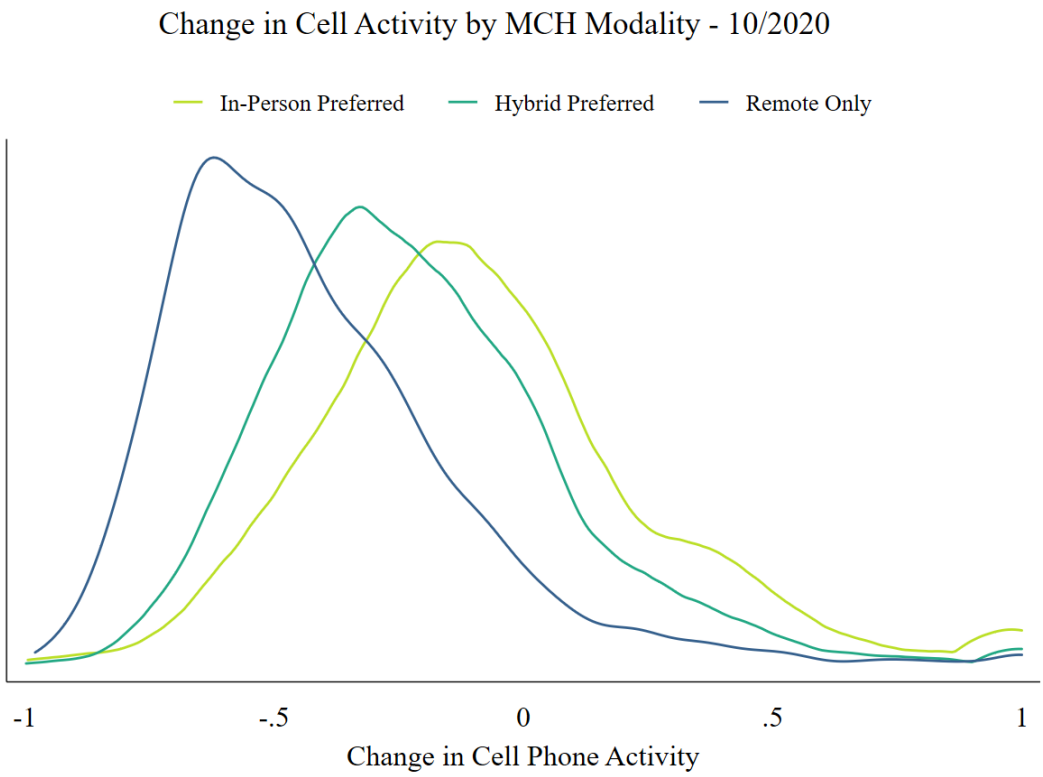
Appendix A: Validation of Learning Options Exposure Estimates

A crucial component of our analysis is the ability to model the availability of in-person learning options to respondents. As shown in our results section, our supply measures of remote learning was significantly related to reported mode of attendance. Data used to create this variable comes from a private company's phone survey of districts. We chose this survey over others because it is, to our knowledge, the most comprehensive source of district reopening decisions. However, as with any survey, there is the risk of measurement error occurring during the data collection process. The survey used to construct this variable does produce slightly different topline numbers (i.e., percent of districts using a remote/hybrid/in-person plan) than surveys collected elsewhere like Burbio's K-12 School Opening Tracker (https://cai.burbio.com/school-opening-tracker/). We believe this may be due to a combination of factors such as differences in coding and sample size.

To validate that our survey data accurately captures the instructional models being used in districts, we utilize innovative cellphone geolocation data based on measures of in-person visits to school buildings developed by Parolin & Lee (2020). In their paper, the authors construct a measure of changes in visits to school buildings relative to a pre-pandemic baseline using anonymized cellphone geolocation information. Importantly, the authors conduct a series of five validation checks to ensure that the data accurately reflects changes in visits to schools. If the survey we use to construct remote learning exposure is accurate, we would expect both the cellphone and survey data to exhibit the same relationship. We show this association in Figure A.1 using overlapping density plots of the relative change in cellphone activity for each district. This visualization shows that the greatest decreases in cellphone activity are among districts classified as remote only by our survey data of school districts, with the least decrease among districts classified as in-person. The similarity of these two separate sources of school re-opening information indicates that our constructed remote only supply measure is likely valid.

Figure A.1 – Changes in Cell Activity by MCH Modality (October 2020)



Appendix B: Technical Appendix

This technical appendix explains in detail the source and construction of variables used in our analysis and provides additional summary statistics representing the correlations between variables. Below are detailed accounts of how each variable was constructed as well as where data used to construct that variable came from. Where appropriate, results of principal-component factor analyses are presented and discussed. Unless otherwise noted, all proportions are in reference to the analytic sample defined in section 3 of the main paper. Tables B.1 and B.2, located at the end of this appendix, provide correlations among respondents' traits and school characteristics.

Race

Our race/ethnicity variable is constructed from a set of binary indicators for race (white, Black, Native America, Asian, Native Hawaiian or Pacific Islander) and ethnicity (whether the respondent identifies as being Hispanic or Latino) included in the UCA wave 15 data. The data for these variables is elicited quarterly as part of the UAS with respondents' most recent data being incorporated at the time of the wave 15 survey. We define a categorical variable to identify respondents as non-Hispanic white, non-Hispanic Black, Hispanic/Latino (of any race), or another race.

Education & Income

Like the race variable described above, respondents' income and level of education are captured quarterly and automatically incorporated into the wave 15 data file at time of survey administration. These factors are captured in two 16-level categorial variables. We collapse this information into two, three-level categorical variables. The resulting education variable categorizes respondents as having a high-school degree or less (36%), some post-secondary education with no degree earned (17%), and some post-secondary degree earned (46%). The resulting income variable categorizes respondents as having a household income less than $40,000 (33%), between $40,000 and $100,000 (41%), and over $100,000 (26%).

Employed

This variable indicates if respondents are currently employed or on leave (sick leave, FMLA leave, etc.) from their employment at the time of the wave 15 survey. Any respondents who indicate a labor status other than employed or on leave are categorized as unemployed including individuals that are retired or disabled.

Grade Level

In wave 15, respondents are asked to identify which grade the randomly selected child they are asked about is in. The options range from kindergarten to 12^{th} grade. We construct a categorical variable with three levels. Children in fourth grade and lower are categorized as attending an elementary school. Middle school children are defined as being in 5^{th} – 8^{th} grades. High school children are defined as being in 9^{th} – 12^{th} grades.

School Quality

We construct our school quality measure using respondents' ratings of their child's school across seven dimensions. These questions are retrospective and ask respondents to assess their child's school prior to the pandemic. Using a 5-point scale, respondents indicate the overall quality of education the school provided as well as the quality of feedback offered to students, quality of instruction in core subjects (science, math, ELA), quality of teacher-student relationships, quality of student engagement, and ability of the school to keep their child healthy. We first reverse code these responses so that a higher value indicates higher quality, and then verify that they capture a unique dimension of school quality using principal components factor analysis. The results of this analysis are shown below. These results show that all school quality questions load onto a latent variable and can be combined. We construct this latent variable using an orthogonal rotation of the factor analysis results.

Factor Analysis for School Quality Measure

	Factor	Uniqueness
Quality of Education	0.8205	0.3267
Quality of Teacher Feedback	0.7768	0.3966
Quality of Science Instruction	0.8175	0.3318
Quality of Mathematics Instruction	0.8432	0.2891
Quality of ELA Instruction	0.8474	0.2820
Students' Relationships with Teachers	0.7771	0.3961
Keeping Students Engaged	0.8136	0.3380
Keeping Students Healthy	0.7020	0.5073

Media & Public Health Trust

To capture media and public health trust, we use data from wave 7[1] of the UCA which asked respondents to rank the trustworthiness of various sources. When selecting news organizations that might contribute to a media trust measure, we excluded two that are viewed as heavily politically aligned to either the Republican or Democratic party as we believed that our election polling data would better capture the influence of such political polarization. We thus devise a media trust variable consisting of respondents' assessment of the trustworthiness of ABC, CBS, CNN, NBC, and national newspapers and a public health trust variable using respondents' assessment of the trustworthiness of the U.S. Department of Health and Human Services (HHS), the Centers for Disease Control and Prevention (CDC), and public health officials generally. We verify that these can be combined using principal-component factor analysis and report these results below.

[1] Wave 7 was conducted between June 10th and July 8th, 2020.

Factor Analysis for Media Trust			Factor Analysis - Trust in Health Orgs.		
	Factor1	Uniqueness		Factor1	Uniqueness
ABC	0.9389	0.1185	HHS	0.8846	0.2176
CBS	0.9384	0.1195	CDC	0.8859	0.2152
CNN	0.8992	0.1914	Officials	0.8570	0.2655
NBC	0.9447	0.1075			
Newspapers	0.8944	0.2000			

Politics

To categorize respondents as likely Trump voters, likely Biden voters, third-party voters, or undecided voters we rely upon UAS's 2020 presidential election tracking poll[2]. We use responses to a question asking, "If the election were held today, who would you vote for?" We recode this variable so that an intention to vote for third-party candidates or candidates that were not running in the 2020 presidential election are supporters of "other candidates." The election tracking poll was administered to respondents approximately every two weeks from August of 2020 until election day in November of 2020. While choice of candidate remained stable among respondents during that time, we select the election poll response that was submitted closest in time to each wave 15 response. Nearly one-third of respondents completed the election poll and wave 15 survey on the same day and over 91% of respondents in the analytic sample completed both within one week of each other.

Perceived COVID-19 Risk

Our measure of perceived COVID-19 risk is constructed from three questions in wave 15 asking respondents to report their perceived probability of contracting COVID-19 in the next three months, being hospitalized because of, and dying from it. We confirm that these three values may be combined into one variable using principal components factor analysis and report the results of this analysis below. While the reported risk of being

[2] https://uasdata.usc.edu/index.php

infected with COVID-19 does appear to be slightly more unique than hospitalization and death, all variables load onto the same latent factor. Following this factor analysis, we use an orthogonal rotation of the results matrix to predict values for the perceived risk index.

Factor Analysis for Perceived Risks.

	Factor	Uniqueness
Risk Infection	0.7984	0.3625
Risk Hospitalization	0.9291	0.1369
Risk Death	0.9156	0.1617

Incidence Rate & Case-Fatality Ratio

We incorporate measures of COVID-19 prevalence and lethality into our analysis using county-level data from the New York Times and U.S. Census Bureau. We use the county-level COVID-19 case and death counts as of the first day of the wave 15 survey (September 30th, 2020) and the estimated population of each U.S. county prior to the pandemic to construct these measures. We define incidence rate as the proportion of a county's pre-pandemic population which has tested positive for COVID-19 and the case-fatality ratio as the proportion of individuals who have tested positive and are confirmed to have died from COVID in that county. To preserve respondent anonymity, we round both values to the nearest tenth prior to merging with UCA respondents.

Urbanity

The three-level urbanity variable we use in our analysis is provided in the UAS election polling data file. The levels – rural, mixed urban, and urban, are defined based upon the proportion of census-designated urbanized areas in a respondents' census-designated zip code tabulation area (ZCTA). If there are no urbanized areas, the respondent is identified as living in a rural location. If there are only urbanized areas in a respondents' ZCTA, they are classified as living in an urban location. Respondents with a mix of urbanized and non-urbanized areas are classified as being "mixed urban."

School Type

In wave 15, respondents indicated if their randomly selected child attended a neighborhood public school, magnet public school, charter school, private/religious school, were homeschooled, or attended another type of school. We recode magnet public schools to be included with neighborhood public schools. Additionally, we remove any respondents who indicate homeschooling on this or other questions. Our resulting analytic sample is composed only of students attending public (86%), charter (6%), and private (8%) schools.

Given a Choice of Modality

Wave 15 respondents who do not report homeschooling their child are asked "Is your child's mode of attendance right now based on a choice your school gave you?" We recode answers to this question to form a binary variable and exclude respondents (N=59) who indicate they are unsure if their school gave them a choice of learning modality from our analysis. Including these respondents and coding them as not being given a choice of modality doesn't change the results significantly.

Share of Schools Only Offering Remote

Using a large database of district reopening plans provided by MCH strategic data[3], we identify learning modalities offered in 78% of public school districts nationwide at the time of the wave 15 survey. For details about how we validated these data, see Appendix A. We then match school districts to census tracts using a crosswalk from NCES[4] and construct a continuous variable representing the prevalence of remote-only learning weighted by district enrollment in each tract. For census tracts with one modality or one district, this becomes a binary variable. While we can identify if districts offer hybrid or in-person learning, we select the remote-only measure as it is less prone to measurement

[3] https://www.mchdata.com/covid19/schoolclosings
[4] https://nces.ed.gov/programs/edge/Geographic/RelationshipFiles

error and more accurately reflects the supply of learning options available to students in that district.

State

Respondents indicate their state of residence on their quarterly update survey. We use this information to include state fixed effects in specifications B-D of our analysis.

Correlations Between Variables

Tables B.1 and B.2 show correlation tables across respondents' traits and school characteristics, respectively. Overall, we observe considerable independent variation that allow identification of our models. The highest correlations among respondents' traits are observed between the media trust and public health trust indexes as well as with the political leanings. Across school characteristics we observe strong correlations between the prevalence of remote learning in local public schools and our outcome variable of mode of learning but no other strong correlations across explanatory variables. Overall, these correlations mitigate any concerns of multicollinearity issues in our models.

Table B.1

Correlations Among Respondents' Traits

Rural	1.00											
Mixed-Urban	-0.47	1.00										
Urban	-0.33	-0.68	1.00									
HS or Less	0.17	-0.14	0.01	1.00								
Ed: Some College	-0.04	-0.02	0.05	-0.25	1.00							
Ed: College Grad	-0.10	0.13	-0.05	-0.61	-0.61	1.00						
Employed	0.04	0.01	-0.05	-0.21	-0.09	0.25	1.00					
Biden Voter	-0.23	-0.04	0.23	-0.09	0.01	0.07	-0.02	1.00				
Trump Voter	0.23	0.07	-0.27	0.04	-0.01	-0.03	0.04	-0.76	1.00			
Media Trust	-0.14	-0.04	0.15	-0.10	0.06	0.03	-0.07	0.46	-0.43	1.00		
Public Health Trust	-0.16	-0.01	0.14	-0.12	0.03	0.08	-0.01	0.36	-0.33	0.52	1.00	
Incidence Rate	-0.15	-0.14	0.27	0.05	0.08	-0.11	-0.04	0.16	-0.19	0.07	0.06	1.00

Table B.2

Correlations Among School Characteristics

In-Person	1.00							
Remote	-0.70	1.00						
Hybrid	-0.29	-0.49	1.00					
School Quality	0.06	-0.07	0.01	1.00				
Public School	-0.05	-0.03	0.10	-0.05	1.00			
Charter School	-0.12	0.16	-0.07	-0.02	-0.64	1.00		
Private School	0.18	-0.11	-0.07	0.08	-0.72	-0.08	1.00	
Given Choice of Modality	0.19	-0.29	0.15	-0.04	0.10	-0.12	-0.02	1.00
Remote Learning Prevalence	-0.38	0.48	-0.18	-0.02	-0.10	0.13	0.02	-0.24	1.00

References

Parolin, Z., & Lee, E. K. (2020). Large socio-economic, geographic, and demographic disparities exist in exposure to school closures and distance learning. *OSF Preprints*. https://doi.org/10.31219/osf.io/cr6gq