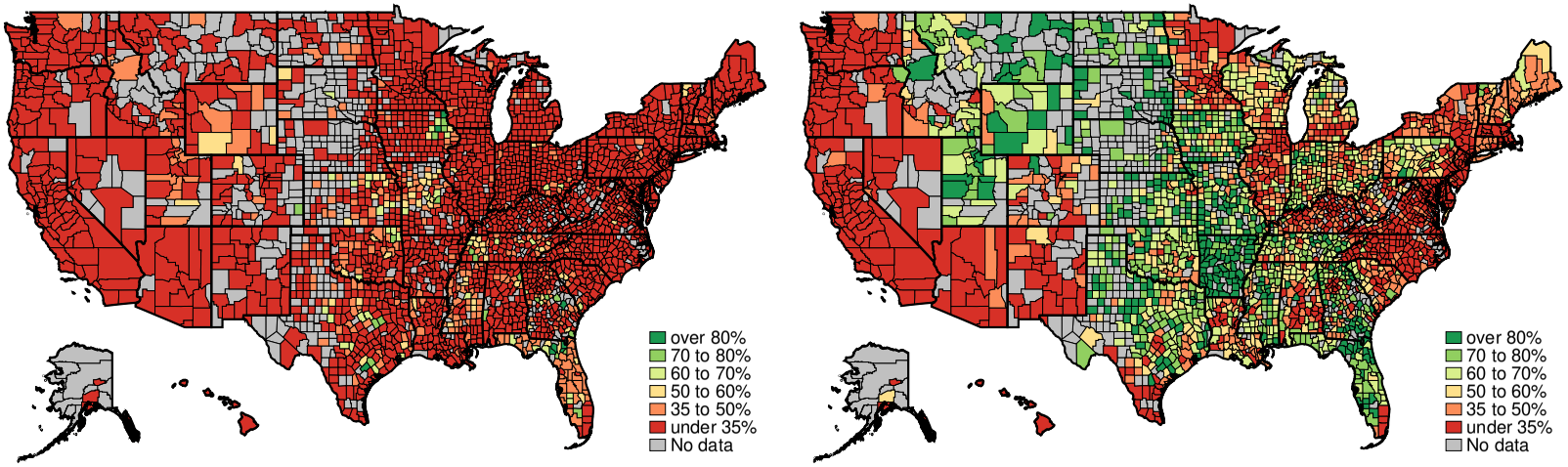
The upper panel of Figure C2 reinforces the finding from Figure C1, that EIPL at the beginning of the pandemic drops sharply across all regions of the country. The left plot in this panel shows county-level EIPL on average from mid-February 2020 until mid-March 2020. It shows levels of EIPL over 80% in almost every county. By contrast, the plot on the right-hand-side of this panel showing EIPL on average from mid-March 2020 to mid-April 2020 is almost uniformly "red". There are a few exceptions, notably in Montana, North and South Dakota, which seem to line up with state-specific responses in the adoption of mitigation strategies at the beginning of the COVID pandemic. For example, the governor of the state of South Dakota adopted an executive order to encourage social distancing and remote work in mid-March of 2020 but resisted imposing a mandatory, state-wide lockdown, and later on ruled out a state mandate on the wearing of face masks in public spaces.

Figure C2: Effective in-person learning across U.S. counties

(a) 4 weeks before (left) and after (right) March 15, 2020

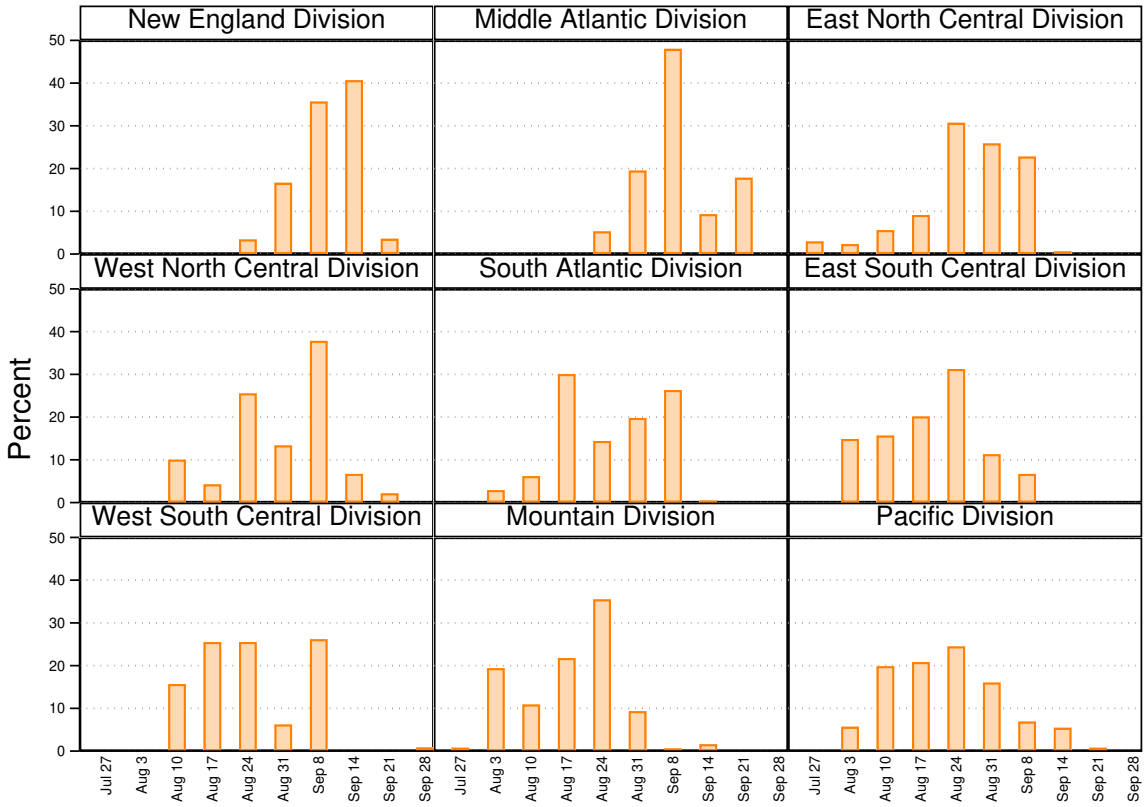


(b) 4 weeks before (left) and after (right) back to school dates according to pre-COVID school district calendars



Notes: The figure shows the student-weighted average county EIPL for all counties for which we have reliable data on at least three schools.

Figure C3: Back to school dates (based on pre-COVID school district calendars), by Census divisions



Notes: The figure shows the percent of public schools in each Census division that would open on the days in question according to pre-COVID school district calendars. U.S. Census Divisions are: New England (CT, MA, ME, NH, RI, VT), Middle Atlantic (NY, NJ, PA), East North Central (IL, IN, MI, OH, WI), West North Central (IA, KS, MN, MO, NE, ND, SD), South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV), East South Central (AL, KY, MS, TN), West South Central (AR, LA, OK, TX), Mountain (AZ, CO, ID, MT, NM, NV, UT, WY), Pacific (CA, OR, WA).

To further analyze the regional and temporal disparities in EIPL, we next look at EIPL around the time when schools *usually* reopen after the Summer break. To this end, we use information from Burbio about the week when most public schools within a county usually reopen. This information, which is reported in Figure C3 by showing back-to-school dates for each of the nine Census divisions, reveals major differences across different parts of the country. For example, in the East South and West South Central divisions, students usually head back to school at the beginning of August, whereas in New England and in the Middle Atlantic division back-to-school dates are typically after Labor Day.[15,16] We use this information (available at the county-level) in the lower panel of Figure C2 to check EIPL in the four weeks before and after back-to-school dates. As can be seen on the left-hand side of the figure, our measure of EIPL indicates close-to-zero in-person learning for most counties before their usual back-to-school dates. That the picture is not uniformly red could reflect inaccuracies of our EIPL measure, but may also be explained by (i) usual back-to-school dates which are not uniform within a county or not well measured by Burbio, (ii) reopening of private schools that may be asynchronous with that of public schools, (iii) schools reopening for in-person learning earlier than what they usually do in normal times. Then, on the right-hand side of the panel, we observe many counties shifting from close-to-zero to much higher levels of EIPL. On the other hand, and as expected, most counties where we measure low EIPL throughout the school year 2020-21 are in red color in the lower panel of Figure C2 before and after usual back-to-school dates.

Table C1: Dispersion in effective in-person learning within counties

Category of county EIPL average **(1)**	Mean county EIPL average **(2)**	Mean county EIPL interquartile range **(3)**
All counties	51.7%	15.5%
0% – 25%	22.6%	11.1%
25% – 50%	38.1%	16.6%
50% – 75%	61.8%	16.6%
75% – 100%	82.5%	11.2%

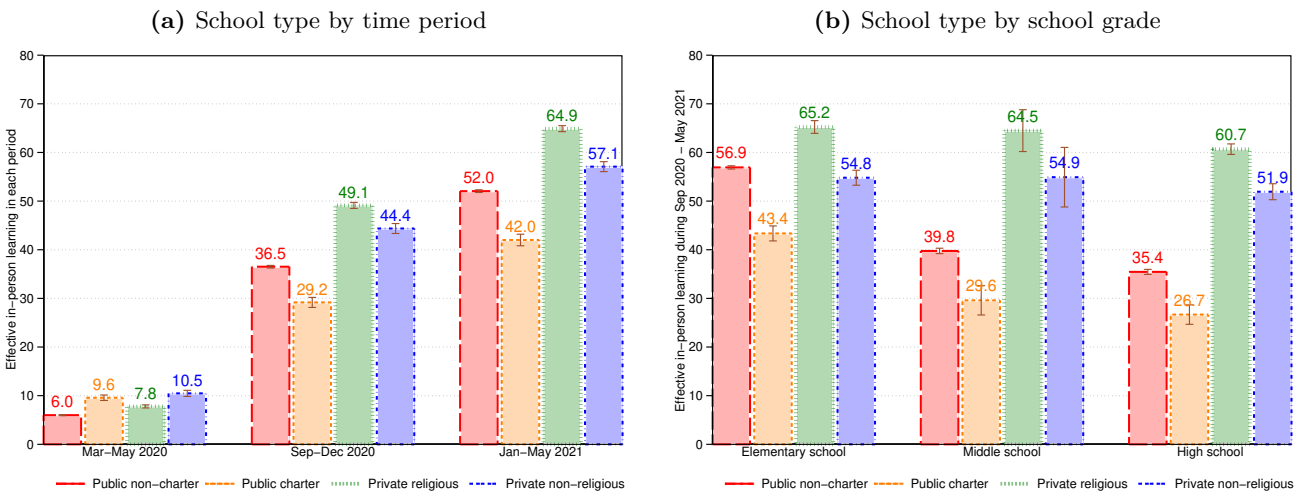
Notes: The table shows the mean county EIPL average and the mean county EIPL interquartile range for groups of counties in a given category of average EIPL from September 2020 to May 2021. The EIPL average for each county is computed as the weighted average of EIPL across schools within the county. The EIPL interquartile range for each county is computed as the difference between the 75th and 25th percentile of the weighted distribution of EIPL across schools within the county. The sample consists of counties with at least three schools with reliable data. The school-specific weights are constructed to keep the sample representative of the population. See Appendix A.2 for a description.

The results in this section focus on differences across regions and/or across counties. At the same time, it is important to emphasize that there are large disparities in EIPL across schools even within counties. Table C1 illustrates this pattern by reporting measures of dispersion within counties. Among all the counties with at least three schools with reliable data, the mean county interquartile range of EIPL across schools is 14% for the 2020-21 school year. Furthermore, the extent of this within-county dispersion is similar across counties with different levels of average EIPL. This suggests that the disparity in EIPL does not only reflect systematic regional differences but may also be driven by school-specific characteristics and local conditions that apply similarly across the country.

[15] Research from the Pew Research Center indicates that these differences are historically related to preferences over teenagers taking on work summer jobs, constraints that limit the time when families can take vacations, and the economic importance of tourism and hospitality industries.

[16] In Figure C1, the shaded area denoting the Summer break of 2020 covers the weeks from May 31st until August 22nd. For many schools across the country, this time interval is only an approximation of the Summer break since, as shown in Figure C3, back-to-school dates are not uniform across regions.

Figure C4: Effective in-person learning by school type and grade



Notes: The figures show student-weighted average EIPL for private schools versus public schools by time period and by school grade.

C.2 Relation of EIPL with school type, grade and locality

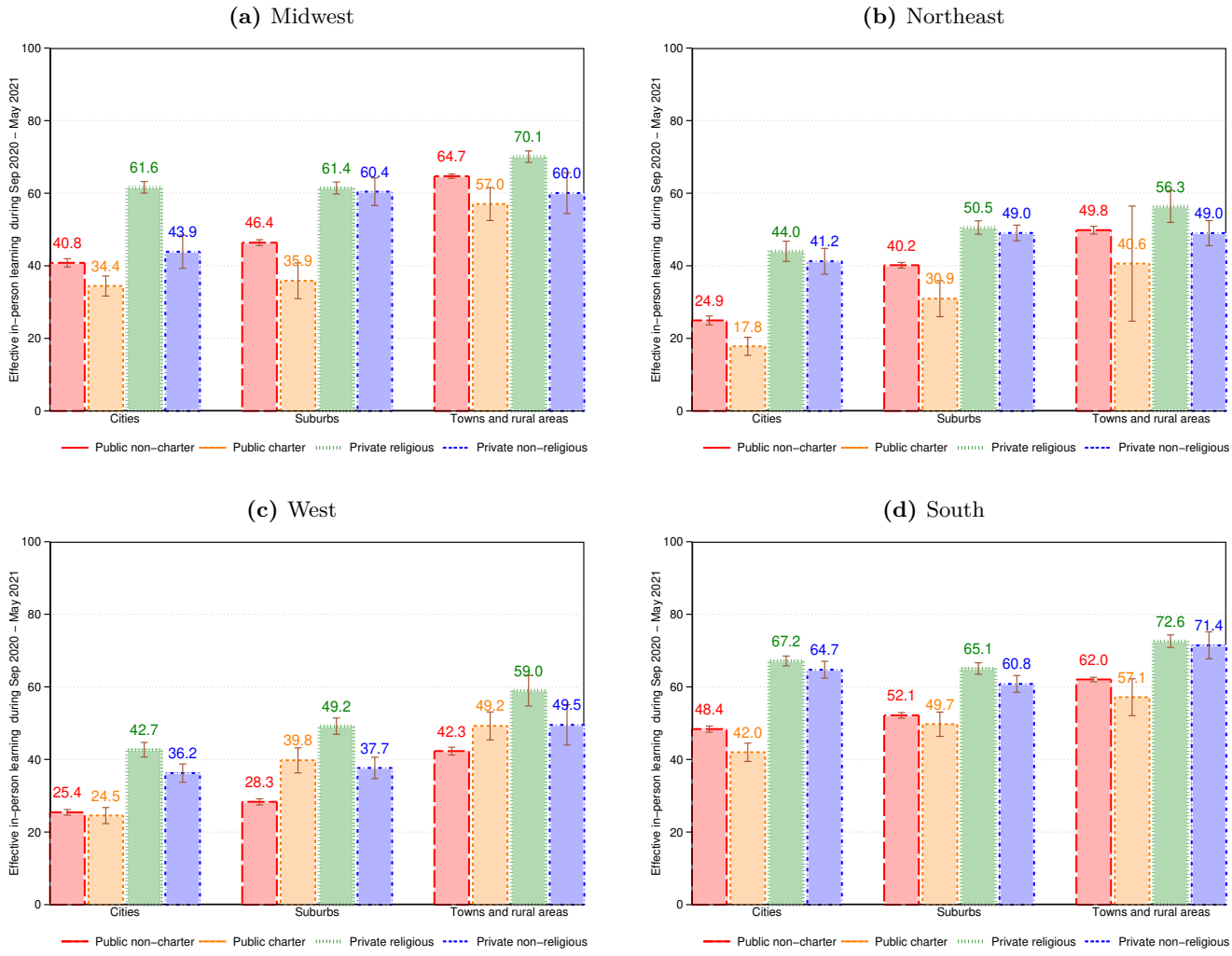
Panel (a) of Figure C4 shows differences in average EIPL by school type and time period. During the first three months of the pandemic, there is almost no difference in EIPL across school types. During both Fall 2020 and Winter/Spring 2021, however, we see substantial differences. Over the entire 2020-21 school year, EIPL is 10% lower for public schools than for private schools, with public charter schools averaging the least EIPL, followed by public non-charter, private non-religious, and private religious schools.

Panel (b) of Figure C4 reports on differences in average EIPL between September 2020 and May 2021 by school type and school grade. Across all four school types, EIPL is highest for elementary schools and lowest for high schools. For private schools, the difference in EIPL across school grades is smaller than for public schools.[17] In other words, the differences in EIPL between public and and private schools that we observe in Panel (a) are in large part due to differences in EIPL at the middle and high school level.

Another potentially important school characteristics to understand EIPL during the pandemic is the type of surroundings in which schools are located. We find that schools that are located in a city have on average lower EIPL than schools in suburbs, which themselves have lower average EIPL than schools located in a town or a rural area. This relationship is robust in the sense that, as shown in Figure C5, it holds true for each school type (public non-charter, public charter, private religious, and private non-religious) within all four regions of the U.S. Interestingly, the relation is weaker in the South for private schools; and the difference between public non-charter and public charter schools is reversed in the western part of the country. Figure C5 also shows that the magnitude of the EIPL gap between public and private schools differs across regions; for instance it is larger in cities of the Northeast region of the country. Notice that the relations of EIPL with school type, grade and locality in Figure C5 represent unconditional averages, and therefore do not control for other predictors that light up in the OLS regressions.

[17] Note that for elementary schools, EIPL is slightly higher for public non-charter schools than for private non-religious schools. This change in ranking of school types compared to the ranking across all school grades is due to geographical differences in the relative prevalence of private non-religious elementary schools.

Figure C5: Effective in-person learning by school type, locale and U.S. region



Notes: The figures show student-weighted average EIPL by school type and and by locale for the different U.S. regions. U.S. regions are: the Northeast (CT, MA, ME, NH, RI, VT, NY, NJ, PA), the Midwest (IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD), the South DE, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX), the West (AZ, CO, ID, MT, NM, NV, UT, WY, CA, OR, WA).

C.3 Additional regression results for Section 5

This section presents results from several regressions summarized in Section 5 of the paper.

We begin in Table C2 with results of regressing EIPL separately on each of the three affluence measures together with the share of non-white students and controls for school type and school grade. Adding these controls does not change the results noticeably, and the estimates on these controls are in line with the results shown in Figure C4. As mentioned in the text, the reason we do not include all three affluence measures together in the regressions is that the high correlation between them make it difficult to interpret the estimates; see Subsection C.4 for additional information.

Table C2: The inverse relationship of effective in-person learning with affluence and race

Dependent variable	Effective in-person learning (EIPL)					
	(a) Public schools			**(b)** Private schools		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Zip-level mean household income	-4.98***			-4.49***		
	(0.47)			(0.62)		
Zip-level share of college educated		-6.87***			-6.81***	
		(0.45)			(0.81)	
Zip-level share of dual-headed households			-3.07***			-2.62***
			(0.71)			(0.79)
School share of non-white students	-19.89***	-19.53***	-20.95***	-6.66***	-6.88***	-7.00***
	(1.34)	(1.36)	(1.58)	(1.02)	(1.04)	(1.11)
School type and grade controls	✔	✔	✔	✔	✔	✔
R-squared	0.09	0.09	0.08	0.04	0.04	0.03
# of counties	2,937	2,937	2,937	1,489	1,489	1,489
# of schools	55,036	55,037	55,036	11,280	11,280	11,280
# of school districts	12,353	12,353	12,353			

Notes: Each column reports coefficients from a weighted OLS regression with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on average school EIPL for the period from September 2020 to May 2021. Panel (a) shows estimates for the public school sample, and panel (b) shows estimates for the private school sample. The school type fixed effects consists of indicators for charter school and non-charter school for the public school sample, and religious school and non-religious school for the private school sample. The school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples.

Next, we focus on the role of (pre-COVID) test scores, school size, and school funding. Except for school enrollment, the different variables are not available for private schools. We therefore focus here on public schools. Column (1) in Table C3 repeats the results from Table C2 above as a reference.[18]; Column (2) adds district-level test scores to the regression; Column (3) adds school size; Column (4) adds school spending and ESSER funding; and Column (5) adds the four variables jointly.

Table C4 analyzes the role of geography. Column (1) in panel (a) repeats the final regression in Table C3 above for reference. Columns (2) and (3) add fixed effects for the state, respectively the county in which the school is located. The consequences of controlling for these more detailed geographical effects are important, raising the explanatory power of the regressions to almost one third, and can be summarized as follows.

First, the inverse relation between EIPL and local affluence is cut in half when the state fixed effect is added, and essentially disappears when the county fixed effect is added. Similarly, the association of EIPL with local education is substantially reduced although it remains negative, implying that even

[18]The estimates for household income, share of college educated and school neighborhood poverty are exactly as in columns (1) - (3) of Table C2. The estimate for share of non-white students is slightly different because this estimate is obtained while controlling jointly for all three measures.

within counties, schools located in zip-codes with a higher share of college-educated households provided on average somewhat lower EIPL. We conclude from these estimates that EIPL is negatively related to affluence and education primarily because less affluent and less educated areas of the county have public schools that provided more EIPL during the 2020-21 school year. Second and contrary to affluence and education, the inverse relation between EIPL and the share of non-white students is unaffected by state fixed effects and is reduced by only about one third by the county fixed effect. So, even within counties and controlling for affluence, education and other school characteristics, there are clear racial differences in that schools with a larger share of non-white students provided on average substantially lower EIPL. Third, the negative coefficient estimate on school size remains unaffected by the state and county fixed effect. The result is interesting because it suggests that smaller schools reopened to in-person learning more quickly than larger schools, perhaps because the logistical challenges of reopening or equity concerns about reopening only certain grades were less important.[19]

Table C3: The role of test scores, school size, and school funding

Dependent variable	Effective in-person learning (EIPL)				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Zip-level mean household income[a]	-4.98***	-6.54***	-4.38***	-5.32***	-5.19***
	(0.47)	(0.56)	(0.47)	(0.53)	(0.59)
Zip-level share of College educated[a]	-6.87***	-9.28***	-6.16***	-7.00***	-7.65***
	(0.45)	(0.43)	(0.48)	(0.47)	(0.49)
Zip-level share of dual-headed households[a]	-3.07***	-3.02***	-2.25***	-3.81***	-3.21***
	(0.71)	(0.81)	(0.70)	(0.74)	(0.75)
School share of non-white students[b]	-18.54***	-15.76***	-17.27***	-16.54***	-12.24***
	(1.71)	(1.58)	(1.63)	(1.56)	(1.43)
Average district-level test scores[b]		4.99***			5.42***
		(0.74)			(0.86)
Student enrollment[b]			-2.26***		-3.63***
			(0.34)		(0.31)
School spending per student[b]				-3.59***	-4.90***
				(0.63)	(0.67)
ESSER funding per student[b]				-2.67***	-0.88
				(0.56)	(0.66)
School type and grade controls	✔	✔	✔	✔	✔
R-squared	0.09	0.10	0.09	0.10	0.11
# of counties	2,920	2,920	2,920	2,833	2,833
# of schools	55,006	55,006	55,006	51,577	51,577
# of school districts	12,352	12,352	12,352	11,119	11,119

Notes: Each column reports coefficients from a weighted OLS regression on the public school sample, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on average school EIPL for the period from September 2020 to May 2021. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. The coefficient estimates for the affluence measures, denoted by [a], are the result of separate regressions with each one of the measures in combination with the other variables below. The coefficient estimates for the other regressors denoted by [b] are the result of regressions where the three affluence measures are included jointly.

[19] As noted above, the regressions control for whether the school is an elementary school, high school, or combined school; but these controls are relatively coarse and there may substantial variations in the number of grades served by a school even within these categories.

Panel (b) of Table C3 repeats the same set of regressions for private schools. As with our analysis of public schools, we find that the role of affluence is substantially reduced (the coefficient on mean household income becomes not statistically different from zero), and the coefficient on the share of non-white students decreases too. The fixed effects raises the explanatory power of the regressions to almost 20%. In sum, panel (b) supports the conclusion that the negative relation between EIPL and affluence and education is driven by less affluent and less educated areas of the county having schools (public, but also private) that provide more EIPL during the 2020-21 school year. On the other hand, the regressions suggest that our findings on the role of school size do not extend to private schools.

Table C4: The importance of geography

Dependent variable	Effective in-person learning (EIPL)					
	(a) Public schools			**(b)** Private schools		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Zip-level mean household income[a]	-5.19***	-2.11***	-0.79***	-4.34***	-1.14**	0.17
	(0.59)	(0.33)	(0.24)	(0.61)	(0.52)	(0.66)
Zip-level share of college educated[a]	-7.65***	-5.33***	-2.94***	-6.66***	-3.30***	-2.11**
	(0.49)	(0.37)	(0.28)	(0.79)	(0.65)	(0.86)
Zip-level share of dual-headed households [a]	-3.21***	-0.65*	-0.06	-2.49***	0.57	1.56**
	(0.75)	(0.37)	(0.29)	(0.78)	(0.60)	(0.65)
School share of non-white students[b]	-12.24***	-11.43***	-6.24***	-6.40***	-4.86***	-3.59***
	(1.43)	(0.85)	(0.67)	(1.08)	(0.64)	(0.70)
Average district-level test scores[b]	5.42***	3.09***	3.81***			
	(0.86)	(0.58)	(0.40)			
Student enrollment[b]	-3.63***	-3.41***	-3.39***	-0.42	-0.15	-0.23
	(0.31)	(0.20)	(0.18)	(0.50)	(0.38)	(0.39)
School spending per student[b]	-4.90***	-0.01	0.42**			
	(0.67)	(0.28)	(0.21)			
ESSER funding per student[b]	-0.88	-1.53***	-1.59***			
	(0.66)	(0.46)	(0.50)			
School type and grade controls	✔	✔	✔	✔	✔	✔
State FE		✔			✔	
County FE			✔			✔
R-squared	0.11	0.22	0.28	0.04	0.10	0.19
# of counties	2,833	2,833	2,833	1,489	1,489	1,489
# of schools	51,577	51,577	51,577	11,280	11,280	11,280
# of school districts	11,119	11,119	11,119			

Notes: Each column reports coefficients from a weighted OLS regression on the public (panel (a)) and private (panel (b)) school samples, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on average school EIPL for the period from September 2020 to May 2021. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. The coefficient estimates for the affluence measures, denoted by [a], are the results of separate regressions with each one of the measures in combination with the other variables below. The coefficient estimates for the other regressors denoted by [b] are the result of regressions where the three affluence measures are included jointly.

Regression results presented in Figure 3 Table C5 presents the estimates reported in Figure 3 of the main text. In order to understand better the role of the county-level regressors, Table C6, presents the results of introducing these regressors in isolation from each other. We begin in Column (1) with a regression on the local and school covariates and on party affiliation of the state governor and the Republican vote share in the 2020 presidential election, while controlling for pre-pandemic ICU bed

capacity, two-week lagged county COVID case and death rates, dummies for various non-pharmaceutical interventions, maximum weekly temperature in the county, population density in the county, and county's rural/urban continuum codes. Column (2) focuses on the effect of vaccination rates for the 2021 part of the sample. Note that the reason why the R-squared remains in the same ballpark is due to the county-level controls. The explanatory power of vaccination rates per se is weak, indicating that vaccination rates leave most of the variation in county-week EIPL unaccounted. Column (3) adds teacher unionization rates and the index of the costs of hiring PK-12 educators. Last, Column (4) adds all five variables together, as in Column (3) of Table C5. Foremost, Table C6 shows that there is little interrelation between the county-level regressors of interest, perhaps with exception of COVID vaccination rates (their role is enhanced in Column (4)) and the local index of costs of hiring PK-12 educators (its role is dampened in Column (4)).

Table C6: Accounting for systematic geographical differences

Dependent variable	Effective in-person learning (EIPL)			
	(1)	**(2)**	**(3)**	**(4)**
Republican governor	12.26***			11.65***
	(1.29)			(1.19)
Share of 2020 Republican voters	11.53***			11.45***
	(0.99)			(1.17)
Lagged COVID vaccination rate		8.51***		12.45***
		(0.86)		(0.75)
Teacher unionization rate			-6.43***	-5.35***
			(1.13)	(0.93)
Local index of costs of hiring PK-12 educators			-8.68***	-3.83***
			(1.05)	(1.10)
Local affluence	✔	✔	✔	✔
School type and grade controls	✔	✔	✔	✔
Other school/district variables	✔	✔	✔	✔
County health, pop. characteristics, weather, NPIs	✔	✔	✔	✔
R-squared	0.21	0.19	0.20	0.23
# of counties	2,774	2,774	2,774	2,774
# of schools	51,007	51,007	51,007	51,007
# of school districts	10,973	10,973	10,973	10,973

Notes: Each column reports coefficients from a weighted OLS regression on the public school sample, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on EIPL for the period from September 2020 to May 2021. Local affluence variables consist of zip-level mean household income, share of adults with College or higher education, share of dual-headed household with children. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. Other school/district variables consist of district-level test scores, school size, school spending per student and ESSER funding per student. County health, pop.characteristics, weather, NPIs consist of pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, population density in the county, dummies for rural-urban continuum codes, maximum weekly temperature in the county, and dummies for various non-pharmaceutical interventions.

In results not reported here, we find that: (i) using the Republican vote share in the 2016 presidential election among our proxies for the general stance towards reopening schools barely change the results, which is unsurprising given the strong persistence in county-level Republican vote shares in the 2016 and 2020 presidential elections; (ii) changing the number of time lags used to measure COVID vaccination, infection and death rates matters for the coefficient on vaccination rates, while suggesting that a 2-weeks

Table C5: Regression results for Figure 3

Dependent variable	Effective in-person learning (EIPL)		
	(1)	**(2)**	**(3)**
Zip-level mean household income[a]	-4.98***	-5.19***	-0.34
	(0.49)	(0.59)	(0.40)
Zip-level share of College educated[a]	-6.87***	-7.65***	-2.57***
	(0.48)	(0.49)	(0.46)
Zip-level share of dual-headed households[a]	-3.11***	-3.21***	-0.39
	(0.71)	(0.75)	(0.38)
School share of non-white students[b]	-17.74***	-12.24***	-4.37***
	(1.71)	(1.43)	(0.88)
Average district-level test scores[b]		5.42***	3.47***
		(0.86)	(0.60)
Student enrollment[b]		-3.63***	-3.23***
		(0.31)	(0.23)
School spending per student[b]		-4.90***	-0.70**
		(0.67)	(0.34)
ESSER funding per student[b]		-0.88	-2.20***
		(0.66)	(0.49)
Republican governor			11.65***
			(1.19)
Share of 2020 Republican voters			11.45***
			(1.17)
Lagged COVID vaccination rate			12.45***
			(0.75)
Teacher unionization rate			-5.35***
			(0.93)
Local index of costs of hiring PK-12 educators			-3.83***
			(1.10)
School type and grade controls	✔	✔	✔
County health, pop. characteristics, weather, NPIs			✔
R-squared	0.09	0.11	0.23
# of counties	2,833	2,833	2,774
# of schools	51,577	51,577	51,007
# of school districts	11,119	11,119	10,973

Notes: Each column reports coefficients from a weighted OLS regression on the public school sample, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on EIPL for the period from September 2020 to May 2021. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. County health, pop.characteristics, weather, NPIs consist of pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, population density in the county, dummies for rural-urban continuum codes, maximum weekly temperature in the county, and dummies for various non-pharmaceutical interventions. The coefficient estimates for the affluence measures, denoted by [a], are the result of separate regressions with each one of the measures in combination with the other variables below. The coefficient estimates for the other regressors denoted by [b] are the result of regressions where the three affluence measures are included jointly.

lag is appropriate;[20] restricting the sample to counties with at least 10 public schools, which reduces the sample size almost threefold, leave the results mainly unchanged.

Share of nonwhite students Our results show that the share of non-white students among the schools' body is associated with lower EIPL during 2020-21. In Table C7, we present additional results regarding this finding by replacing the share of non-white students by shares of students of different races or ethnicities. Columns (1), (3) and (5) of panel (a) repeat the results from Table C5 for reference, while columns (2), (4) and (6) show the estimates for the different races/ethnicities; Panel (b) performs a similar analysis for private schools. Table C7 shows that the coefficient on the school share of non-white students is mainly driven by the share of Hispanic students, and to a much lesser extent by that of Black students. We hypothesize that the important role of the share of Hispanic students is related to the ethnic makeup of schools in states of the South-western part of the country (California, New Mexico), where as shown in Figure 1 EIPL has remained very low throughout the school year of 2020-21.

District vs. school-level test scores Our main regression uses pre-COVID test scores at the level of school districts for reasons of data availability (see Subsection A.5). In Table C8, we show effects of using school-level test scores, which are available for about 38,000 schools in our dataset vs. 52,000 for the district-level test scores. Columns (1) and (2) repeat results from Columns (2) and (3) of C5 for reference. As can be seen the addition of county-level controls make the coefficient on test scores shrink from to 5.6 to 3.5. Then, in Columns (3) and (4), we restrict the regression to schools for which we also have school-level test scores available from SEDA. The coefficient on district-level test scores goes up by about 1 unit. Finally in Columns (5) and (6), we replace district-level test scores by school-level test scores. The magnitude of the effects of test scores changes, but not by much. The effects of the addition of county-level controls is similar to that obtained under our main regression.

Other indicators of local affluence Table C9 presents additional results of introducing other indicators of local affluence. The indicators considered are: the Opportunity Atlas (OA)'s measure of upwards mobility as measured by the mean household income rank for children whose parents were at the 25th percentile of the national income distribution, where incomes for children are measured as mean earnings in 2014-2015 when they were between the ages 31-37; OA's fraction of children born in 1978-1983 birth cohorts with parents at the 25th percentile of the national income distribution who were incarcerated in 2010; OA's measured average rent for two-bedroom apartments in 2015; and EDGE's school neighborhood poverty estimates. Columns (1), (3), (5) and (7) run a quasi-univariate regression, where only the local affluence measure of interest is included in the regression along with the school share of non-white students and school type and grade controls. These regressions are thus similar to those reported in Table C2. Columns (2), (4), (6) and (8) of Table C9 add all the covariates included in our main regression, and are therefore comparable to results shown in Table C5.

Table C9 confirms that an inverse relationship between EIPL and affluence holds with regards to incarceration rates and neighborhood poverty: public schools in areas with *higher* rates of incarceration and public schools in *poorer* neighborhoods (i.e. school with a higher index) provided on average *lower* EIPL during the pandemic. The relation with housing prices (as captured by average rents of two-bedroom apartments) is also consistent with out main results. The role of upward mobility is more difficult to fathom because there may not be a clear correlation between this indicator and the measures of local affluence considered in our main analysis. The coefficient is negative in the quasi-univariate regression, then turns positive when introducing the controls, but in absolute term the effect on EIPL

[20] When using contemporaneous values of COVID vaccination, infection and death rates, the effect of vaccination rates is less pronounced – it is reduced by half –, and infection rates exert a negative effect on EIPL. On the other hand, with a lag of one month, the effect of COVID vaccination rates is very close to the baseline estimates. It is unclear how best to measure the dynamic relationships between the COVID health variables and EIPL, but in all instance the regressions show that the vaccination campaign is positively related to EIPL in a statistically and economically significant way.

Table C7: Robustness check: racial makeup of schools

Dependent variable	Effective in-person learning (EIPL)											
	(a) Public schools						**(b) Private schools**					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Non-white students	-17.74***		-12.24***		-4.37***		-6.37***		-6.40***		-3.25***	
	(1.71)		(1.43)		(0.88)		(1.08)		(1.08)		(0.75)	
Black students		-2.40***		-0.91***		-0.77**		-1.17***		-1.18***		-0.86***
		(0.34)		(0.35)		(0.32)		(0.34)		(0.35)		(0.31)
Hispanic students		-9.48***		-6.41***		-1.86***		-1.66***		-1.66***		-0.84***
		(1.00)		(0.78)		(0.58)		(0.51)		(0.51)		(0.30)
American indian students		-0.07***		-0.02		-0.10***		-0.00		-0.00		-0.00
		(0.02)		(0.02)		(0.03)		(0.00)		(0.00)		(0.00)
Hawaiian students		-0.07***		-0.02		0.10		-0.00**		-0.00**		-0.00
		(0.02)		(0.02)		(0.07)		(0.00)		(0.00)		(0.00)
Asian students		-1.92***		-1.73***		-0.65***		-1.65***		-1.65***		-0.57**
		(0.14)		(0.14)		(0.10)		(0.27)		(0.27)		(0.25)
Students with 2 or more races		-2.53***		-2.19***		-0.61*		-1.74***		-1.73***		-0.74***
		(0.47)		(0.44)		(0.32)		(0.30)		(0.29)		(0.23)
Local affluence controls	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
School type and grade controls	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Other school/district variables			✔	✔	✔	✔			✔	✔	✔	✔
County controls					✔	✔					✔	✔
R-squared	0.09	0.09	0.11	0.11	0.23	0.23	0.04	0.04	0.04	0.04	0.19	0.19
# of counties	2,833	2,833	2,833	2,833	2,774	2,774	1,489	1,489	1,489	1,489	1,489	1,489
# of schools	51,577	51,577	51,577	51,577	51,007	51,007	11,280	11,280	11,280	11,280	11,280	11,280
# of school districts	11,119	11,119	11,119	11,119	10,973	10,973						

Notes: Each column reports coefficients from a weighted OLS regression on the public (panel (a)) and private (panel (b)) school samples, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on EIPL for the period from September 2020 to May 2021. Local affluence variables consist of zip-level mean household income, share of adults with College or higher education, share of dual-headed household with children. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. Other school/district variables consist of district-level test scores, school size, school spending per student and ESSER funding per student. County controls consist of pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, population density in the county, dummies for rural-urban continuum codes, maximum weekly temperature in the county, and dummies for various non-pharmaceutical interventions.

Table C8: Robustness check: District-level vs. school-level test scores

Dependent variable	Effective in-person learning (EIPL)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Average district-level test scores	5.61***	3.48***	6.46***	4.10***		
	(0.74)	(0.61)	(0.77)	(0.67)		
Average school-level test scores					4.18***	2.55***
					(0.73)	(0.49)
Student enrollment		-3.20***		-4.51***		-4.56***
		(0.24)		(0.38)		(0.38)
School spending per student		-0.76**		-1.63***		-1.28***
		(0.30)		(0.38)		(0.39)
ESSER funding per student		-2.06***		-1.74***		-3.21***
		(0.46)		(0.54)		(0.50)
School-level test scores available			✔	✔	✔	✔
Local affluence	✔	✔	✔	✔	✔	✔
School type and grade controls	✔	✔	✔	✔	✔	✔
County controls		✔		✔		✔
R-squared	0.11	0.28	0.08	0.25	0.08	0.25
# of schools	51,577	51,577	37,651	37,651	37,651	37,651
# of school districts	11,119	11,119	8,971	8,971	8,971	8,971
# of counties	2,833	2,833	2,729	2,729	2,729	2,729

Notes: Each column reports coefficients from a weighted OLS regression on the public school sample, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2.4. The regressions are estimated on EIPL for the period from September 2020 to May 2021. Local affluence variables consist of zip-level mean household income, share of adults with College or higher education, share of dual-headed household with children. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. County controls consist of pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, population density in the county, dummies for rural-urban continuum codes, maximum weekly temperature in the county, and dummies for various non-pharmaceutical interventions.

is quite limited. Also, in Table C9 as in the main regression, the estimates of race become smaller after adding geographic controls, reflecting the fact that schools in suburban and town/rural areas provided on average higher EIPL, and suburban and town/rural areas are on average less affluent, have a smaller share of college-educated households, and have a smaller population of non-white students.

Table C9: Robustness check: Other indicators of local affluence

Dependent variable	Effective in-person learning (EIPL)							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
OA upward mobility of 25th pctile children	-2.84***	1.24***						
	(0.64)	(0.44)						
OA incarceration rate of 25th pctile children			3.30***	0.60**				
			(0.48)	(0.27)				
OA rent for two-bedroom apartments					-7.70***	-1.83***		
					(0.51)	(0.50)		
EDGE school neighborhood poverty index							5.73***	0.52*
							(0.36)	(0.30)
School share of non-white students	-19.74***	-3.62***	-19.56***	-3.82***	-15.15***	-3.39***	-20.37***	-3.90***
	(1.51)	(0.94)	(1.37)	(0.93)	(1.15)	(0.90)	(1.38)	(0.96)
Local affluence		✔		✔		✔		✔
School type and grade controls	✔	✔	✔	✔	✔	✔	✔	✔
Other school/district variables		✔		✔		✔		✔
County controls		✔		✔		✔		✔
R-squared	0.07	0.23	0.07	0.23	0.09	0.23	0.08	0.23
# of counties	2,833	2,774	2,833	2,774	2,833	2,774	2,833	2,774
# of schools	51,577	51,007	51,577	51,007	51,577	51,007	51,577	51,007
# of school districts	11,119	10,973	11,119	10,973	11,119	10,973	11,119	10,973

Notes: Each column reports coefficients from a weighted OLS regression on the public school sample, with standard errors clustered at the county level in parentheses and school weights calculated as explained in Appendix A.2. The regressions are estimated on EIPL for the period from September 2020 to May 2021. Local affluence variables consist of zip-level mean household income, share of adults with College or higher education, share of dual-headed household with children. The school type fixed effects consists of indicators for charter school and non-charter school, and the school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school for both samples. Other school/district variables consist of district-level test scores, school size, school spending per student and ESSER funding per student. County controls consist of pre-pandemic ICU bed capacity, two-week lagged county COVID case and death rates, population density in the county, dummies for rural-urban continuum codes, maximum weekly temperature in the county, and dummies for various non-pharmaceutical interventions. OA (Opportunity Atlas) upward mobility is the mean household income rank for children whose parents were at the 25th percentile of the national income distribution where incomes for children are measured as mean earnings in 2014-2015 when they were between the ages 31-37; OA incarceration rate is the fraction of children born in 1978-1983 birth cohorts with parents at the 25th percentile of the national income distribution who were incarcerated in 2010. See Appendix A.5 for details.

C.4 Description of school-level regression variables

In Section 5, we measure the effects of local and school/district variables on EIPL by looking at the effects of going from the 25th to the 75th percentile of the distribution of a variable. Table C10 presents descriptive statistics for the sample of public and private schools of our analysis. In particular, observe that private schools are on average located in more affluent areas, have a lower proportion of non-white students and are much smaller than public schools in terms of student enrollment.

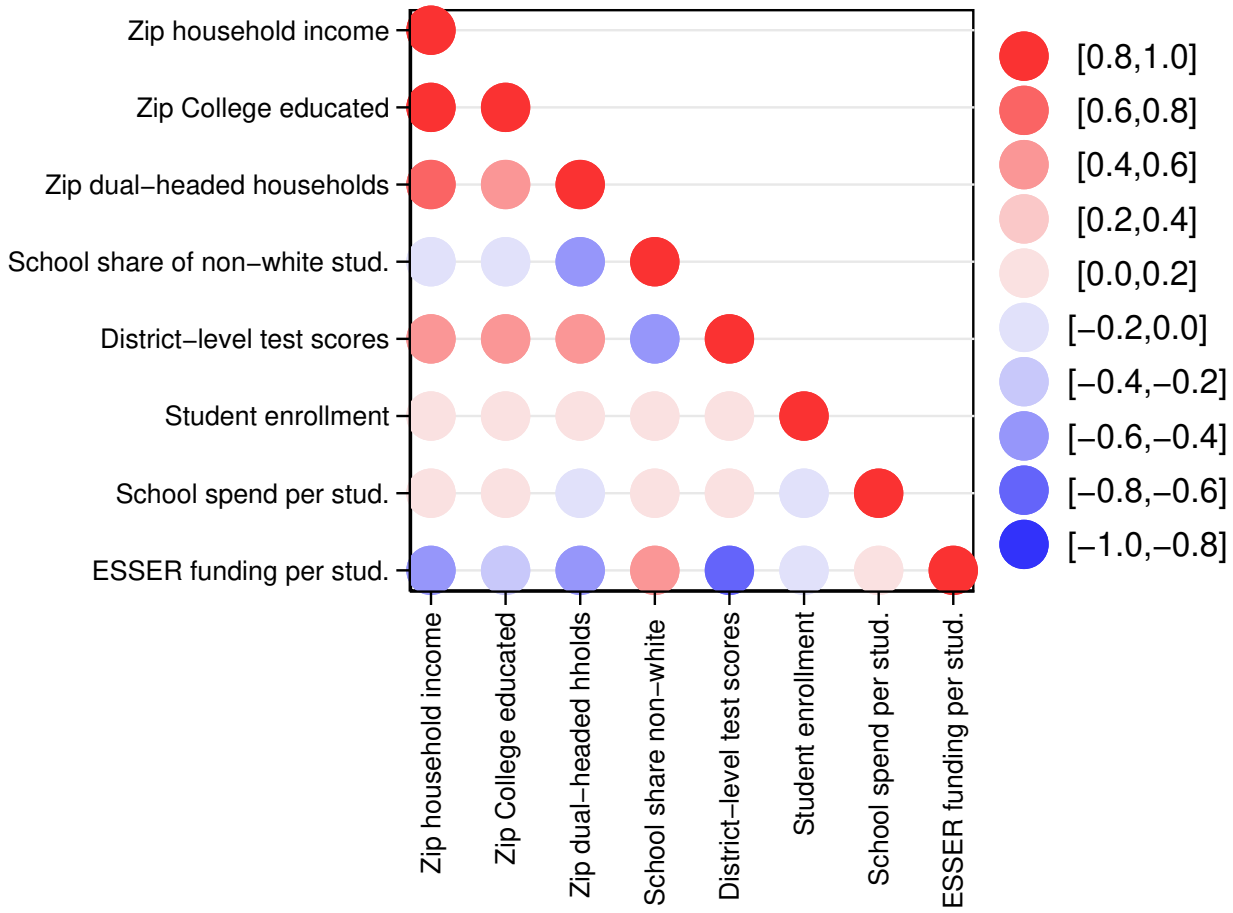
Table C10: Descriptive statistics of the school-level regression variables

	Mean	St. Dev.	Percentile			Min.	Max.
			25th	50th	75th		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
				(a) Public schools			
Zip-level mean household income	75,027	29,719	55,774	67,287	86,209	7,770	432,067
Zip-level share of college educated	0.27	0.15	0.16	0.23	0.35	0.01	0.89
Zip-level share of dual-headed households	0.70	0.14	0.62	0.71	0.80	0.00	1.00
School share of non-white students	0.47	0.32	0.18	0.42	0.76	0.00	1.00
Average district-level test scores	0.00	0.32	-0.20	0.00	0.20	-1.25	1.25
Student enrollment	615	469	338	502	733	6	8,327
School spending per student	12,321	4,484	9,417	11,339	14,083	112	49,957
ESSER funding per student	3,229	2,391	1,534	2,806	4,193	0	30,189
				(b) Private schools			
Zip-level mean household income	86,503	39,765	59,828	75,054	103,386	22,512	397,509
Zip-level share of college educated	0.35	0.18	0.21	0.31	0.47	0.02	0.90
Zip-level share of dual-headed households	0.70	0.15	0.62	0.72	0.81	0.05	0.99
School share of non-white students	0.37	0.31	0.12	0.28	0.56	0.00	1.00
Student enrollment	211	260	42	138	272	6	4,312

Notes: The table reports the mean, standard deviation (St. Dev.), the 25th, 50th, 75th percentiles, and the minimum (Min.) and maximum (Max) values of the right-hand side variables of the school-level regressions. All statistics are computed with the school weights calculated as explained in Appendix A.2.

To complement panel (a) of Table C10, Figure C6 presents the correlations between the regressors. As the top left corner shows, the first two affluence measures (mean household income and share of adults with College or higher education) are highly correlated with each other, with a correlation coefficient of 0.83. The share of dual-headed households with children is also positively related with these measures: the correlation coefficients are respectively at 0.60 and 0.47. The school share of non-white students is negatively related to the affluence measures. The correlations range from -0.46 (correlation with the share of dual-headed households with children) to -0.05 (correlation with zip-level share of adults with College or higher education). The other interesting correlations in this figure are those between ESSER funding per student and the affluence measures. The correlations are negative and within the -0.58 to -0.35 range. Since the share of non-white students is negatively related to the affluence measures, it is positively correlated with ESSER funding per student (correlation of 0.43). There is also a strong inverse relation between ESSER funding per student and pre-COVID district level test scores, with a coefficient correlation of -0.69.

Figure C6: Cross-correlations of the school-level regression variables



Notes: The figure show the cross-correlations of the variables used in the school-level regressions. Correlations are computed with the school weights calculated as explained in Appendix A.2.

References

Dena Bravata, Jonathan H Cantor, Neeraj Sood, and Christopher M Whaley. Back to school: The effect of school visits during COVID-19 on COVID-19 transmission. *NBER Working Paper 28645*, 2021.

MIT Election Data and Science Lab. County Presidential Election Returns 2000-2020, 2018. URL https://doi.org/10.7910/DVN/VOQCHQ.

Erin M Fahle, Benjamin R Shear, Demetra Kalogrides, Sean F Reardon, Richard DiSalvo, and Andrew D Ho. Stanford Education Data Archive (Version 4.1), 2021. URL http://purl.stanford.edu/db586ns4974.

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, 2021.