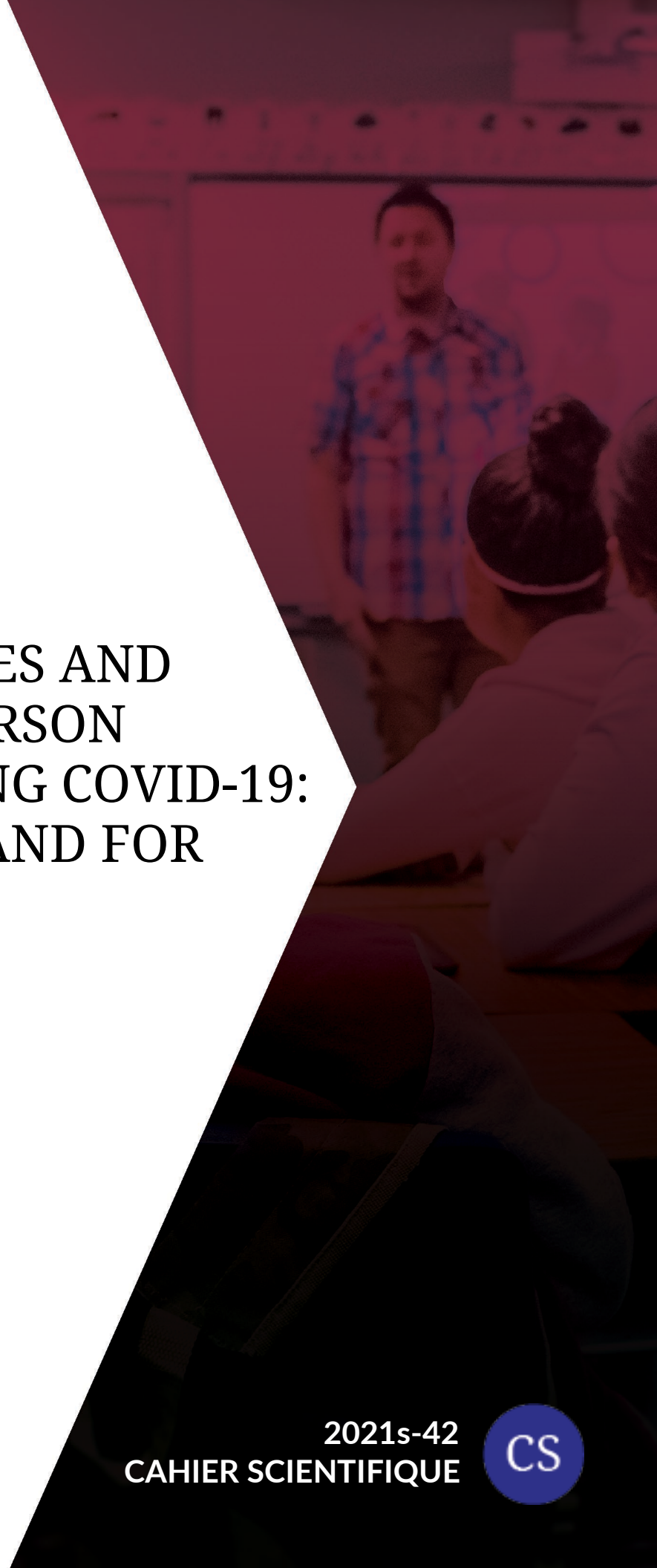


SCHOOL CLOSURES AND EFFECTIVE IN-PERSON LEARNING DURING COVID-19: WHEN, WHERE, AND FOR WHOM

ANDRÉ **KURMANN**
ETIENNE **LALÉ**

2021s-42
CAHIER SCIENTIFIQUE

CIRANO
Knowledge into action

Center for Interuniversity Research and Analysis on Organizations

The purpose of the **Working Papers** is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Les cahiers de la série scientifique visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

ISSN 2292-0838 (online version)

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

André Kurmann et Etienne Lalé

Résumé

Nous apparions des données de mobilité obtenus à partir de téléphones portables à un échantillon représentatif de près de 70 000 écoles aux États-Unis, et combinons ces données avec des informations sur les modes d'apprentissage scolaire pour construire une mesure de l'apprentissage effectivement réalisé en présentiel (EIPL) pendant la pandémie de COVID-19. Nous augmentons ensuite ces données avec plusieurs bases de données administratives afin de documenter les différences d'EIPL dans le temps, selon les régions et en fonction des caractéristiques individuelles des écoles. Nous obtenons trois résultats principaux. Premièrement, alors que l'EIPL a chuté en deçà de 20 % par rapport à son niveau pré-pandémique au printemps 2020 dans toutes les régions des États-Unis, l'EIPL a ensuite fortement varié au cours de l'année scolaire 2020-2021, atteignant plus de 80 % dans certaines villes du sud alors qu'il se maintenait en deçà de 20 % dans certaines villes de la côte ouest. Deuxièmement, une part substantielle de cette variation est expliquée par les caractéristiques observables des écoles : (i) les écoles publiques ont fourni en moyenne moins d'EIPL que les écoles privées ; (ii) les écoles situées dans des localités plus riches et plus instruites et les écoles comptant une plus grande proportion d'élèves non blancs ont fourni en moyenne moins d'EIPL ; et (iii) les écoles publiques ayant des dépenses par élève plus élevées avant la pandémie, ayant reçu un montant d'aide d'urgence aux écoles élémentaires et secondaires (ESSER) par élève plus importante, et les écoles ayant un plus grand nombre d'élèves ont fourni en moyenne moins d'EIPL. Troisièmement, l'association négative de l'EIPL avec la richesse, l'éducation et les dépenses scolaires pré-pandémiques est due en grande partie à des différences régionales systématiques qui sont corrélées aux préférences politiques. En revanche, l'association négative de l'EIPL avec la part d'élèves non blancs d'une école et le financement ESSER persiste au sein même des comtés et en contrôlant la richesse et l'éducation locales. Ces tendances sont importantes pour comprendre les facteurs qui ont conduit aux disparités dans les fermetures d'écoles et pour évaluer l'impact de la perte d'apprentissage en présentiel pendant la pandémie sur le niveau d'éducation futur, les inégalités des revenus et la croissance économique.

Mots-clés: COVID-19, Fermetures et réouvertures d'écoles, Apprentissage efficace en présentiel, Inégalité

JEL Codes/Codes JEL: E24, I24

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

November 15, 2021

Abstract

We combine cell phone data on foot-traffic to a highly representative sample of almost 70,000 schools in the U.S. with information on school learning modes to estimate a measure of effective in-person learning (EIPL) during the COVID-19 pandemic. We then match the data with various administrative records to document differences in EIPL over time, across regions, and by individual school characteristics. We find three main results. First, while EIPL dropped to below 20% of its pre-pandemic level across all regions of the U.S. during Spring 2020, EIPL varied widely during the 2020-21 school year, ranging from less than 20% in some cities on the West Coast to more than 80% in some cities in the South. Second, a substantial part of this variation is accounted for by observable school characteristics: (i) public schools provided on average less EIPL than private schools; (ii) schools in *more* affluent and educated localities and schools with a *larger* share of non-white students provided on average *lower* EIPL; and (iii) public schools with *higher* pre-pandemic spending per student, *higher* district-level Elementary and Secondary School Emergency Relief (ESSER) funding per student, and *larger* student enrollment provided on average *lower* EIPL. Third, the negative association of EIPL with affluence, education and pre-pandemic school spending is driven in large part by systematic regional differences that are correlated with political preferences. In contrast, the negative association of EIPL with a school's share of non-white students and ESSER funding persists even within counties and controlling for local affluence and education. These patterns are important for our understanding of the factors that led to the large disparities in school closures and the impact of in-person learning loss during the pandemic on future educational attainment, income inequality, and economic growth.

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 Burbio, Return to Learn (in particular Nat Malkus), and Safegraph for generously making their data available. 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, evidence is emerging that remote instruction constituted at best an imperfect substitute to in-person instruction and led to substantial learning losses and social-emotional harm with possibly large adverse long-term effects, especially for students from disadvantaged backgrounds.[1,2] A key input for analyzing these consequences is the availability of precise and consistent estimates of effective in-person learning at the individual school level. The goal of this paper is to provide such estimates and analyze the extent to which differences in effective in-person learning during the pandemic reflect school characteristics as opposed to systematic geographic differences that are not directly related to schools.

We match anonymized cell phone data from Safegraph on foot traffic to the population of public and private schools from the National Center for Education Statistics (NCES). We then map changes in foot traffic during the pandemic to information on school learning mode by Burbio and Return2Learn to construct a measure of effective in-person learning (EIPL) at the individual school level. After removing low-quality matches and records with sparse or noisy data, we end up with a dataset of weekly EIPL for almost 70,000 schools that is highly representative of the universe of schools in the U.S. We organize our analysis of this dataset into three parts.

In the first part, we provide new evidence on the disparities in in-person learning 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 and Texas such as Jacksonville or Houston, 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 less than 25% over the same time period.

Concurrently, we find that even for narrow geographical areas, there are sizable differences in EIPL across individual schools. For instance, the mean interquartile range of average EIPL across schools within a given county amounts to 14%. This suggests that some schools returned to in-person learning much more quickly than others not just because of systematic regional differences, but also because of school-specific factors that apply similarly across the country.

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

1. Public schools provided on average 10% less EIPL than private schools, with public charter schools

[1] See Bravata et al. [2021], Chernozhukov et al. [2021b], Ertem et al. [2021], and the references therein for systematic assessments of the effect of school closures on subsequent COVID-19 infections.

[2] See Curriculum Associates [2021], Dorn et al. [2021], Kogan and Lavertu [2021], Lewis et al. [2021] and references therein for evidence on the learning losses and social-emotional harm on students caused by the pandemic. See Agostinelli et al. [2020], Fuchs-Schündeln et al. [2021] or Jang and Yum [2020] for simulations of the adverse long-term consequences of learning losses in the form of lower educational attainment and reduced earnings potential, especially for students from less affluent and less educated backgrounds.

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 substantially lower in more affluent and more educated localities, and for schools with a larger share of non-white students.

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

In the third part of the analysis, we ask to what extent these results are driven by systematic regional differences that are not directly related to schools. We find that once we control for county fixed effects, the negative association of EIPL with local affluence, education, and school spending per student largely disappears. We show that this result is not driven by whether the school is located in a city, suburb, or town/rural area. In other words, schools in less affluent parts of the country with lower education and lower spending per student reopened for in-person learning more quickly. We show that this result is to a substantial part explained by the presence of a Republican governor and the margin by which the county voted Republican in the 2020 presidential election. COVID vaccination rates also predict higher county EIPL but overall have little explanatory power, whereas the COVID health situation as measured by county-level case and death rates is unrelated to EIPL. The results suggest that school reopenings were driven in large part by policies and general attitudes towards school openings that align with local political preferences.

In contrast, we find that the negative association of EIPL with a school's share of non-white students and ESSER funding persists even within counties and after controlling for local affluence and other school characteristics. A school at the 75th percentile of the share of non-white student distribution provided on average 7% lower EIPL during the 2020-21 school year than a school at the 25th percentile; and a school at the 75th percentile of the ESSER funding distribution provided on average 3% lower EIPL than a school at the 25th percentile. These results are striking 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 probably closest to ours are Bravata et al. [2021], Chernozhukov et al. [2021a], and Ertem et al. [2021] who use Safegraph, respectively Burbio data to predict the effect of school reopenings on infection rates and deaths, Dee et al. [2021] who use Burbio data to study the effect of school closings on school enrollment, and Parolin and Lee [2021] who use Safegraph data to estimate school closures and analyze their relationship with demographic and socio-economic composition of a school's student body. 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 foot traffic or learning mode data. While mobility data as provided for example by Safegraph covers a large and representative share of all schools in the U.S., it is not clear what a given decline in foot traffic to a school represents in terms of lost in-person learning.[3] Furthermore, attributing cell phone

[3]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

data to a particular location is inherently challenging, thus raising questions about the reliability of the information. The learning mode data from Burbio and Return2Learn, in turn, covers only public schools and provides only averages at the county-, respectively the district level. This limits the analysis of the relationship of EIPL with individual school characteristics and local conditions surrounding the school. Even more importantly, learning modes are reported as percentages of student time spent in traditional, hybrid, or virtual learning mode. Yet the extent to which traditional learning is fully spent in person and how much hybrid learning contributes to in-person learning is unclear. By combining the two data sources, we attempt to address these limitations and perform important validity checks about the quality of the different datasets.

Second, we relate our estimate of EIPL to a host of indicators measuring not just the demographic and socio-economic characteristics of a school's student body, but also school funding and local conditions surrounding the school. We document striking inverse associations of EIPL with local income, education, and school funding 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. The results raise critical questions about the response of schools in different parts of the country to the pandemic, and the patterns of EIPL that we uncover are important for our understanding of the impact of in-person learning loss during the pandemic on future educational attainment, income inequality, and economic growth. See Fuchs-Schündeln et al. [2021] for an application in this respect. 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 in slowing down the spread of COVID-19.

2 Data sources, matching, and sample restrictions

This section describes the different datasets used to construct our EIPL measure. We start with an overview of the Safegraph data, how we match it to administrative information on individual schools from the NCES, and how we construct changes in school visits for a restricted sample that aims to reduce measurement error. Then we provide an overview of the Burbio and Return2Learn data and compare their learning mode estimates.

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. based on anonymized cell phone pings from over 40 million devices.[4] From this large set of POIs we retain all places with North American Industry Classification System (NAICS) code 611110 ("Elementary and Secondary Schools") that have weekly foot traffic data (visits). We then match these places by school 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 in the NCES, we lose about 22,000 schools, but the matched sample remains highly representative of the

remained in remote-learning mode) or if students get dropped off and picked up by parents instead of using buses, then changes in Safegraph foot traffic data alone would overestimate EIPL during the pandemic.

[4] A cell phone ping is the process of determining the current location of a cell phone.

universe of schools in terms of demographic and geographic makeup. Appendix A provides details of the matching procedure and reports statistics on the representativeness of the sample.

2.2 Constructing changes in school visits

The Safegraph data provides weekly visit counts for each school by dwell time. There are $D = 7$ dwell time intervals (less than 5, 5 to 10, 11 to 20, 21 to 60, 61 to 120, 121 to 240, more than 240 minutes), Denoting weekly visits counts as $v_{j,t}(d)$ for $d = 1, \ldots, D$, the total visits count for school j in week t is $v_{j,t} = \sum_{d=1}^{D} v_{j,t}(d)$.

As shown in Appendix A, prior to the pandemic, both aggregate total visit counts and aggregate visits longer than 240 minutes per day decline markedly during the weeks of Thanksgiving, Christmas, and Summer break. And consistent with the public health emergency on March 13, 2020, both visits series drop precipitously during the week of March 15 to March 21, 2020 and remain substantially lower thereafter.

At the same time, we find that aggregate visits are trending upward prior to the pandemic, which is in large part driven by the increase in cell phone devices covered by Safegraph. Furthermore, closer inspection of the data reveals that visit counts are subject to substantial variation over time, both for individual schools and for in the aggregate. While some of these high-frequency variations are due to holidays and other breaks, another part of the variations reflects the inherent difficulty of attributing cell phone pings to a particular POI.

To address these measurement issues, we normalize weekly visits with the weekly count of Safegraph devices at the county level, and then construct the change in school visits as the percent difference in dwell-time weighted, normalized visits relative to the average over a reference period preceding the pandemic; i.e.

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

where $n_{c(j),t}$ denotes the normalization by the count of devices during week t in county $c(j)$ in which school j is located; and $\omega_j(d) = \frac{\sum_{t=t_0}^{t_1} v_{j,t}(d)}{\sum_{t=t_0}^{t_1} v_{j,t}}$ measures the importance of visits of dwell time d for school j during reference period $t = t_{-1}, \ldots, t_0$ (beginning of November 2019 through the end of February 2020, excluding the weeks of Thanksgiving, Christmas and New Year). The percent change in school visits is then defined as

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

where $\widetilde{v}_{j,0} = \frac{1}{t_0 - t_{-1} + 1} \sum_{t=t_{-1}}^{t_0} \widetilde{v}_{j,t}$ is the average over the reference period.

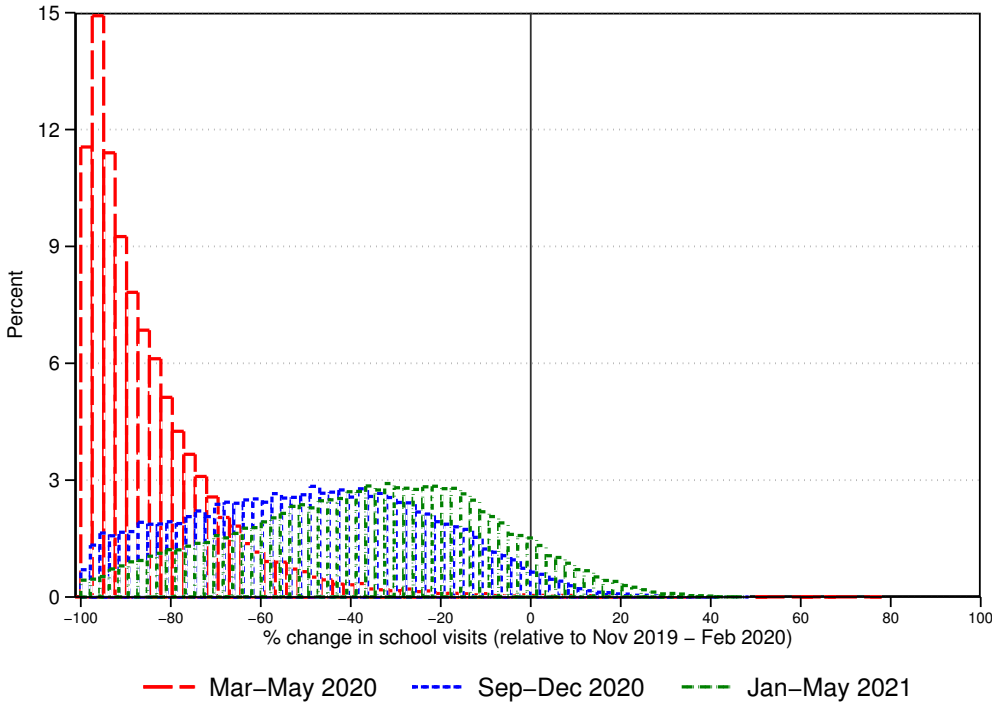
In an effort to further reduce measurement error, we drop schools with sparse or noisy visit data and apply weights to ensure that the remaining sample of roughly 70,000 schools remains representative of the full sample of schools in the U.S. See Appendix A for details. Moreover, we top-code $\Delta\widetilde{v}_{j,t}$ at 100%, and 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 $\Delta\widetilde{v}_{j,t+1}$.[5]

[5]This adjustment implements the assumption that during the school year 2020-21, schools did not reopen for only one week at a time.

Compared to Bravata et al. [2021], Chernozhukov et al. [2021b] and Parolin and Lee [2021] who consider year-over-year changes in raw visits, our construction of dwell-time weighted normalized changes in school visits relative to an average over a reference period prior to the pandemic has the advantage that it is not influenced by the upward trend due to increased device coverage and that it reduces the influence of holidays that fall on different weeks across years as well as other idiosyncratic variations. Furthermore we apply apply stringent sample selection criteria for schools with sparse or noisy data. Appendix B provides further discussion.

Figure 1 presents histograms of the distribution of the average change in school visits $\Delta\widetilde{v}_{j,t}$ during three subperiods. The figure clearly shows that relative to the pre-pandemic baseline, school visits declined massively during March-May 2020, and were still significantly lower during September-December 2020 as well as during January-May 2021. At the same time, the dispersion in visit changes across schools in each of the subperiods is large, reflecting the differences in school closings during the pandemic. See Appendix A for further results on the distribution of school visit changes.

Figure 1: Distribution of changes in school visits for selected subperiods

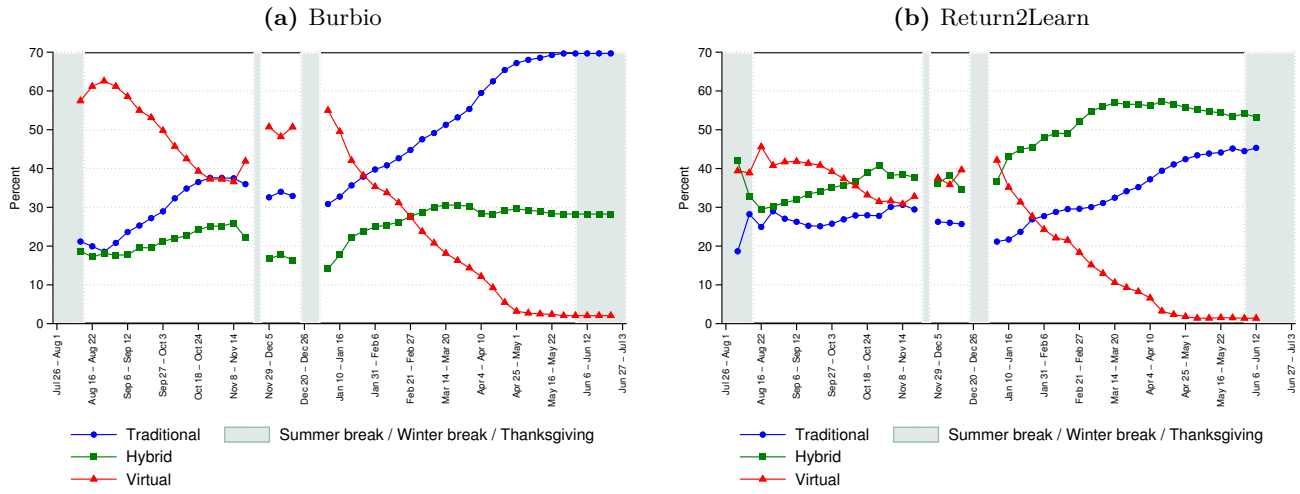


Notes: The figure show the distribution of average dwell-weighted normalized changes in school visits during March-May 2020, September-December 2020, and January-May 2021, relative to average pre-pandemic (November 2019 through February 2020) school visits.

2.3 Burbio and Return2Learn learning mode data

Burbio is a private company that collects data on school learning mode from publicly available sources for 1,200 public school districts representing 47% of U.S. K-12 student enrollment in over 35,000 schools in all 50 states. The data is aggregated to the county level and primarily used for commercial purposes, but the company also uses the data to publish a weekly School Opening Tracker and generously agreed to

Figure 2: 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.

share the data with us and other researchers. The data consists of weekly indicators from August 2020 to June 2021 of the share of public school students engaged in either "Traditional", "Hybrid", or "Virtual" learning mode, where Hybrid means that students attend school 2-3 days per week in-person.[6]

Return2Learn is another school learning mode tracker constructed by researchers from the American Enterprise Institute and the College Crisis Initiative center at Davidson College, and is generously shared with researchers. Like Burbio, the Return2Learn data consists of weekly indicators of the share of public school students engaged in one of the three learning modes, although their definitions are somewhat different.[7] In particular, "Hybrid" means that either students in some grades can return to buildings in person while other grades can only return in a hybrid or remote mode, or all students can return to in-person school for four days or less. Return2Learn provides weekly data 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 information sources used to construct the three learning modes, it is important to compare the Burbio and Return2Learn trackers. As shown in Figure 2, while the learning mode percentages averaged across counties from the two trackers generally show a similar evolution over time, in particular with regards to the decline in virtual learning in 2021, there are also important level differences, especially with regards to 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. These differences are even more pronounced for some regions of the U.S. and make it difficult to use the trackers without further information for a quantitative assessment of effective in-person learning.[8] Furthermore, note that since

[6]The Burbio data also contains another category called "Undecided". Since it is only rarely used, we ignore it and rescale the three main learning modes such that they always sum to 100 for each county and week. See https://about.burbio.com/methodology/ for further details.

[7]Return2Learn also uses a different terminology, classifying the different learning modes as "Fully in-person", "Hybrid", and "Fully remote". See https://www.returntolearntracker.net/about/ for details.

[8]We are not the first to point out the difficulty of distinguishing between "Traditional" and "Hybrid" learning, and that

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 effective in-person learning between school types (e.g. public vs. private or elementary vs. secondary) or demographic and socioeconomic surroundings of individual schools.

3 From school visits to effective in-person learning

To construct a measure of effective in-person learning (EIPL), we map changes in school visits from Safegraph to changes in school learning mode from Burbio and Return2Learn. This is the main methodological contribution of the paper and proceeds in two steps. First, we aggregate changes in weekly visits to the pertinent level (the county level for Burbio, respectively the district level for Return2Learn) to estimate the mapping. Second, we use the estimates to compute measures of EIPL at the individual school level.

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 week t that students in county c effectively spend in in-person learning as

$$EIPL_{c,t} = T_{c,t} + \gamma H_{c,t}, \tag{1}$$

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 estimate their relationship with a linear regression

$$EIPL_{c,t} = \alpha + \beta\Delta\widetilde{v}_{c,t} + \varepsilon_{c,t},$$

or equivalently

$$T_{c,t} = \alpha + \beta\Delta\widetilde{v}_{c,t} + \gamma H_{c,t} + \varepsilon_{c,t}. \tag{2}$$

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.

We estimate (2) using the weeks from September 6, 2020 to December 13, 2020 excluding the week of Thanksgiving. The reason we are not using data for Winter and Spring 2021 is that during this period, schools increasingly moved away from virtual learning. But when $V_{c,t} \approx 0$, traditional learning becomes 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. During Fall 2020, in contrast, there are changes across all three learning modes, which enables us to identify the mapping between $T_{c,t}$ and $\Delta\widetilde{v}_{c,t}$, controlling for $H_{c,t}$.

Table 1 reports the results of the estimation, both for the Burbio county level mapping (first panel) and the Return2Learn district level mapping (second panel). To save on space, the table only shows estimates for the case when we restrict $\alpha = 100$ as implied by our measures for the pre-pandemic period

this leads to inconsistencies between different trackers of school learning mode. See Camp and Zamarro [2021].

Table 1: Regression of "Traditional learning" on changes in school visits

Dependent variable	Traditional learning ($T_{c,t}$)			
	(a) Burbio		**(b)** Return2Learn	
	(1)	**(2)**	**(3)**	**(4)**
Change in school visits ($\Delta\widetilde{v}_{c,t}$)	1.09***	1.09***	1.00***	1.03***
	(0.02)	(0.02)	(0.02)	(0.02)
Hybrid learning ($H_{c,t}$)	-0.42***	-0.40***	-0.49***	-0.46***
	(0.02)	(0.02)	(0.00)	(0.00)
R-squared	0.63	0.65	0.60	0.61
# of geographic units	2,783	1,741	7,533	2,413
# of students (in thousands)	44,668	42,726	41,478	32,088
% of all public-school students	87.9	84.1	81.6	63.1

Notes: Safegraph, Burbio and Return2Learn data for the Fall term 2020 (weeks of September 6, 2020 to December 19, 2020, excluding the week of Thanksgiving). All regressions are weighted by student enrollment at the county (Burbio) or district (Return to learn) level. Standard errors are clustered at the county (Burbio) or district (Return to learn) level. In columns (1) and (3), all counties and districts with available school visits data are included (subject to the sampling restrictions for schools with sparse or noisy visits data described in Section (2)). In columns (2) and (4), all counties and districts with data for at least 5 schools are included.

when schools are fully in person (i.e. $T_{c,t} = 100$, $\Delta\widetilde{v}_{c,t} = 0$, and $H_{c,t} = 0$).[9] In columns (1) and (3), we include all geographic units (counties or districts) with available data (subject to sample restrictions on Safegraph data described above). For both datasets, the mapping between learning modes and school visit changes is tightly estimated, with a R^2 of 0.6 or higher and highly significant coefficients. A 1 percentage point decline in school visits reduces EIPL by 1.09 percentage points according to Burbio and 1.00 percentage points according to Return2Learn, while the estimated fraction of hybrid learning spent in person is 0.42 according to Burbio and 0.49 according to Return2Learn, or 2 to 2.5 days out of a 5 day school week.

In columns (2) and (4), we restrict the sample to the counties and districts for which we have reliable school visits data for at least 5 schools.[10] The results are robust across the different specifications, confirming that there is a tight linear relationship between change in school visits and Burbio and Return2Learn learning modes. Also note that the estimated γ is consistently lower with the Burbio data than with the Return2Learn data. This difference reflects the fact, discussed above, that Burbio is on average more likely to classify a student to be in traditional in-person learning mode, which in turn implies that hybrid learning spent in person is on average lower than according to Return2Learn.

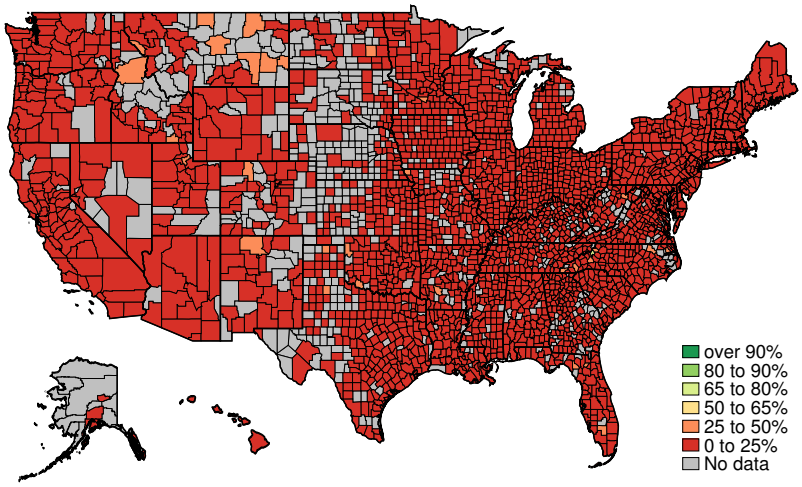
To refine the mapping, we reestimate (2) separately for different regions. By doing so, we allow for differences in how a given learning mode translates into EIPL. As shown in Appendix C, while the fit of the regression is high across all regions, there are substantial differences in the estimated coefficients, especially with regards to the share of hybrid learning spent in person. This should not be surprising since different districts pursued different hybrid school policies and these differences appear to vary systematically across regions of the U.S.

[9] When we run the regression unrestricted, α is estimated close to 100.

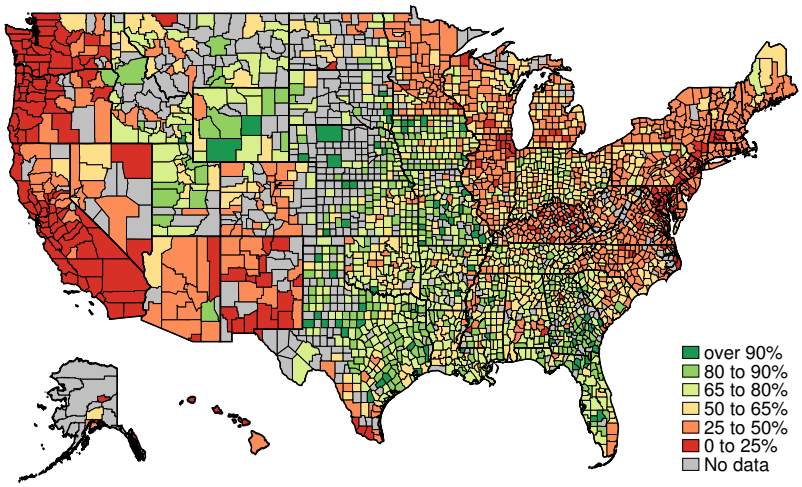
[10] Since many counties have more than one public school district and also contain independent private and public charter schools, this restriction reduces the Return2Learn sample by a larger proportion.

Figure 3: County-level loss of effective in-person learning

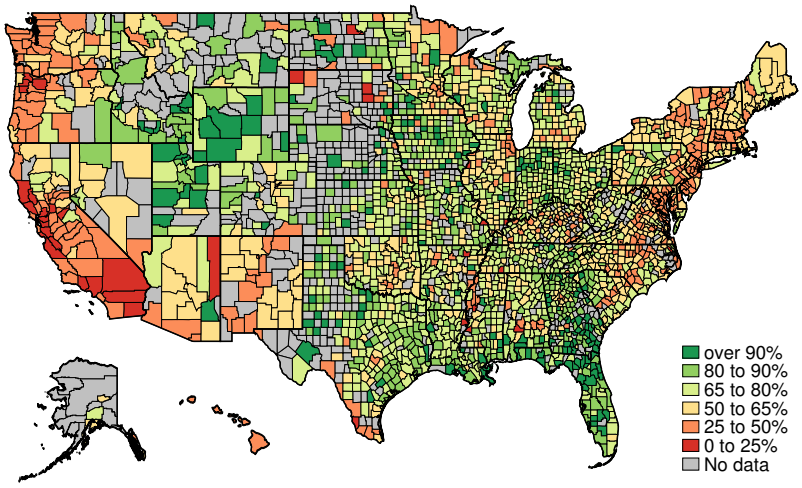
(a) March 2020 to May 2020



(b) September 2020 to December 2020



(c) January 2021 to May 2021



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

Given these estimates, we compute $EIPL_{j,t}$ at the individual school level by projecting visit changes $\Delta\widetilde{v}_{j,t}$ based on the mapping from each region that yields the highest R^2. This effectively assumes that the relationship between learning modes and school visits for all schools in a given region is the same, including for private schools.

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

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

Figure 3 shows the map of the resulting county EIPLs for three different time periods. From March to May 2020, EIPL was greatly reduced across most of the U.S., averaging between 0% and 20% of its pre-pandemic level, as many schools closed completely for in-person learning. From September to December 2020, EIPL recovered partially but very unequally across regions. While EIPL increased to between 40% and 60% on average in the Southern, Midwestern, and Central Northern parts of the country as many schools returned to at least partial in-person instruction, EIPL in the Northeastern and Western parts remained stuck in the 0% to 40% range as a large fraction of schools continued in full virtual mode. From January through May 2021, EIPL generally recovered further but large regional disparities persist, with counties located mostly on the East and West Coast continuing to average EIPL below the 50% mark.[11]

Table 2 provides another illustration of the large 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 cities in the U.S. by population. In cities in Florida and Texas such as Jacksonville or Houston, EIPL averaged 75% or higher, whereas in cities in California, Oregon and Washington such as Los Angeles, Portland or Seattle, EIPL averaged less than 25%. These differences are striking and suggest that state-wide policies and attitudes towards school closing may have played an important role.

Table 2: The top 10 and bottom 10 cities of effective in-person learning

Rank	CBSA name	EIPL	Rank	CBSA name	EIPL
1	Jacksonville, FL	84.5%	41	San Francisco-San Mateo-Redwood City, CA	24.9%
2	Fort Worth-Arlington, TX	80.0%	42	Sacramento--Arden-Arcade--Roseville, CA	23.7%
3	Houston-Baytown-Sugar Land, TX	79.9%	43	Las Vegas-Paradise, NV	22.5%
4	Dallas-Plano-Irving, TX	79.5%	44	San Diego-Carlsbad-San Marcos, CA	22.4%
5	Tampa-St.Petersburg-Clearwater, FL	76.7%	45	Portland-Vancouver-Beaverton, OR-WA	21.2%
6	San Antonio, TX	74.2%	46	Seattle-Bellevue-Everett, WA	21.0%
7	Orlando, FL	71.6%	47	San Jose-Sunnyvale-Santa Clara, CA	18.7%
8	Austin-Round Rock, TX	71.0%	48	Los Angeles-Long Beach-Santa Ana, CA	18.0%
9	Cincinnati-Middletown, OH-KY-IN	65.4%	49	Oakland-Fremont-Hayward, CA	15.6%
10	Atlanta-Sandy Springs-Marietta, GA	59.7%	50	Riverside-San Bernardino-Ontario, CA	13.8%

Notes: The table shows the top-10 and bottom-10 Core-Based Statistical Areas (CBSAs) in terms of average EIPL from September 2020 to May 2021 among the 50 largest CBSAs by population. EIPL for each CBSA is computed as the student-weighted average across schools with reliable data.

[11] Appendix C provides additional evidence on the temporal disparities across regions by showing weekly time series of average EIPL for the nine different U.S. Census divisions.

At the same time, it is important to emphasize that there are large disparities in EIPL across schools even within counties. As shown in Table 3, 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.

Table 3: 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	57.4%	14.3%
0% – 25%	21.8%	10.4%
25% – 50%	38.7%	16.3%
50% – 75%	61.5%	15.2%
75% – 100%	83.5%	10.0%

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.3 for a description.

5 What accounts for the disparities in EIPL?

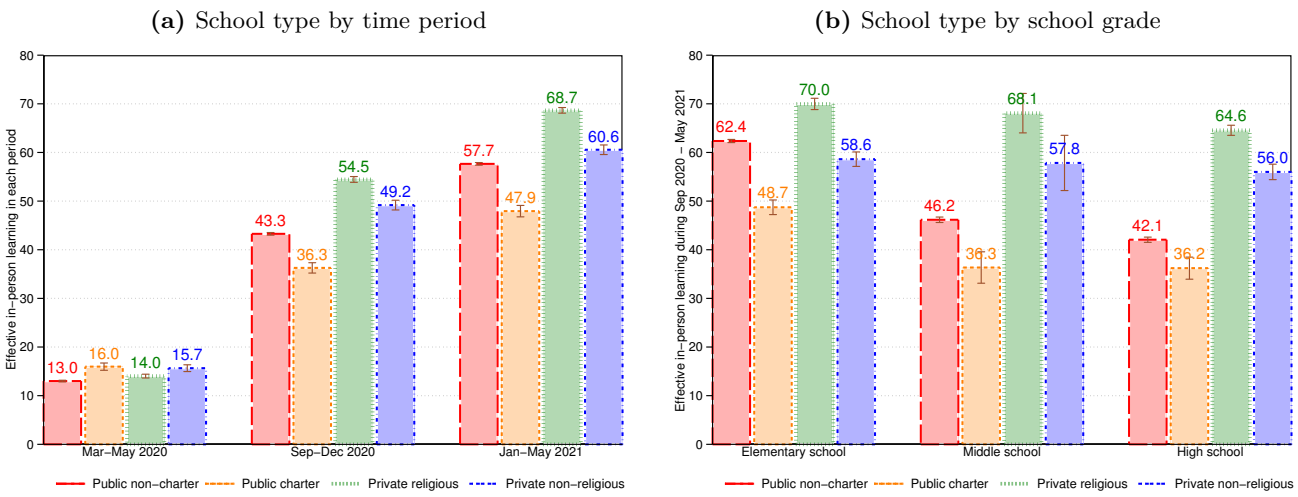
Why did some schools return to in-person learning more quickly than others and what explains the large geographic differences? To answer these questions, we begin by analyzing how various observable school characteristics and local conditions surrounding the school correlate with EIPL. Then we return to geography and examine the extent to which the correlation between EIPL and school characteristics, respectively local conditions are due to systematic regional differences, and what may account for these 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 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.

5.1 School type and grade

Perhaps the most obvious observable school characteristics are school type and grade. NCES designates each school as either a public non-charter, public charter, private non-religious, or private religious school, and either elementary school, middle school, high school, or a combination thereof.[12] We use this information to group individual schools accordingly and compute average student-weighted EIPL for each group.

[12]Elementary school generally comprises grades K through 5, middle school comprises grades 6 through 8, and high school comprises grades 9 through 12.

Figure 4: 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.

Panel (a) of Figure 4 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.

This ranking, which is not driven by other observable factors, may come as a surprise for two reasons. First, public charter schools are typically independent whereas public non-charter schools belong to school districts that, for some urban areas, are comprised of several hundred schools. One could have expected that being independent 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.

Panel (b) of Figure 4 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.[13] 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.

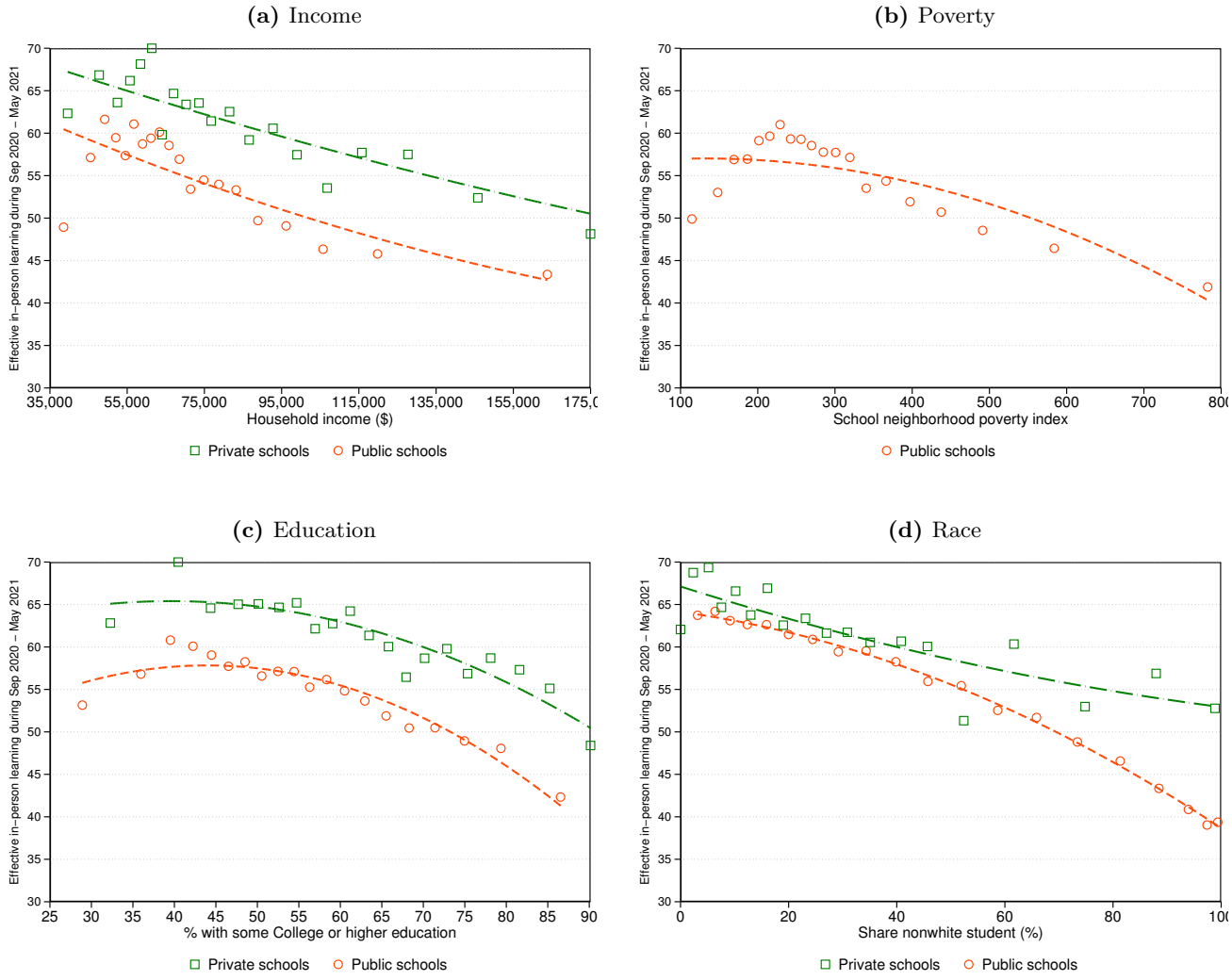
5.2 Affluence, education, and race

The next dimension we consider is local affluence, education, and race. We measure local affluence either as average household income in the zip-code of the school or school neighborhood poverty. Zip-code

[13]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.

average household income is based on a 2016-2019 average from the American Community Survey (ACS), while school neighborhood poverty is an estimate constructed by NCES for public schools, also based on data from the ACS. For education, we use the share of households with some college education or higher in the zip-code of the school, again taken from the ACS. For race, finally, we use the school's share of non-white students, provided in the NCES data for the 2018-19 school year.

Figure 5: Effective in-person learning by local affluence, education, 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) school neighborhood poverty, (c) zip-code average share of college educated, and (d) school share of non-white students. The school neighborhood poverty index is available only for public schools and ranges from 0 (poorest) to 1000 (richest). Observations are weighted with the school-specific sampling weights discussed above.

We begin with nonparametric binned scatterplots of the unconditional relationship between average school EIPL from September 2020 to May 2021 and the different measures of affluence, education, and race. As shown in panel (a) of Figure 5, there is a striking inverse relationship between EIPL and household income: schools in zip-codes with *high* average household income provided on average *lower* levels of EIPL during the 2020-21 school year. Consistent with the above results, EIPL is on average higher for private

schools than for public schools but interestingly, the relationship of EIPL with household income and education is otherwise very similar. Panel (b) confirms that an equivalent inverse relationship between EIPL and affluence holds with regards to neighborhood poverty: public schools in *poorer* neighborhoods (i.e. school with a lower index) provided on average *higher* EIPL during the pandemic.[14]

Panel (c) makes a similar point for education: schools in zip-codes with a *high* share of college-educated people provided on average *lower* levels of EIPL. Given that affluence and education are highly correlated, this may be expected but it is still remarkable.

Finally, as shown in panel (d), there is also a striking inverse relationship between EIPL and the share of non-white students of a school. For schools with close to 0% of non-white students, EIPL averaged about 65% during the 2020-21 school year, 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 40% for public schools and just below 55% for private schools. Interestingly, the gap between public and private schools with respect to race is small for schools with up to about 50% of non-white students and then opens up gradually as the share of non-white students increases towards 100%.

Given the general association of poverty with race, the inverse relationship of EIPL with both affluence and race may come as a surprise. But according to our data, the share of a school's non-white students is essentially uncorrelated with local affluence and education (see Appendix C). In other words, the negative relationship of EIPL with affluence and education is, at least on average, independent of the negative relationship of EIPL with the share of non-white students.

To illustrate this result and assess the relative importance of affluence and education versus race, 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 that the sample is representative of the full set of schools; standard errors are clustered at the county level; and the coefficients are scaled so that they show directly the implied change in EIPL of going from the 25th percentile to the 75th percentile of the distribution of a variable.[15]

Table 4 reports the 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.[16] Adding these controls does not change the results noticeably, and the estimates on these controls are in line with the results shown in Figure 4.

[14] The poverty index ranges from 0 (poorest) to 1000 (richest). This index is not available for public schools.

[15] More precisely, all right-hand side variables are expressed as deviations from the mean, normalized by the interquartile range. Appendix C provides descriptive statistics for the different variables.

[16] Since local affluence and education are highly correlated with each other, regressing EIPL on all of them jointly would not noticeably increase the explanatory power but would make it difficult to interpret the different estimates.

Table 4: 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)**
Zip-level average household income	-6.07***			-4.51***	
	(0.61)			(0.62)	
Zip-level share of college educated		-7.52***			-5.17***
		(0.71)			(0.80)
School neighborhood poverty			-6.03***		
			(0.47)		
School share of non-white students	-13.97***	-14.02***	-17.13***	-5.75***	-6.03***
	(2.11)	(2.19)	(1.85)	(1.08)	(1.11)
School type and grade controls	✔	✔	✔	✔	✔
R-squared	0.07	0.06	0.07	0.04	0.04
Observations	55,036	55,037	53,996	11,280	11,280

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.3. The regressions are estimated on average school EIPL for the period from September 2020 to May 2021. Columns (1) to (3) show estimates for the public school sample, and columns (4) and (5) show 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.

Panel (a) shows the estimates for the public school sample. The estimates all tell the same story as the unconditional scatterplots above. EIPL is inversely related to local affluence, local education, and the school's share of non-white students in statistically significant and quantitatively important ways. According to Column (3), for instance, a school located in a zip-code at the 75th percentile of the income distribution and with a student body at the 75th percentile of the non-white distribution is predicted to have had 20% lower EIPL during the 2020-21 school year than a school at the 25th percentile of the two distributions. Almost three quarters of this difference comes from racial composition of the school, which is remarkable.

Panel (b) shows the estimates for the private school sample (estimates for school neighborhood poverty index are missing because this index is not available for private schools). As for public schools, EIPL of private schools is inversely related to affluence and race. Consistent with the binned scatterplot shown above, however, the estimated association of EIPL with the share of non-white students is only about one third as important for private schools as for public schools.

The regressions confirm the nonparametric result of Figure 5 that there are large differences in EIPL by local affluence, local education, and racial composition of a school. At the same time, the R-squared remains below 0.1 for all of the regressions, which means that differences in local affluence and race account for only a relatively small share of the variation in EIPL across schools. We therefore proceed by adding other observable school characteristics to explore how they affect the results.

5.3 Public school/district size and school funding

The additional school characteristics we consider are school/district size and school funding. We have also considered other observables such as pre-pandemic school test scores and several alternative demographic and socio-economic indicators describing the neighborhood of the school. All of these variables are highly correlated with the above measures of local affluence and therefore do not add significant explanatory power.

For school size, we use school enrollment, and for district size, we use the number of schools in the district. Both variables are available for the 2018-19 school year from the NCES data and are close to orthogonal with local affluence. In turn, district size is positively correlated with share of non-white students, reflecting the fact that school districts are typically larger in urban areas that are racially more diverse.

For school funding, we consider both pre-pandemic school spending per student and district-level ESSER funding by student. Pre-pandemic school spending per student is available for the 2018-19 school year from the National Education Resource Database on Schools of the joint Edunomics lab / Massive Data Institute at Georgetown University. Interestingly, this variable is only weakly correlated with the different measures of local affluence above. ESSER funding by public school districts is available from Malkus [2021b] and the Return2Learn team. Quite surprisingly, district-level ESSER funding per student is essentially unrelated to pre-pandemic school spending per student – an observation to which we return below – but it is negatively correlated with local affluence and education and positively correlated with the share of non-white students.

Except for school enrollment, the different variables are not available for private schools. We therefore focus here on public schools and report a limited set of results for private schools in Appendix C. As above, we estimate OLS regressions of EIPL on affluence and race and then add the different school funding and school/district size variables. This allows us to assess their association with EIPL conditional on affluence and race. See the appendix for nonparametric binned scatterplots of the unconditional relation with EIPL.

Table 5 shows the results. Since local affluence and education are highly correlated with each other, we report regression estimates where we include each of these measures one-by-one. The estimates for the other left-hand side variables (race, school/district size, and school funding) barely change across the regressions; so for these variables, we report estimates controlling for local affluence and education jointly. Column (1) repeats the results from Table 4 above as a reference.[17] Column (2) adds school and district size to the regression; column (3) adds school spending and ESSER funding; and column (4) adds the four variables jointly.

[17] The estimates for household income, share of college educated and school neighborhood poverty are exactly as in columns (1) - (3) of Table 4. The estimate for share of non-white students is slightly different because this estimate is obtained while controlling jointly for all three measures.

Table 5: The (un)importance of public school/district size and school funding

Dependent variable	Effective in-person learning (EIPL)			
	(1)	**(2)**	**(3)**	**(4)**
Zip-level average household income[a]	-6.07***	-5.31***	-5.46***	-4.15***
	(0.61)	(0.56)	(0.60)	(0.58)
Zip-level share of college educated[a]	-7.52***	-6.87***	-8.37***	-6.74***
	(0.71)	(0.81)	(0.61)	(0.70)
School neighborhood poverty[a]	-6.03***	-5.40***	-6.40***	-5.00***
	(0.47)	(0.47)	(0.50)	(0.49)
School share of non-white students[b]	-17.10***	-15.09***	-12.93***	-10.02***
	(1.83)	(1.49)	(1.81)	(1.64)
Student enrollment[b]		-1.63***		-3.37***
		(0.40)		(0.35)
District number of schools[b]		-0.60*		-0.42
		(0.34)		(0.30)
School spending per student[b]			-4.82***	-5.92***
			(0.68)	(0.68)
ESSER funding per student[b]			-3.91***	-3.75***
			(0.60)	(0.59)
School type and grade controls	✔	✔	✔	✔
R-squared	0.08	0.08	0.09	0.10
Observations	53,978	53,978	51,043	51,043

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.3. 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.

EIPL is estimated to be lower for larger schools and for schools belonging to larger districts although this latter estimate is not significant.[18] Interestingly, EIPL is also negatively associated with school spending and ESSER funding per student. The former result may have been expected given the positive correlation of school spending per student with local affluence and the inverse relation of EIPL with local affluence found above. Nevertheless, it raises the question as to why public schools with higher per-student spending prior to the pandemic were less likely to reopen for in-person learning than lesser funded schools. The negative 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 both of these points below.

The final important observation about these regressions is that adding school/district size and school

[18]The negative yet insignificant estimate for district size is not driven by public charter schools, which are typically independent or part of a small network, since the regressions control for school type. To the contrary, as we have seen above, public charter schools provided on average lower and not higher EIPL.

funding reduces the negative relation of EIPL with local affluence, education, and race only modestly. Even after controlling for these additional school characteristics, a school located in a zip-code at the 75th percentile of the income distribution and with a student body at the 75th percentile of the non-white distribution is predicted to have had about 14% lower EIPL during the 2020-21 school year. This suggests that affluence and race are not simply proxies for school size or school spending per student.

5.4 Geography

In the last part of the analysis, we return to geography and ask how much of the relation of EIPL with school characteristics and local conditions is driven by systematic regional differences, and what factors not directly related to schools can account for these differences.

Table 6: The importance of geography

Dependent variable	Effective in-person learning (EIPL)			
	(1)	**(2)**	**(3)**	**(4)**
Zip-level average household income[a]	-4.15***	-3.51***	-1.25***	-0.12
	(0.58)	(0.57)	(0.28)	(0.22)
Zip-level share of college educated[a]	-6.74***	-5.85***	-3.53***	-1.47***
	(0.70)	(0.71)	(0.35)	(0.28)
School neighborhood poverty[a]	-5.00***	-4.35***	-1.82***	-0.46**
	(0.49)	(0.50)	(0.27)	(0.20)
School share of non-white students[b]	-10.02***	-8.68***	-10.80***	-6.85***
	(1.64)	(1.55)	(0.65)	(0.62)
Student enrollment[b]	-3.37***	-3.09***	-3.05***	-3.24***
	(0.35)	(0.33)	(0.21)	(0.20)
District number of schools[b]	-0.42	-0.37	-0.28	0.08
	(0.30)	(0.31)	(0.23)	(0.12)
School spending per student[b]	-5.92***	-5.73***	-0.16	0.49**
	(0.68)	(0.69)	(0.27)	(0.21)
ESSER funding per student[b]	-3.75***	-3.92***	-2.83***	-3.18***
	(0.59)	(0.58)	(0.32)	(0.39)
School type and grade controls	✔	✔	✔	✔
Locale FE		✔	✔	✔
State FE			✔	
County FE				✔
R-squared	0.10	0.10	0.26	0.31
Observations	51,043	51,043	51,043	51,043

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.3. 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.

Table 6 reports the results of reestimating the above regressions with various geographic controls. As before we perform these estimations for the sample of public schools and relegate results for private schools to the appendix. Column (1) repeats the final regression in Table 5 above for reference. In column (2), we add controls for whether a school is located in a city, a suburb, a town, or a rural area, as designated in the NCES data. The estimates for affluence and race become somewhat smaller, 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. But overall, these reductions in the coefficients and the additional explanatory power of these geographic locale controls remain modest, indicating that the results above are not driven by systematic differences between cities, suburbs, and town/rural areas.

Columns (3) and (4) 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 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]

Fourth, the state fixed effect completely absorbs the negative association of EIPL with pre-pandemic school spending per student, while the county fixed effect turns the association between the two variables mildly positive. This indicates that school spending per student did not play a decisive role for EIPL overall but instead picked up systematic differences across states. In contrast, the inverse relation between EIPL and ESSER funding per student remains unaffected by geography. In other words, even within counties, schools in districts with more ESSER funding per student provided on average less EIPL.

The remaining question is what factors explain why schools in some states and counties of the U.S. provided substantially higher EIPL than others. From Figure 3 we know that these places are located primarily in the central and southern parts of the U.S. – places that in general were more favorable towards reopening the economy despite potential health risks and at the same time have seen lower

[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.

COVID vaccination rates. To assess the relative importance of these factors for EIPL, we re-estimate the above regression as a school-week panel and extract the county-week fixed effect.[20] Then we regress the resulting county-week fixed effects on party affiliation of the state governor and the Republican vote share in the 2020 presidential election as proxies for the general stance towards reopening schools as well as weekly vaccination rates.[21] In addition, we control for the county's COVID health situation with pre-pandemic ICU bed capacity, two-weeks lagged COVID case and death rates, and maximum weekly county temperature. As before, right-hand side variables except the indicator for state governor are scaled so that the estimated coefficients directly reflect the predicted change in the county fixed effect of moving from the 25th percentile to the 75th percentile of a variable's distribution. All regressions are weighted by pre-pandemic county population.

Table 7 reports the results.

Table 7: Accounting for systematic geographical differences

Dependent variable	County-week fixed effect			
	(1)	**(2)**	**(3)**	**(4)**
Republican governor	21.55***		22.89***	21.52***
	(1.64)		(1.76)	(1.82)
Share of 2020 Republican voters	11.31***		11.96***	12.39***
	(0.91)		(1.03)	(1.06)
Lagged COVID vaccination rate		7.18***	12.32***	10.65***
		(1.44)	(1.15)	(1.50)
ICU bed capacity				0.91*
				(0.54)
Lagged COVID case rate				-0.74*
				(0.39)
Lagged COVID death rate				0.37
				(0.25)
Maximum weekly temperature				3.24**
				(1.26)
NPI controls				
R-squared	0.32	0.02	0.37	0.38
# of counties	2,829	2,819	2,819	2,814
# of weeks	35	35	35	35

Notes: Each column reports coefficients from a weighted OLS regression of the county-week fixed effects, with standard errors clustered at the county level in parentheses and observations weighted by county population. The county-fixed fixed effect is extracted from the regression of EIPL against school and local characteristics (see Column (5) of Table 6) for the period from September 2020 to May 2021. The NPI controls consist of county-week indicators for stay-at-home orders, restrictions on public gatherings and mask mandates.

We begin in Column (1) with a regression of the county-week fixed effect on party affiliation of the state governor and the Republican vote share in the 2020 presidential election. The estimated coefficients

[20] Running the regression as a county-week panel does not affect any of the reported estimates since all of the left-hand side variables are time-invariant and predetermined.

[21] County-level data on the presidential elections is obtained from the MIT election lab. Data for COVID cases, deaths, and vaccination rates come from the New York Times, the Johns Hopkins Coronavirus Resource Center, and the Centers for Disease Control and Prevention, and county-level counts of ICU beds are obtained from the Kaiser Health News; see Appendix A.1 for details about the data sources.

are large and together account for one third of the variation in the county fixed effects. Everything else constant, schools located in a county with a Republican governor provided more than 20% higher EIPL during the 2020-21 school year. Likewise, schools located in a county at the 75 percentile of the distribution of share of Republican voters in the 2020 presidential election provided more than 10% higher EIPL than schools located in a county at the 25th percentile.

Column (2) considers the effect of vaccination rates for the 2021 part of the sample. All else equal, schools in counties at the 75th percentile of the distribution of COVID vaccination rates averaged 7% higher EIPL than schools in counties at the 25th percentile. However, the explanatory power of this regressor is weak, indicating that vaccination rates leave most of the variation in county-week EIPL unaccounted.

Columns (3) combines the two different factors. Since Republican party voting preferences and vaccination rates are negatively correlated, this pushes up the respective estimates. Column (4), finally, adds the COVID health controls. Except for temperature, the effect of these health controls are quantitatively unimportant and surrounded by considerable uncertainty, suggesting that the local COVID health situation on average did not play an important role for school reopenings and EIPL.

6 Conclusion

Using a new empirical approach that combines comprehensive data on school visits with information on school learning modes, 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 school characteristics and local conditions.

While the results from our regression analysis should not be interpreted as causal, they nevertheless reveal interesting patterns that raise important questions on why EIPL in some areas and for some students was so much lower during the 2020-21 school year. We highlight three questions in particular that merit further investigation:

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 systematic geographic differences, which in turn are correlated with political preferences. But why would more democratic-leaning areas have been more reluctant to return students 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. This explanation contrasts, however, 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 more affluent and educated parents) provided more EIPL than public schools. So, clearly more work is needed on this front. But no matter the explanation, it remains that students in less affluent and less educated areas of the U.S. received on average more EIPL, which may

mitigate some of the negative consequences of school closures for long-term earnings inequality (see Fuchs-Schündeln et al., 2021 for a simulation of these effects).

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 white and non-white students that existed already before the pandemic.

3. Why did schools in districts with more ESSER funding per student provide less EIPL? One possible reaction to this remarkable result is that without ESSER funding, schools would have been closed for even longer. Yet, the negative relationship arises even within counties and despite controlling for neighborhood poverty and race, 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.[22]

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.

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.

Andrew Camp and Gema Zamarro. Determinants of ethnic differences in school modality choices during the COVID-19 crisis. *EdWorkingPaper No. 21-374*, 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.

Curriculum Associates. Academic achievement at the end of the 2020-21 school year. Technical report, Curriculum Associate Research Brief, June 2021.

[22]The remaining unspent ESSER funds could still be put to good use, for instance by improving school infrastructure or providing summer learning programs for students who fell behind, 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. As pointed out by Malkus [2021a], however, ESSER funding comes with very few constraints and it is unclear to what extent school districts will use the funds for their intended purpose.

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.

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.

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.

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

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, http://www.returntolearntracker.net/esser/, 2021b.

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.

Appendices

A Data appendix

This appendix presents additional information about our data sources (Subsection A.1); the Safegraph foot traffic data and our algorithm to match NCES datasets to the Safegraph data (Subsection A.2); the procedure to construct school weights (Subsection A.3); a comparison between the matched data and NCES datasets (Subsection A.4); and an examination of the school-level visits data (Subsection A.5).

A.1 Data sources

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).

EDGE data. The Education Demographic and Geographic Estimates (EDGE; see https://nces.ed.gov/programs/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 are type indicators (City, Suburban, Town, and Rural). We use EDGE data to complement the NCES datasets. First, we use the 2019-2020 geocodes to improve the reliability of the match between the CCD/PSS files and Safegraph data (see Subsection **??**). Second, we use the school neighborhood poverty estimates from EDGE. The estimates are constructed based on the data from the Census Bureau's American Community Survey which allow to compute income-to-poverty ratio (IPR). IPR is 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.

ACS data. The socio-demographic and income variables (household income at the county and ZIP-code levels, share of individuals with some College or higher educational attainment) 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. There is a one-to-one mapping from CBGs to county FIPS codes, allowing us to directly aggregate the estimates to the county level. For the ZIP-code level, on the other hand, we use the ZIP-TRACT crosswalk provided by the U.S. Housing and Urban Development (HUD)'s Office of Policy Development and Research (see https://www.huduser.gov/portal/datasets/usps_crosswalk.html). We aggregate to the ZIP-code level using the

so-called total ZIP ratio, which is the ratio of all addresses for each Census tract 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).

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.

A.2 Details on the Safegraph data and matching algorithm

Safegraph data. Each POI in Safegraph's data is identified by a unique persistent `safegraph_place_id`. A POI is essentially 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.[23] 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 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.

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:[24]

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.[25]

[23] 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.

[24] 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.

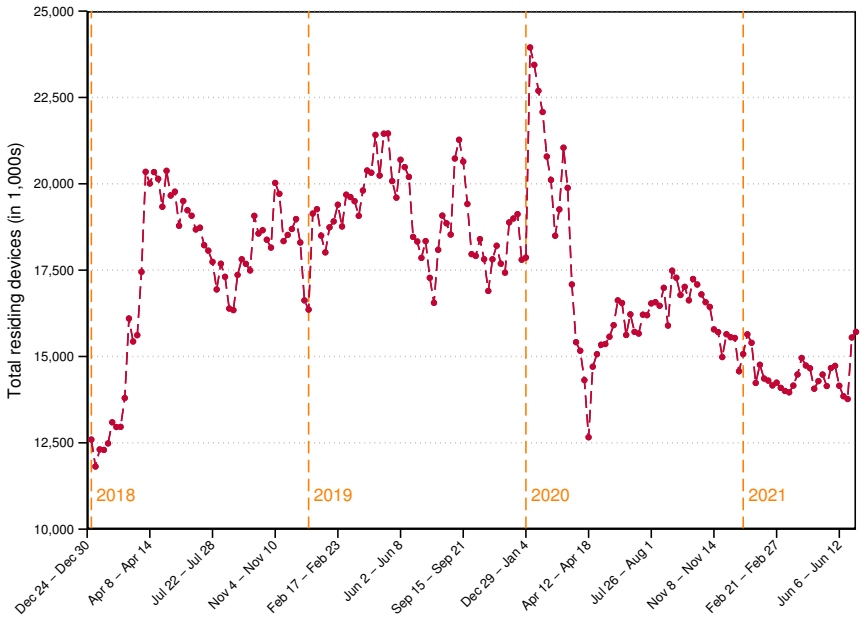
[25] Consider for instance Safegraph's "Big Spring Lake Kinderg Sch" described in Footnote 24. The name of this school in

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. 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).

Normalization and sample selection. When working with Safegraph's foot traffic data, there is an important concern that changes in visits counts over time can be driven by changes in the sample of cell phone devices that Safegraph uses to detect foot traffic. 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.[26] As can be seen, following large variations in the

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").

first two quarters of 2018, the sample 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

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.

[26] 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.

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 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.3 School weights

As explained in Section 2 of the main text and Subsection A.2 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 $\hat{\Pr}\{y_j = 1\}$.[27] 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).

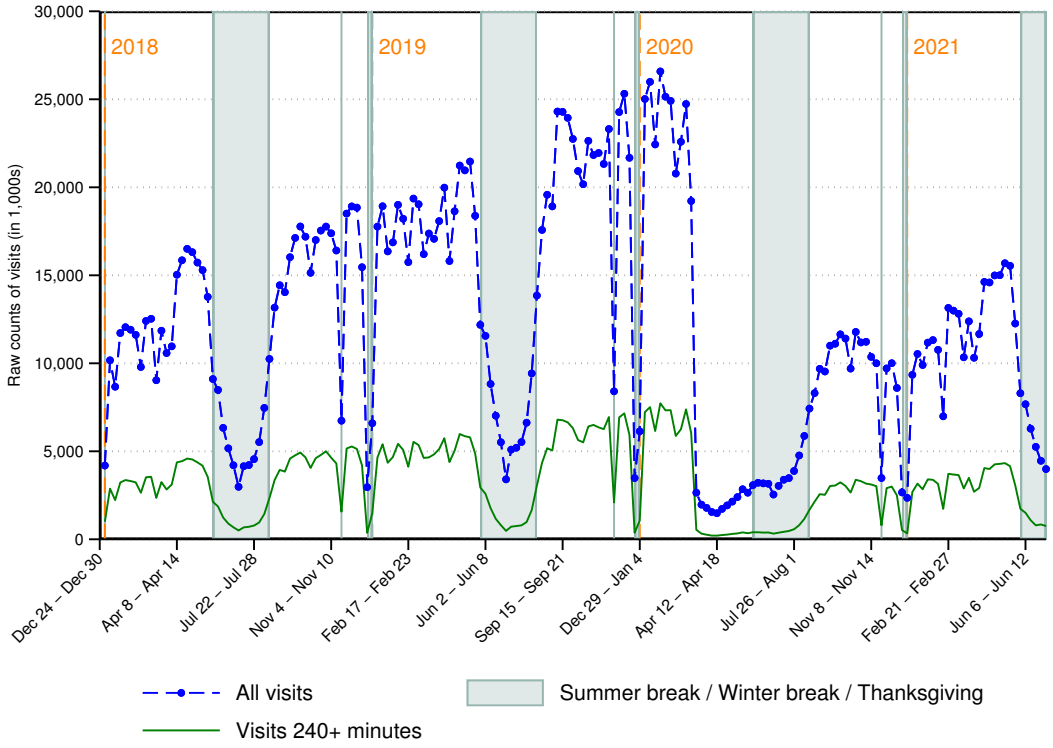
A.4 Comparison of selected sample to the NCES universe of schools

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

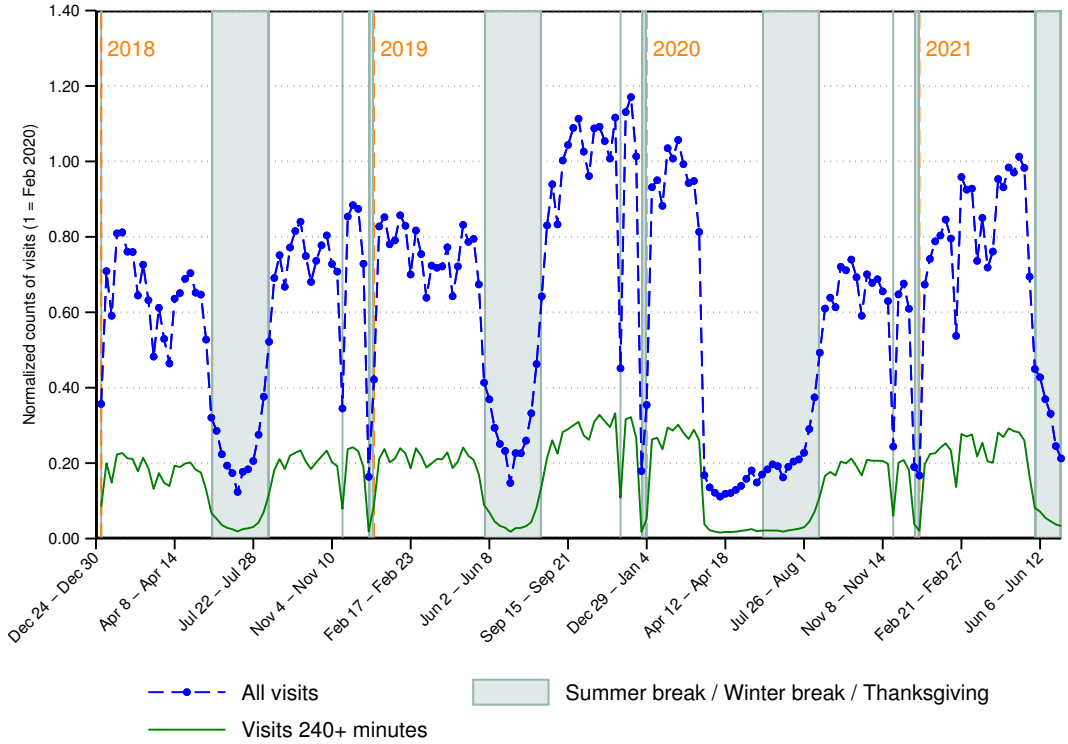
[27] Since the CCD contains the universe of public schools, the sampling weight of public schools is 1 and therefore the adjusted weight is 1 divided by the probability of selection into the "in scope" dataset. Across all schools, the final weights that we obtain range from 1.24 to 138.6 with an average 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 $\hat{\Pr}\{y_j = 1\}$.

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").

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[c]	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. ([a]) Includes special education, alternative, and other schools not classified by grade span. ([b]) NCES enrollment numbers in public schools include imputations for public school prekindergarten enrollment in California and Oregon.

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: ([a]) Matched refers to schools matched to Safegraph data. ([b]) In scope refers to schools matched to Safegraph data and with visits data that is neither too sparse or too noisy (see Subsection A.2). Except for the sample count, all the statistics for the in scope data are based on the weights calculated in Subsection A.3. ([c]) % Free lunch and % Reduced-price lunch denote the share of students who are eligible for free and reduced-price lunches, respectively.

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.

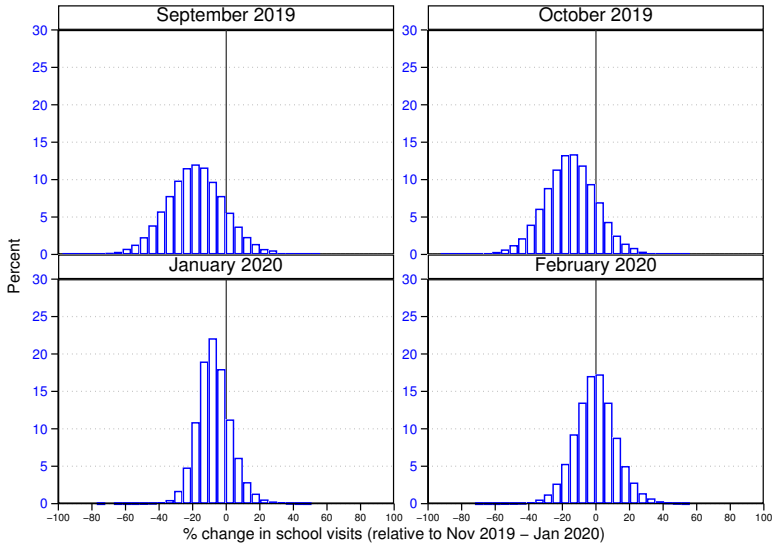
Through direct merges, matching and fuzzy-name matching (see Appendix A), we combine the CCD-PSS file to Safegraph data. Out of 124,583 schools in the CCD-PSS file, we obtain 102,774 matches to the Safegraph data for which the quality of the match is high enough to be considered as reliable. We call 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 $\Delta\widetilde{v}_{j,t}$. 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.3, which by construction make the in-scope dataset representative of the universe of CCD-PSS schools.

A.5 A closer look at the school visits data

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 (Sections 2 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 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 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.

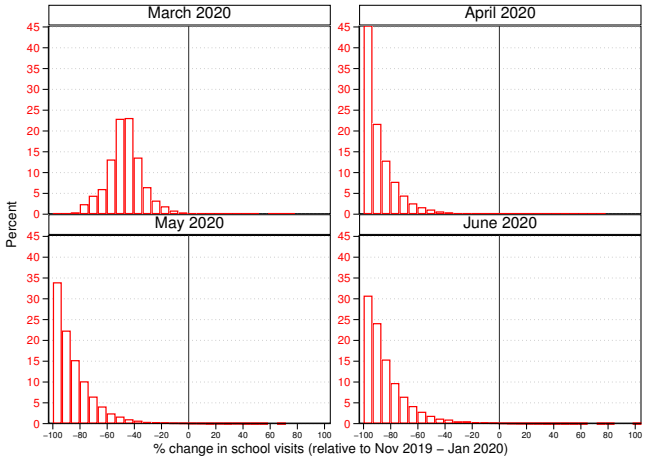
Figure A4 plots the distribution of average $\Delta\widetilde{v}_{j,t}$ during different months of the COVID-19 pandemic. The upper panel focuses on the first 4 months of the pandemic. The top left plot in this panel shows that $\Delta\widetilde{v}_{j,t}$ does well in capturing week-to-week variations: 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.[28]

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 vertical scale is the same in the two panels. We see a slight decrease in November and December relative to September-October 2020, which is partly due to the fact that both months include a week of vacation (Thanksgiving in November 2020, Christmas in December 2020). The lower panel shows a gradual 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.

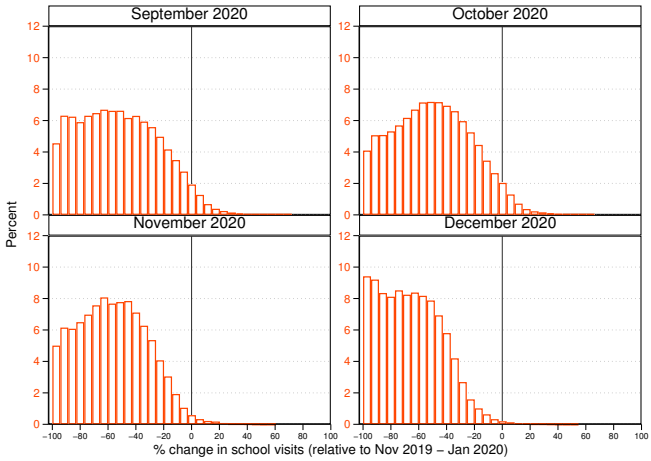
[28] In April, May and June in the upper panel of Figure A4, we observe some schools with changes in visits not lower than -60% or -80%. To understand how this relates to the upper map of county-level loss of EIPL (Figure 3), recall that: 1) these changes in visits are translated into EIPL by being multiplied by a coefficient that can be greater than 1, 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 3 since the data is averaged over longer periods of time.

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

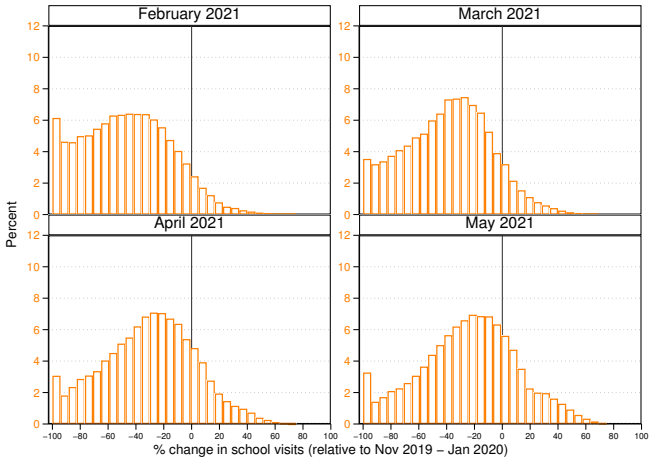
(a) March 2020 to May 2020



(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.

B Comparison with other Safegraph-based indicators

This section overviews two approaches to measuring school activity during the pandemic using visit data from Safegraph. The first one 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:

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

In this equation, $v_{j,m}$ denotes visits to school j during month m, ϑ denotes some threshold value, and $\mathbb{1}\{.\}$ is the indicator function. Thus it flags school j as being closed in 2020 if its year-over-year growth rate in visits drops by at least ϑ%. In their paper (and in the Internet repository providing their data) Parolin and Lee [2021] use a threshold value ϑ of 50%.[29] 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.[30] More importantly, this 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.[31,32] 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.

[29] 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 $\vartheta = 75\%$.

[30] 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).

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

[32] 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.

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

		(a) Burbio		**(b)** Return2Learn	
Tracker:		Virtual learning	Not virtual	Virtual learning	Not virtual
Eq. (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.

C Additional tables and figures

C.1 Mapping from School visits to EIPL at the region level

To create a tighter mapping from Safegraph school visits to EIPL, we re-estimate Equation (2) separately for each U.S. Census division. We proceed in two steps. We first search for the sample time window (starting from the week of September 6, 2020) that yields the better mapping as measured by the R^2 of the regression. Indeed, recall from Section 3 that in order to estimate the relationship between traditional learning ($T_{c,t}$) and school visits while controlling for the hybrid learning ($H_{c,t}$), we require the variation of $T_{c,t}$ to be not perfectly related to that of $H_{c,t}$. The extent to which this condition applies depends on how hybrid school policies have been implemented, and thus is likely to vary across regions. In the second step, we estimate (2).

Table C1 reports the results. First, in all regions with exception of the West South Central division (and East South Central division in Burbio), we obtain R^2 that are comparable to or even higher than the R^2 of the base regressions from Table 1. Second, we observe variations across regions in the coefficient on Hybrid learning, consistent with the idea the implementation of hybrid school policies displays some regional variation. Third, with exception of the West South Central division, the region-specific coefficients on changes in school visits are in line with those from Table 1 estimated nationwide.

C.2 Regional disparities in EIPL over time

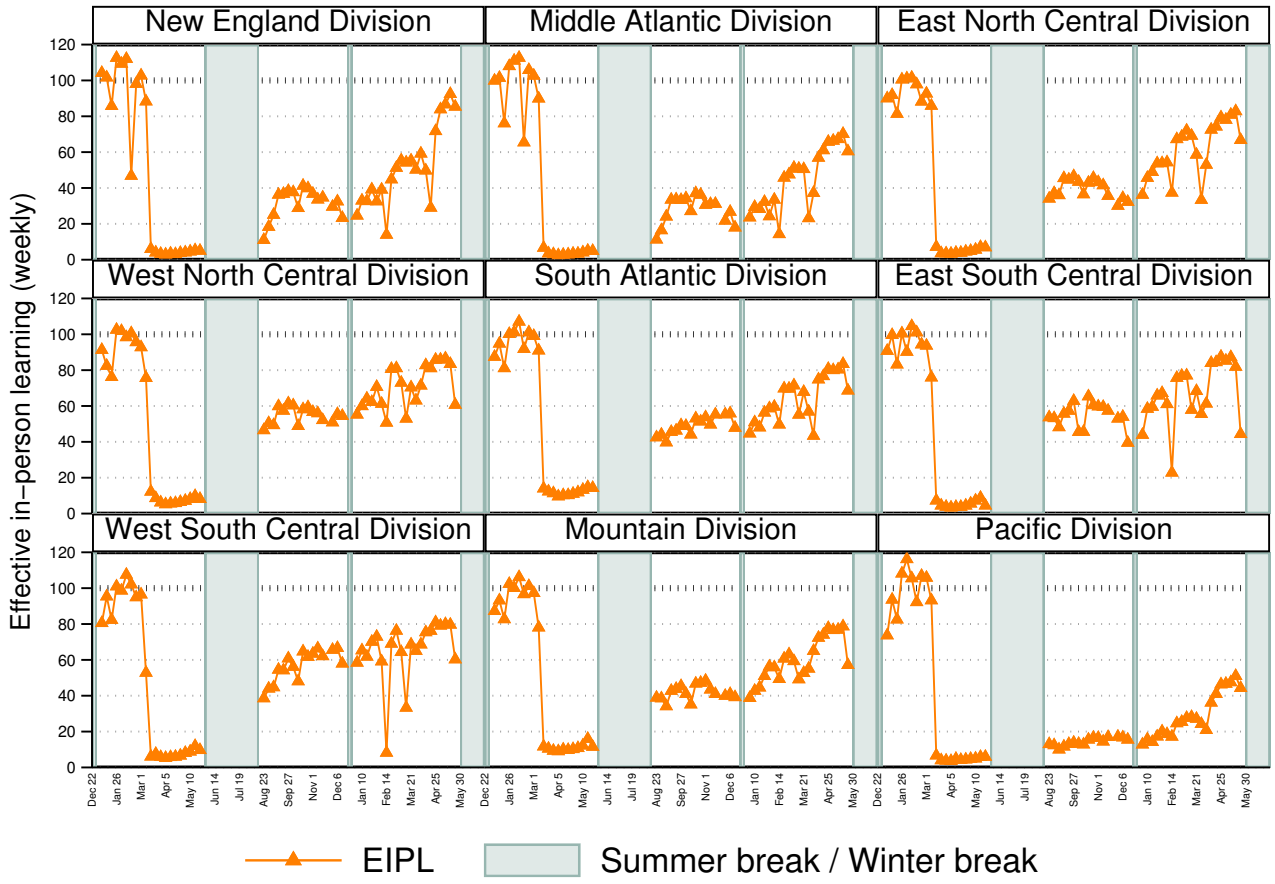
Figure C1 summarizes the temporal and geographic variation in EIPL by averaging weekly student-weighted EIPL for each of the nine U.S. Census Divisions. While EIPL drops to near zero for all divisions between March and May 2020, we see large differences during the 2020-21 school year. EIPL in states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

Table C1: Regression of "Traditional learning" on changes in school visits, by Census division

Dependent variable	Traditional learning ($T_{c,t}$)								
	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
					(a) Burbio				
Change in school visits ($\Delta\widetilde{v}_{c,t}$)	1.13***	1.25***	1.18***	1.20***	0.99***	1.12***	0.74***	0.98***	1.21***
	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)	(0.06)	(0.10)	(0.02)
Hybrid learning ($H_{c,t}$)	-0.36***	-0.29***	-0.49***	-0.50***	-0.45***	-0.44***	-0.60***	-0.34***	-0.33***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.05)	(0.06)	(0.06)	(0.02)
R-squared	0.72	0.77	0.58	0.61	0.59	0.44	0.43	0.63	0.90
Sample: From Sep 6, 2020 to	Jan 30, 2021	Dec 5, 2020	Apr 3, 2021	Jan 23, 2021	Feb 13, 2021	Feb 13, 2021	Nov 14, 2020	Jan 30, 2021	May 29, 2021
					(b) Return2learn				
Change in school visits ($\Delta\widetilde{v}_{c,t}$)	1.16***	1.27***	1.11***	1.00***	0.97***	1.07***	0.45***	1.07***	1.19***
	(0.05)	(0.02)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.04)	(0.01)
Hybrid learning ($H_{c,t}$)	-0.30***	-0.34***	-0.39***	-0.52***	-0.49***	-0.60***	-0.82***	-0.47***	-0.21***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.01)
R-squared	0.81	0.83	0.66	0.62	0.59	0.59	0.13	0.72	0.85
Sample: From Sep 6, 2020 to	Dec 5, 2020	Nov 14, 2020	Dec 19, 2020	Dec 12, 2020	Feb 13, 2021	Feb 6, 2021	Jan 23, 2021	Nov 14, 2020	Apr 17, 2021

Notes: Safegraph, Burbio and Return2Learn data for counties or district with at least 5 schools. All regressions are weighted by student enrollment at the county (Burbio) or district (Return to learn) level. Standard errors are clustered at the county (Burbio) or district (Return to learn) level. In columns (1) and (3), all counties and districts with available school visits data are included (subject to the sampling restrictions for schools with sparse or noisy visits data described in Section (2)). In columns (2) and (4), all counties and districts with data for at least 5 schools are included

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).

C.3 Relation between EIPL and cities vs. rural areas

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 very robust. As shown in Figure C2, 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 much 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 C2 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.

C.4 Description of school-level regression variables

In Section 5, we measure the effects of local and school/district variable on EIPL by looking at the effects of going from the 25th to the 75th percentile of the distribution of a variable. Table C2 presents descriptive statistics for the sample of public and private schools used in these regressions. In particular, observe that private schools are on average located in more affluent areas, have a lower proportion of nonwhite students and are much smaller than public schools in terms of student enrollment.

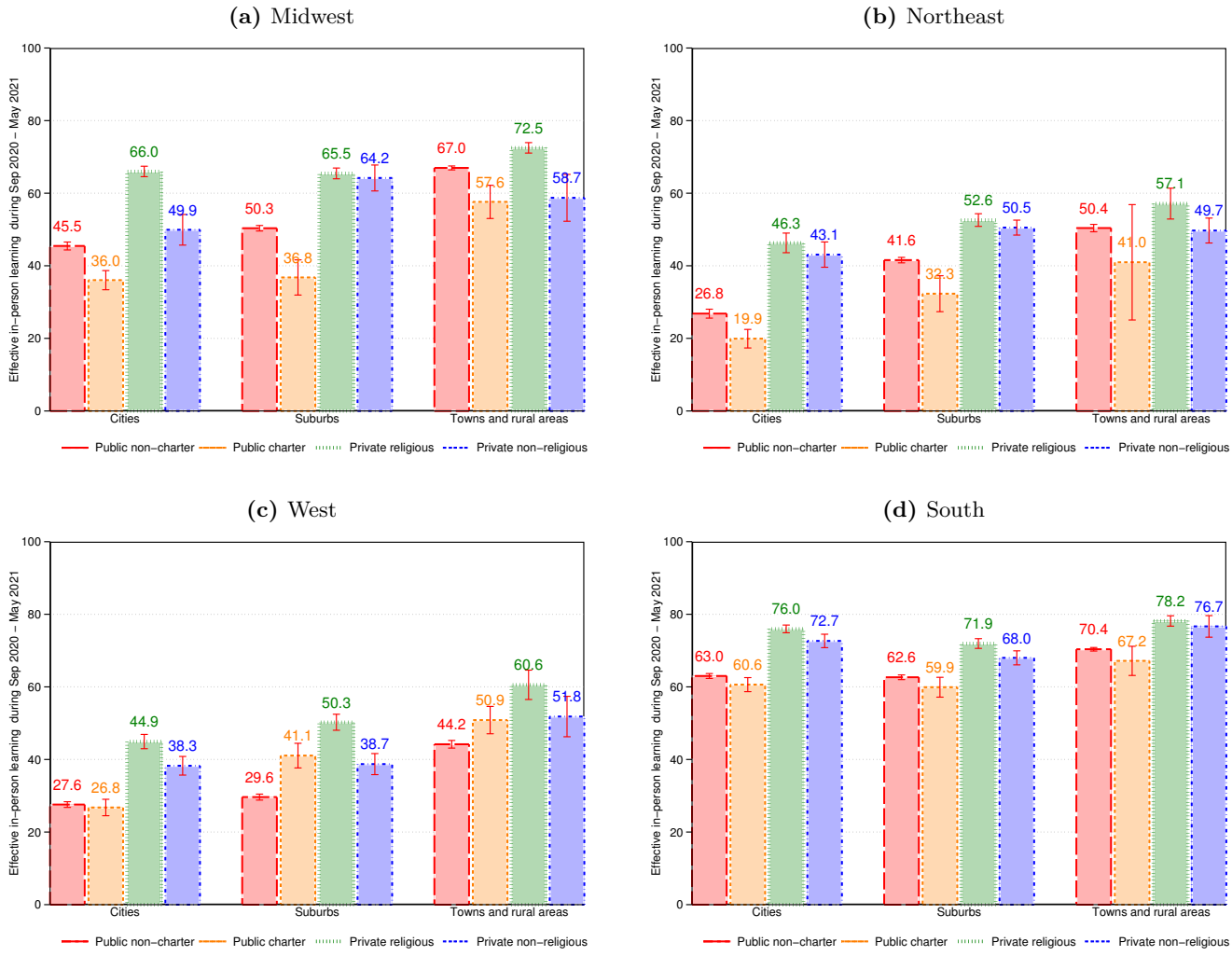
Table C2: Descriptive statistics of the school-level regression variables

	Mean (1)	St. Dev. (2)	25th pctile (3)	50th pctile (4)	75th pctile (5)	Min. (6)	Max. (7)
				(a) Public schools			
Zip-level average household income	75,046	29,740	55,784	67,297	86,235	7,770	432,067
Zip-level share of college educated	0.56	0.15	0.46	0.55	0.67	0.11	0.97
School neighborhood poverty	318	159	210	278	382	40	997
School share of non-white students	0.47	0.32	0.18	0.42	0.75	0.00	1.00
School spending per student	12,321	4,476	9,413	11,336	14,087	112	49,957
ESSER funding per student	3,228	2,390	1,534	2,804	4,177	0	30,189
Student enrollment	616	467	339	503	734	6	8,327
District number of schools	44	93	5	12	39	1	735
				(b) Private schools			
Zip-level average household income	86,503	39,765	59,828	75,054	103,386	22,512	397,509
Zip-level share of college educated	0.62	0.15	0.51	0.62	0.74	0.15	0.97
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 (pctile), 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.3.

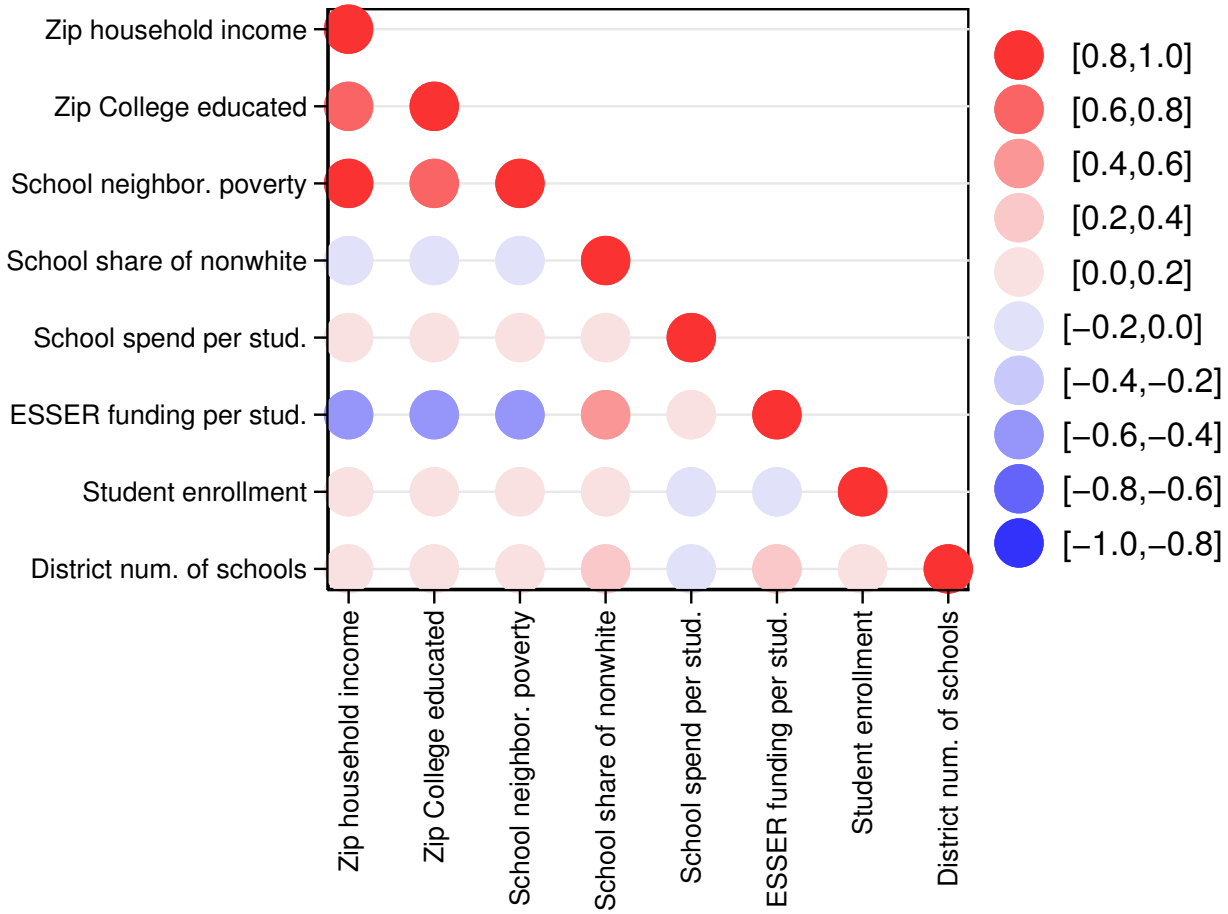
To complement panel (a) of Table C2, Figure C3 presents the correlations between the regressors. As the top left corner shows, the affluence measures are highly correlated with each other, with correlation coefficients ranging between 0.71 and 0.83. The school share of non-white students is negatively related to the affluence measures. The correlations range from -0.19 (correlation with school neighborhood poverty) to -0.10 (correlation with zip-level average household income). 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 -0.5 to -0.4 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).

Figure C2: 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).

Figure C3: 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.3.

C.5 Additional regression results

Regression results for private schools. Table C3 presents regression results for the sample of private schools. Columns (1) and (2) of the table repeat the results of panel (b) of Table 4 for reference. Columns (3) and (4) add controls for schools size, as measured by student enrollment, and indicators for whether a school is located in a city, a suburban, a town, or a rural locale. In line with our analysis of the effects of school/district size for public schools, these controls have no impact on the measures of local affluence and race of the school's student body. The effect of school size is generally negative but imprecisely estimated. In Columns (5) and (6), we introduce fixed effects for the county in which the school is located. As with our analysis of public schools, we find that the role of affluence is substantially reduced (the coefficient on average 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 above 20%. In sum, Table C3 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.

Table C3: Robustness: The importance of geography for private schools

Dependent variable	Effective in-person learning (EIPL)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Zip-level average household income	-4.51***		-4.19***		-0.17	
	(0.40)		(0.43)		(0.45)	
Zip-level share of college educated		-5.17***		-4.52***		-1.45**
		(0.56)		(0.60)		(0.60)
School share of non-white students	-5.75***	-6.03***	-5.22***	-5.57***	-3.23***	-3.57***
	(0.59)	(0.60)	(0.58)	(0.59)	(0.52)	(0.53)
Student enrollment			-0.48	-0.62*	-0.38	-0.33
			(0.34)	(0.34)	(0.25)	(0.25)
School type and grade controls	✔	✔	✔	✔	✔	✔
Locale FE			✔	✔	✔	✔
County FE					✔	✔
R-squared	0.04	0.04	0.04	0.04	0.21	0.21
Observations	11,280	11,280	11,280	11,280	11,280	11,280

Notes: Each column reports coefficients from a weighted OLS regression with standard errors in parentheses and school weights calculated as explained in Appendix A.3. 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 religious school and non-religious school. The school grade fixed effects consist of indicators for elementary vs. middle vs. high. vs. combined school. The locale FE consists of indicators for cities, suburbs, towns and rural areas.

County-week fixed effects regressions. Table C4 presents several robustness checks of the regression of the county-week fixed effects against party affiliation, vaccination rates, and COVID health controls. The first column repeats column (5) of Table 7 for reference. In Column (2), we use the Republican vote share in the 2016 presidential election among our proxies for the general stance towards reopening schools. The results barely change, which is unsurprising given the strong persistence in county-level Republican vote shares in the 2016 and 2020 presidential elections. In Columns (3) and (4), we change the number of time lags used to measure COVID vaccination, infection and death rates. When using contemporaneous values as is done in Column (3), the effect of vaccination rates is less pronounced – it is reduced by half –, and infection rates exert a negative effect on EIPL. The effects of party affiliation of

the state governor and of the Republican vote share remain unchanged. On the other hand, with a lag of one month (Column (4)), the effect of COVID vaccination rates is very close to the baseline estimates. It is unclear how bets 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. Last, in Column (5) of Table C4, we assess the sensitivity of the results to sample size, by restricting the analysis to counties with at least 10 public schools. This reduces the sample size almost threefold. The results remain unchanged, with exception of pre-pandemic ICU bed capacity that exert a positive role on EIPL.

Table C4: Robustness: Accounting for systematic geographical differences

Dependent variable	County-week fixed effect				
	Baseline	2016 votes	No lag	4-weeks lag	$\geq$ 10 schools
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Republican governor	17.08***	17.04***	16.84***	16.75***	16.76***
	(1.71)	(1.71)	(1.69)	(1.70)	(1.88)
Share of Republican voters	11.34***	11.52***	11.59***	10.97***	12.44***
	(0.96)	(0.93)	(0.95)	(0.93)	(1.07)
Lagged COVID vaccination rate	9.26***	9.30***	4.65***	8.14***	10.12***
	(1.31)	(1.31)	(1.19)	(1.18)	(1.49)
ICU bed capacity	0.81*	0.76*	0.86*	0.79*	2.24***
	(0.48)	(0.45)	(0.51)	(0.47)	(0.48)
Lagged COVID case rate	-0.02	0.13	-1.87***	0.73*	0.12
	(0.33)	(0.34)	(0.32)	(0.42)	(0.37)
Lagged COVID death rate	0.44*	0.54**	-0.08	0.80***	0.37
	(0.24)	(0.24)	(0.22)	(0.24)	(0.33)
Maximum weekly temperature	2.40**	2.65***	2.58***	3.06***	2.07*
	(1.00)	(0.98)	(0.88)	(1.09)	(1.07)
NPI controls	✔	✔	✔	✔	✔
R-squared	0.43	0.43	0.42	0.42	0.46
# of counties	2,814	2,814	2,814	2,814	1,078
# of weeks	35	35	35	35	35

Notes: Each column reports coefficients from a weighted OLS regression of the county-week fixed effects, with standard errors in parentheses and observations weighted by county population. The county-fixed fixed effect is extracted from the regression of EIPL against school and local characteristics (see Column (5) of Table 6) for the period from September 2020 to May 2021. The NPI controls consist of county-week indicators for stay-at-home orders, restrictions on public gatherings and mask mandates.