The Effect of COVID-19 School Closures on College Enrollment

Garrett Gonta

University of California San Diego

Econ 191: Senior Essay Seminar

Professor Emanuel Vespa and Professor Julie Berry Cullen

December 16, 2023

The Effect of COVID-19 School Closures on College Enrollment

Background

Attending college is a huge step towards the future for many Americans and the decision to enroll in university has a huge effect on the life they will eventually live. Whether a person might go to college to get a job that might pay more, be in an industry they might enjoy, or for simply an impactful social experience, the decision to go to college will be life-changing for those who do. College as a whole is a huge decision for many, especially in the academic sense. Whether someone isn't sure they are academically prepared for college or whether they had a positive or negative experience within their primary education will affect a student's decision to attend university in a significant way. School closures due to COVID-19 and the Global Pandemic were incredibly significant to these factors and would likely have some sort of effect on whether people decided to attend university. Analyzing different learning styles employed during the Pandemic is significant in this case because it draws important clues to human behavior and decision-making when it comes to large significant changes in the way students learn. The presence of social distancing during the Pandemic on K-12 education likely affected students' view of college and education as a whole. These factors and potential connections between the COVID-19 pandemic and someone's decision to enroll in a university are quite valuable to study and knowing how the Pandemic might have affected these decisions would give key details to the human condition.

Research Question

How did School Closures caused by the COVID-19 Pandemic affect College and University enrollment in the Short and Long Term?

Literature Review

Current research conclusions on this topic many times contradict each other leading to differing deductions about the effect of the virus and the various education reforms associated with it concerning university new student enrollment and continuing re-enrollment. The National Student Clearinghouse using data for enrollment in the Fall of 2019 and 2020 found interesting changes in college enrollment rates for students before and after COVID-19. The organization found using statistics that there was a 3 percent decrease in overall college enrollment, a 16.1 percent decrease in freshman enrollment, and a 22.7 percent decrease in freshman enrollment for community colleges (National Student Clearinghouse, 2020). This data does support that there was an overall decrease in enrollment during this period but also has many flaws in its implementation. Firstly this data is found only using statistical methods, meaning that it gives no information on any possible causal relationship between the Pandemic and college enrollment. Also, this data cannot reveal any possible variables that can be causing these changes in enrollment such as the widespread change in school learning models. Due to the data only having enrollment numbers for Fall 2020 and Fall 2019, this data can only focus on the immediate changes in enrollment between these two years. Although these data conclusions do provide the potential baseline for some sort of causal effect, this data would have to be expanded on through an econometric lens to draw any useful conclusions. A study by Economic Professors at the University of California Santa Cruz, George Bulman and Robert Fairlie, analyzed how COVID-19 affected community college enrollment in California and found similar conclusions. The researchers used an exploratory data approach to find that there was a reported 11% decrease in community college enrollment for this state between fall 2019 and fall 2020, and a 7% decrease in enrollment from fall 2020 to fall 2021 (Bulman & Fairlie, 2022). The researchers dive deep into student demographics, fields of study, and transferability of courses to analyze the various, specific enrollment changes over these year periods. Similarly, Bulman and Fairlie in their study do not try to find any potential causes or correlated variables for this enrollment decrease. These studies act as an important baseline for understanding the existence of enrollment changes from before to after the Pandemic, but this data needs to be expanded on to be able to notice the actual significance of these differences in enrollment and any possible factors that led to these observed changes.

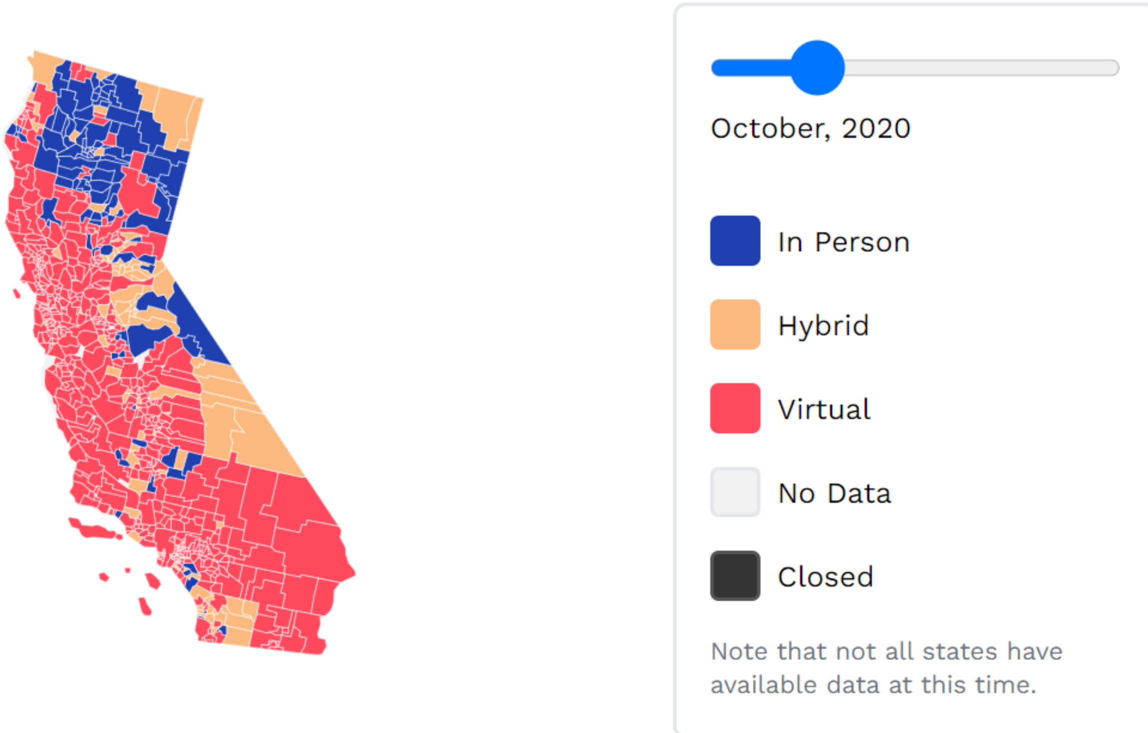
While some data-based studies, as looked at above, concur with the idea that enrollment decreased across the board, survey-based data at the school level found contradicting conclusions. One study disagrees with both the previous conclusions and argues that COVID-19 caused no significant change in people's decisions in continuing enrollment. The paper that points this out is "Investigating the Influence of COVID Related Worry on University Enrollment Intentions" by Christopher Thomas and Kristie Allen. This paper is quite intriguing with its choice of model that it uses to analyze this effect. This paper uses a psychological, survey-based methodology that asked a group of students from a university whether or not they were planning on re-enrolling in the university for Fall 2020 and whether their personal worries about contracting the virus had any effect on re-enrollment plans. This study found that the sample's opinion over COVID had not changed their re-enrollment plans and even concluded this for those most worried about the virus (Thomas & Allen, 2021). The piece identified that for this sample of college students, COVID had not impacted their plans to re-enroll next fall, leading to evidence that college enrollment wouldn't change at this university. This survey-based study leaves much to be desired in terms of research. Because the information was collected via a survey, students who stated they were going to re-enroll for the fall semester might have not after all. This survey also consisted of a small sample of students at a single college which won't tell an accurate story of actual enrollment changes across all American universities. A similar study was conducted by Economic researchers, Estaban Aucejo of Arizona State University, Jacob French of New York University, and Basit Zafar of the University of Michigan, which had some differing results. This piece used a survey-based approach to find what experiences or activities Arizona State University (ASU) students were willing to pay for after the COVID outbreak and how much certain experiences are worth to those students. They figure this out using a willingness-to-pay (WTP) approach where the researchers identify the worth of certain activities on campus by looking at how much students are willing to spend on that activity compared to the average cost of attending the university. Using this method, the researchers found that the students' WTP for in-person instruction was 4.22% of the average net cost of attending ASU, and their WTP for in-person social activities was 8.06% (Aucejo et al., 2023). These values are fairly significant, meaning that if the data can be assumed for the rest of the university's students, the average ASU students' willingness to pay for in-person activities was 12.28%. Although this only represents a smaller portion of the net cost of attending the university, this is definitely baseline evidence for a possible correlation between school closure status after COVID-19 with the decision of whether someone would want to enroll in a university given schools are not open. With this type of data, one could assume fairly significant enrollment decreases when the university is either virtual or closed. Despite these favorable conclusions, the sample of 1,150 students at ASU used for this research gives large limitations to what can be said about people's enrollment decisions after COVID across America. The piece gives the closest data and information to establish a connection between school closures and college enrollment, but this is simply a conclusion that can be inferred from the paper as the researchers only focus on what the students are willing to pay for university rather than looking at actual changes in enrollment because of the change in learning models. This piece will focus on building upon or countering the ideas in these research papers using an economic lens and focusing on school closures during the COVID-19 Pandemic. The work in this study will fill the gap that is created when analyzing issues using only statistical analysis or survey-based data collection by using econometric modeling to find a possible causal effect of the sudden change in school learning models during this world-changing virus on the decision to attend college.

Hypothesis

When thinking about how changes in learning models due to the Pandemic affected college enrollment, it would be important to study the short-term and long-term effects separately leading to a split hypothesis. One could assume from the data presented that, in the short-term, enrollment would likely decrease as a result of school closures with varying effects depending on the schools' closure status. It would be assumed that the further away a school was from in-person instruction, the larger these enrollment decreases would be. Thus it would be assumed that the least to most affected would be schools that are hybrid, then virtual, and finally closed. In the long term, hybrid and virtual schooling would likely have little to no effect on college enrollment in future years, especially when compared to schools being closed which most likely would result in a negative effect on university enrollment in the long term. This comes down to assumptions over curriculum and knowledge gaps over COVID-19. Although learning experiences were not the same between hybrid and virtual schooling when compared to in-person instruction, the same or very similar curriculum still came through to the students during this time. Students of schools that were closed during the pandemic likely would have missed out on quite a lot of important instruction and probably would feel unprepared for university. Thus looking at the long term, the hypothesis comes down to mostly curriculum losses based on the different learning models.

Data Description

When conducting an economic study to analyze how enrollment rates changed as a result of pandemic school closures, it is extremely significant to find data that would be able to evaluate these differences. Arguably the most important data would be the base of this project, school closures during the COVID-19 Pandemic. This data was found using the COVID-19 School Data Hub (2023), which compiled reports from schools and school districts across the country on which learning model they were using during the 2020-2021 school year. The piece focuses on this school year as this entire period was affected by COVID and had the most variation between state policies compared to just looking at the Spring of 2020 where schools were mostly closed all across the country or the 2021-2022 school year which was primarily open for a lot of areas in the U.S. Because the study is trying to dive into differences between instructional styles, this school year will be the most imperative to study. The data set identifies four major learning models that were used during this time, in-person instruction, hybrid schooling, virtual learning, and closed schools. This data was documented at the school or district level depending on individual state mandates for this type of reporting. This also goes for the frequency of disclosing this data as the data was reported at the weekly, bi-weekly, monthly, quarterly, yearly, or bi-annually level depending on the state. Seen below is a visual representation of the learning model data for California.



As shown, the California data visualization is represented at the school district level and has monthly reporting. California, as visualized, was largely imploring virtual or hybrid models throughout the school year. Because of the discrepancies in the reporting of this data due to state differences, and the fact that most enrollment decisions come mostly in the fall of each year, this study will focus on school years as a time parameter. School years in this study are identified as starting in July of a year and ending in June of the following year. For example, data in the 2015-2016 school year would be data reported from July 2015 to June 2016. Also due to reporting in some states coming from school districts and others at the school level, this data will be centralized around the school district level which will still be plenty specific enough to draw possible conclusions from, while also making cross-America data analysis possible. This data will be important in establishing consistent independent variables needed for the implementation of this study.

Another extremely important piece of this project would be identifying changes in college enrollment over time. The data for this is best collected using the Integrated Public Use Microdata Series Current Population Survey (IPUMS CPS) (Flood et al., 2023), which runs surveys across the US to find a variety of data, including college enrollment data. From this data set, enrollment data at the individual level, organized by counties, was found from the years 2012-2023. Data was chosen to start in the 2012-2013 school year specifically because it would rule out any possible bias from any prominent recent events in the late 2000s and early 2010s such as the recession of 2008 which could have caused some unnecessary biases. The data will also be cleaned thoroughly to help find changes that COVID-19 school closures could have caused. The first major cleaning would be limiting the age group that this study is looking at. The National Education Center for Education Statistics reports that 82% of college students during Fall 2021 were aged 18-34 (U.S. Department of Education, 2021). Thus this piece will be looking at only those ages because that demographic would be the most likely to attend university and thus give us less noise when looking for changes in enrollment due to school closures. Another important factor that changes people's chances of going to college is their highest educational attainment. Thus this piece is limited to only individuals who graduated high school or have some years of college but have no college degree. This is identified during the school year the data is collected from. This inclusion will further help to limit the sample to get a better idea of how school learning models affected the choices of enrolling in college from those most likely to be deciding whether or not to attend. The enrollment data is represented through the variable “dumEnroll” which is equal to 1 when someone is enrolled in university either part-time or full-time, and a 0 if they are not enrolled at all. The purpose of using a binary variable for this study will be explained further in the methodology. The data as stated before is represented at the individual level but to compare school closures for these individuals, the study needs to be zoned by locations. The most useful location information from the IPUMS CPS for this study would be looking at the county level, this way the data is specific enough to identify differences across a certain state while also limiting bias, explained more in the methodology. A summarization of this data is shown below for Alameda County, CA for better understanding.

statefip	county	schYear	dumEnroll
california	6001	2013	.3165468
california	6001	2014	.300813
california	6001	2015	.4262295
california	6001	2016	.5
california	6001	2017	.4565217
california	6001	2018	.4029851
california	6001	2019	.3516484
california	6001	2020	.2605634
california	6001	2021	.4117647
california	6001	2022	.4037267
california	6001	2023	.2521739

This summarization shows that each county, with sufficient enough reporting, has data for school years from 2012 to 2023 and contains enrollment data for each of those counties. However, it is important to note that in the actual data set, the enrollment data is not summarized by county and school year as it is in the visualization and in fact is identified at the individual level. Thus this data set contains information for individual enrollment for 18 to 34-year-olds with high school diplomas for each school year in each county.

One important discrepancy between these two data sets is the locational data. As explained above, the COVID-19 School Data Hub's school learning model information is represented at the school or district level, while the IPUMS CPS data is grouped by all counties across the United States. Because these two sets' locational data don't align, a third dataset is used just for the data merging portion. The National Center for Education Statistics (U.S. Department of Education, 2023) contains important translational data in their set “School Locations & Geoassignments.” This set contains the county location for all school districts across the United States. Thus this set is used as a translator, changing the COVID-19 school closure data from being grouped at the school district level to the county level. This merge allows school closures to be represented at the county level, which enables the merging of the learning model data and the enrollment data. To perform this merge efficiently it is important to note that school learning models are averaged for each county to allow lossless enrollment data merging. Thus this third data set, although not necessarily adding variables of information, is imperative to reaching the final dataset that this study will explore.

After all of these merges there needs to be some slight data cleaning and adding until the final set is reached. The first step, before the summarization of closure data explained above, is to add variables to identify learning model usage in the study. This is done using dummy variables. The dummy variables dumVirtual, dumHybrid, dumClosed, and dumOpen all identify whether a school was virtual, hybrid, closed, or open respectively in a certain period using a 1 or 0 to identify its status. Then at this point, the school closure data is averaged across the entire county for the whole school year. Thus with this collapsing of the data, the meaning of the dummy values change. Now these values represent the percentage of the 2020-2021 school year that the schools in each county were virtual, hybrid, closed, or open. This is best understood with the visualization below of Alameda County, CA once again.

schYear	county	dumVirtual	dumHyb...	dumClose	dumOpen
2021	6001	.909602	.072523	.009448	.008427

This splice of data explains that the schools in Alameda County were 90.96% virtual, 7.52% hybrid, 0.95% closed, and 0.84% open during the 2020-2021 school year. Also, the creation of dummy variables used to identify the previous years' school learning model is important to the short-term analysis, as further identified in the methodology. This is accomplished by creating dummy variables lastYearVirtual, lastYearHybrid, lastYearClosed, and lastYearOpen which are simply the data of the previous dummy variables shifted up by one year. After all of these pieces of data are spliced together and dummy variables are created, the data is finally cleaned through balancing. The balancing of the data is important in identifying what counties have all the data necessary for a proper model to be run. These requirements are that the county must have enrollment data from the school years between 2012-2023 and also have school learning model data from the 2020-2021 school year. Due to the data merging, and not consistent reporting from either the IPUMS CPS or the COVID-19 school data hub, only 188 counties fit the needed specifications to run this data. The implications of this will be explained in the limitations. After this last piece of data cleaning, the final set is created. Thus the final data set is best described as having school closure data represented through dummy variables that define the percentage of the year that a specific county was using a certain learning model which is matched up with the enrollment status of 18 to 34-year-old high school graduates that live in that certain county.

Method

As referenced above, looking at both the short and long-term effects of school closures will give a greater idea of the effect that changes in learning models had at different periods. This analysis will be achieved through the use of two vastly different models. When trying to find the short-term effect of school closures, it will be important to analyze the immediate next-year change that switching learning models caused. In the context of this study, because the school year when these closures happened was 2020-2021, when trying to study the immediate one-year effects of school closure status, the significant year of study for college enrollment will be the 2021-2022 school year. The short-term model, as shown below, is a fixed-effects model that evaluates changes in enrollment based on learning model usage while controlling across counties and school years.

$$dumEnroll_{c,t} = \beta_0 + \beta_1 dumVirtual_{c,t-1} + \beta_2 dumHybrid_{c,t-1} + \beta_3 dumClose_{c,t-1} + \gamma_1 dum2013_t + \cdots + \gamma_9 dum2021_t + \gamma_{11} dum2023_t + \delta_2 county2_c + \cdots + \delta_C countyC_c + u_{it}$$

The model's endogenous variable, dumEnroll, is a variable used to identify the likelihood, identified as a probability, of an individual from a particular county and school year to be enrolled in college based on the defined exogenous factors. The critical exogenous variables would be the dummy variables for each learning model. These dummy variables represent the percentage of time that a county's schools were either virtual, hybrid, or closed during the 2020-2021 school year, and are equal to 0 if not in that school year. A piece that is crucial to the model would be the "t-1" subscript for the school closure dummy variables. This inclusion allows for the study of changes in enrollment rates in college due to last year's learning model. This one-year change is the basis of the short-term model. The dummy variables' coefficients are the most significant piece to the study because they identify the extent of the effect that a learning model had on college enrollment in that county the next year. For example, if the coefficient for $dumVirtual_{i,t-1}$, β_1, is equal to 0.1, then if a particular county's schools were virtual for the entire school year, the average 18-34-year-old high school graduate of that county would be 10% more likely to attend college the next year compared to if the county's schools would have been in-person for that school year. This coefficient is what will be the heart of the study and will give key details about any significant changes in enrollment rates. The rest of this model represents the fixed effects part of the regression, in which the model controls for changes in enrollment for each county and each school year. This is key to being able to rule out most omitted variable bias in this model. By controlling for omitted variable bias in this way, any significant deductions about the dummy variables from this model are assumed to be casual. These fixed effects control for differences between counties such as differences in local legislature, distances from colleges, and attitudes towards college. And the fixed effects for the school years, simply control for differences in the enrollment rate between various years. The last factors of the model would be the intercept term and the error term. The intercept term identifies a baseline of college enrollment given the excluded variables in the model. Because the model is a dummy-fixed effects regression, it is easy to visualize that the excluded variables' effects described in the intercept term would be the effects of open schools, the 2020-2021 school year, and the first county in the data, Baldwin County, Alabama. The short-term model will analyze the effect of differences in school learning models on the next school year's college enrollment while controlling for omitted variable bias through fixed effects.

When trying to find enrollment changes in the long run due to differences in school closure models, a model must be made that would be able to analyze the effects of varying types of schooling years after the targeted school year of 2020-2021. The econometric model that would best determine these long-term changes would be an event study model. This model would compare changes in the enrollment rate due to school closures status to previous enrollment trends. This event study model is shown below.

$$dumEnroll_{c,t} = county_c + schYear_t + \sum_{k \neq -1} fixedVirtual_c \times 1\{t = k\}\beta_k$$

$$+ \sum_{k \neq -1} fixedHybrid_c \times 1\{t = k\}\gamma_k + \sum_{k \neq -1} fixedClosed_c \times 1\{t = k\}\delta_k + u_{c,t}$$

The endogenous variable in the event study is dumEnroll which again represents the estimated chance that a certain 18-34-year-old high school diploma-holding individual would be attending university given a certain county and school year. The exogenous variables and the summations they are included in are the impactful differences between this model and the short-term model because they represent the interaction between a certain learning model and time. The exogenous variables, fixedVirtual, fixedHybrid, and fixedClosed, all represent the percentage of the 2020-2021 school year that a county used a certain learning model. But unlike the short-term model, these percentage values aren't only represented in the target year and instead are fixed for the county across all the years of the study. This allows for the interaction between the learning models and specific school years. This interaction term means the regression will analyze the relationship between a county that used a specific learning model in the target year and each school year separately. This interaction will find the effects that certain learning models had on college enrollment given previous enrollment trends for those counties who used the same learning model. This will identify whether changes in enrollment due to a learning model were actually significantly different compared to previous trends and whether those changes lasted in the long term. This is done by creating a list of percentage dummy values for fixedVirtual, fixedHybrid, and fixedClosed for each year of study in the data. This is accomplished through the summations found in the model. In this study there was an analyzed 11 school years from 2012-2023, thus the summation terms will create 10 individual proportion variables for each learning model, one for each of the school years of study, excluding a baseline school year of 2019-2020. These 10 variables for each learning model will be compared with one another to check whether pre-event trends existed for areas that used certain learning models during the 2020-2021 school year. For there to be long-term effects of changes in learning models, the data must reveal that there was no difference in enrollment rate due to learning models from the years before the baseline year to the baseline year and that there were significant differences between the years after the baseline year and the baseline year. In this context, for there to be long-term effects of a particular type of schooling, the school years between 2012-2019 must be statistically the same as the 2019-2020 school year, and the school years between 2020-2023 must be statistically different from the 2019-2020 school year. This mathematically is described as that, for example for virtual schooling, $\beta_{-2} = \beta_{-3} = \beta_{-4} = \beta_{-5} = \beta_{-6} = \beta_{-7} = \beta_{-8} = 0$, and $\beta_1 \neq 0$ & $\beta_2 \neq 0$ & $\beta_3 \neq 0$ within a certain confidence level. If these assumptions for the coefficient between the virtual interaction terms are found to be true in the data, then the long-term model will find that there is in fact a long-term effect that virtual schooling had on university enrollment. The extent of this effect will be found in the coefficients of the after-event interactions. So in this case, the coefficients for the school years 2020-2021, 2021-2022, and 2022-2023. These coefficients, similar to the short-term model, will identify the effect of learning models on the change in the chance of enrollment for a demographic-fitting individual in a specific county for that school year. For example, if a county was completely virtual for the 2020-2021 school year and had virtual schooling interaction coefficients of 0.2 for 2020-2021, 0.1 for 2021-2022, and .05 in 2022-2023, assuming that these values are significant, the chance that an 18-34-year-old high school graduate would be attending college would increase by 20% for 2020-2021, increase by 10% for 2021-2022, and increase by 5% in 2022-2023. By being able to evaluate all the variables for each year separately, the model allows for the analysis of the long-term effect of the differing school learning models during COVID-19. Again, as stated above, many counties changed school learning styles throughout the target school year, thus for these counties, the effect that learning model changes had on the chance of enrollment for the demographic would be a combination of all the after-baseline coefficients for each of the learning styles. Thus, assuming the data is significant, the combined effect that virtual, hybrid, and closed schooling had on enrollment rates in the long term would be seen through coefficients, β_1, β_2, β_3, γ_1, γ_2, γ_3, δ_1, δ_2, and δ_3. Another important part of this data would be the fixed-effects variables for this model. These fixed effects are identified by "county" and "schYear" and these interact with the regression in the same way seen in the short-term model. These fixed-effects variables will account for any differences between counties and school years, thus allowing for any findings about the long-term effects of the types of schooling to be free of most omitted variable bias. The event study model, through the use of interaction terms, will see how learning models used during the 2020-2021 school year affected enrollment in the following years and whether or not the changes in schooling types had a significant impact on enrollment in the long-term.

Results

The results are also naturally split between the short-term and long-term methods. The short-term county and school-year fixed effects regression is calculated in Stata, shown below.

```
Fixed-effects (within) regression               Number of obs     =    223,924
Group variable: county                          Number of groups  =        188

R-squared:                                      Obs per group:
     Within  = 0.0020                                         min =        127
     Between = 0.0710                                         avg =    1,191.1
     Overall = 0.0027                                         max =     20,427

                                                F(13, 223723)     =      34.58
corr(u_i, Xb) = 0.0454                          Prob > F          =     0.0000

------------------------------------------------------------------------------
   dumEnroll | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
lastYearVirtual |   .0263328   .0081586     3.23   0.001      .010342    .0423235
 lastYearHybrid |   .0222291   .0095217     2.33   0.020     .0035667    .0408914
  lastYearClose |  -.1350402   .0914604    -1.48   0.140    -.3143003    .0442199
     dum2013 |   .0759769   .0070798    10.73   0.000     .0621007    .0898531
     dum2014 |   .0641958   .0070832     9.06   0.000     .0503128    .0780787
     dum2015 |   .0576053   .0071111     8.10   0.000     .0436678    .0715427
     dum2016 |   .0567602   .0072742     7.80   0.000     .0425029    .0710176
     dum2017 |   .0674473   .0072364     9.32   0.000     .0532641    .0816305
     dum2018 |   .0605166   .0073982     8.18   0.000     .0460164    .0750168
     dum2019 |   .0423393   .0074304     5.70   0.000     .0277759    .0569028
     dum2020 |   .0314032   .0059694     5.26   0.000     .0197033     .043103
     dum2021 |   .0205743    .005702     3.61   0.000     .0093986      .03175
     dum2023 |   .0151689   .0059528     2.55   0.011     .0035016    .0268363
       _cons |   .2351055   .0053884    43.63   0.000     .2245443    .2456667
-------------+----------------------------------------------------------------
     sigma_u |  .08374313
     sigma_e |  .43594096
         rho |  .03558816   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(187, 223723) = 30.18               Prob > F = 0.0000
```

This regression output displays a large amount of key data that provides a summary of the overall information as described in the methodology. The most important pieces of data in this regression would be the coefficients and t values for the dummy learning models. When trying to find the short-term effect that some learning models had on college enrollment it is important to contextualize the coefficients and t values. For virtual schooling, the computed coefficient is 0.0263 and the t-value is 3.23. This t-value reveals that the coefficient found is significant at the 99.8% level, meaning that virtual schooling in the 2020-2021 school year did in fact have a significant impact on the college enrollment rate in the 2021-2022 school year for the study's key demographic. The extent of this effect is found within the coefficient which is equal to 0.0263. This is best understood in percentage terms. This coefficient means for counties whose schools were completely virtual during the 2020-2021 school year, the chance that a random 18-34-year-old with a high school diploma from that county would be in college in 2021-2022 was increased by 2.63%. This may seem like a small change, but considering the baseline college enrollment rate in 2021-2022 was .235 or 23.5%, found using the coefficient for _cons, this 2.63% increase in the chance these individuals attend college equates to an increase in enrollment of 11.19% for 2021-2022 due only to the fact schools were virtual the year before. An important note considering this calculation of the percentage increase in enrollment in a county is that this metric would change based on the county fixed effect values. Thus this calculation of the increase in a county's enrollment is determined in the baseline county, meaning its value would slightly differ depending on the county of focus. A similar result is found for hybrid schooling. The coefficient for the dummy variable for hybrid schools was 0.0222, while the t-value was 2.33. This variable is found significant at the 98% level which is also quite high. The significant effect that hybrid had was defined by its coefficient of 0.0222. This value reveals that for counties with schools that were entirely hybrid during the 2020-2021 school year, there would be a 2.22% increase in any demographically fitting individual's chance of attending college the next year compared to if those counties' schools were in-person instead. This coefficient, when compared to the baseline intercept, helps to report an increase in 9.44% of the demographic enrolled in college the next year in the baseline county. This result also makes logical sense when compared with the effect of virtual schools. One would expect that the hybrid model, the in-between of in-person and virtual instruction, would naturally have an effect that is bounded by these two learning models. Because the baseline in this regression is open schools and the coefficient for virtual schools is 0.0263, it would be assumed that the effect of hybrid schooling would be between 0 and 0.0263. And in this model, this is indeed true. The final variable of interest, schools that were closed, had greatly contrasting results compared to the previous learning models. The regression discovered a coefficient of -0.135 and a t-value of -1.48. The observed t value for this dummy variable reveals only significance at the 85% level. This value is just slightly too low to confirm any strong causal conclusions about closed schools, especially when compared to the substantial significance levels found for virtual and hybrid. But an interesting piece of data found would be the coefficient for closed schools. This coefficient, although not largely significant, is quite impactful to the regression in terms of scale being equal to -0.135. This coefficient if assumed true, questionable due to only 85% significance, would mean a demographic-fitting individual of a county where schools were entirely closed during the 2020-2021 school year would experience a decrease in the chance they would be attending college the next year by 13.5%, which is quite extreme. Especially when considering the baseline next year's enrollment rate of 0.235 once again, this decrease of 13.5% would result in a whopping 57.44% decrease in the enrollment rate for the demographic if a county’s school was entirely closed during the school year, assuming 85% significance is sufficient.

The analysis of each individual variable is important to understanding the meaning of the coefficients but many counties during this 2020-2021 school didn’t follow a certain learning model throughout the entire year. Thus it is imperative to understand what it means when a county would change between various learning models and how that would affect next year's enrollment rate. Firstly it is useful to understand whether or not the combination of the three learning models is actually significantly different than the baseline in-person learning structure.

This can be accomplished using a simple F-test for significance through stata between the three variables as shown below.

```
. test lastYearVirtual lastYearHybrid lastYearClose

 ( 1)  lastYearVirtual = 0
 ( 2)  lastYearHybrid = 0
 ( 3)  lastYearClose = 0

       F(  3,223723) =    5.14
            Prob > F =    0.0015
```

This three-variable F-test calculates an F value of 5.14 which it defines is statistically significantly down to the 99.8% level. Thus, the data reveals that it is extremely likely that the combinations of these learning models have an effect on short-term college enrollment different from in-person instruction. Thus when looking at counties that change learning models throughout the year, it is assumed these estimated next year enrollment rates are in fact significantly caused by these differing models. To visualize this multivariable effect better, learning model data from Alameda County, CA will be used again. Above it was shown that this county was 90.96% virtual, 7.52% hybrid, 0.95% closed, and 0.84% open during the 2020-2021 school year. Thus plugging these percentages into the dummy variables, the regression can determine the learning models' effect on next year's college enrollment. This is seen through solution, .9096(0.0263) + 0.0752(0.0222) + 0.0095(-0.135) = 0.0239 + 0.0017 – 0.0013 = 0.0243. Thus the model states that the combination of all the learning models used in Alameda County's schools increased the chance an 18-34-year-old high school graduate in the county was enrolled in college in the 2021-2022 school year by 2.43%. This multivariate model is important because most counties in America went through huge changes in the learning models they used throughout the year.

Possible causes for these effects seen by varying learning models are hard to decipher. There likely is some correlation between the K-12 school closures observed in this study and nearby college closures. For example, if a county's schools are primarily virtual during the observed school year, colleges nearby that county would likely have a higher chance of being virtual during the next school year compared to counties that had in-person K-12. The colleges that are virtual in that area would likely drive enrollment rates as they are either able to accept more students online or students are more incentivized to enroll due to easier access to education. This would apply well to the model's findings, as areas with hybrid or virtual schools did see a statistically significant increase in their enrollment rate in the next year. If this possible hypothesis is true also for closed schools, one would expect a county where schools were primarily closed to have nearby colleges closed during the next school year. This would naturally lead to negative effects on enrollment for the next year which is indeed seen within the Stata analysis. Another possible reason for these changes would be the greater amount of free time allotted to areas where schools were not in-person. This assumption comes from the idea that non-in-person instruction would allow many to have more free time by cutting away commuting time, allowing multitasking during online courses, etc. One could assume that students attending high schools that were primarily online or hybrid would possibly have more time to apply to college, and thus enrollment rates the next year would increase. This makes logical sense when compared to the increase in next year enrollment from areas with virtual and hybrid schools. However, the possible negative effect that closed schools might have had on the enrollment rate wouldn't be explained by this deduction. In this study, by using K-12 school closure data on the enrollment rate of college, the short-run effects of the learning models can be measured extremely well. However, this top-down approach to this project makes deductions over why exactly different school learning models are affecting short-term enrollment rates difficult.

The long-term model, as identified through the event study fixed effects regression was computed through Stata and is seen below.

```
Fixed-effects (within) regression               Number of obs     =    223,924
Group variable: county                          Number of groups  =        188

R-squared:                                      Obs per group:
     Within  = 0.0024                                         min =        127
     Between = 0.0338                                         avg =    1,191.1
     Overall = 0.0031                                         max =     20,427

                                                F(40, 223696)     =      13.63
corr(u_i, Xb) = -0.0526                         Prob > F          =     0.0000
```

dumEnroll	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
yl8virtual	.020593	.0197709	1.04	0.298	-.0181575	.0593434
yl7virtual	.0259457	.0197769	1.31	0.190	-.0128165	.064708
yl6virtual	.0204576	.0199909	1.02	0.306	-.018724	.0596393
yl5virtual	.0319675	.0208929	1.53	0.126	-.0089821	.0729171
yl4virtual	.0725358	.0204204	3.55	0.000	.0325124	.1125593
yl3virtual	.0227306	.0212136	1.07	0.284	-.0188474	.0643087
yl2virtual	.0179621	.0213334	0.84	0.400	-.0238508	.059775
ylea0virtual	.0292687	.0120046	2.44	0.015	.0057399	.0527974
ylea1virtual	.0543856	.0120495	4.51	0.000	.0307688	.0780024
ylea2virtual	.0486087	.0136624	3.56	0.000	.0218307	.0753867
yl8hybrid	.0139266	.0239495	0.58	0.561	-.0330137	.060867
yl7hybrid	.05346	.0238972	2.24	0.025	.0066221	.1002979
yl6hybrid	.0915082	.0238336	3.84	0.000	.0447949	.1382216
yl5hybrid	.0858394	.024652	3.48	0.000	.0375221	.1341568
yl4hybrid	.0522596	.0242085	2.16	0.031	.0048116	.0997076
yl3hybrid	.0853399	.0250341	3.41	0.001	.0362737	.1344061
yl2hybrid	.0129987	.025153	0.52	0.605	-.0363005	.062298
ylea0hybrid	.0515418	.0139864	3.69	0.000	.0241288	.0789548
ylea1hybrid	.0664852	.0140174	4.74	0.000	.0390115	.093959
ylea2hybrid	.048378	.0159771	3.03	0.002	.0170633	.0796926
yl8close	-.5366298	.2181515	-2.46	0.014	-.9642012	-.1090584
yl7close	-.6341616	.2185002	-2.90	0.004	-1.062416	-.2059067
yl6close	-.6467448	.2251146	-2.87	0.004	-1.087964	-.2055259
yl5close	-.5616727	.237489	-2.37	0.018	-1.027145	-.0962003
yl4close	-.8429028	.2339427	-3.60	0.000	-1.301424	-.3843811
yl3close	-.398484	.2508569	-1.59	0.112	-.8901571	.0931892
yl2close	-.4210895	.2539019	-1.66	0.097	-.9187308	.0765518
ylea0close	-.3245062	.1262821	-2.57	0.010	-.572016	-.0769965
ylea1close	-.5646197	.1296277	-4.36	0.000	-.8186867	-.3105526
ylea2close	-.8636275	.1511303	-5.71	0.000	-1.159839	-.5674159

The regression gives quite interesting results. Again, the learning models are found to have a long-term effect on the enrollment rate in the demographic if all school years coefficients for the learning models from 2012-2019 are statistically equal to the baseline year of 2020-2021 and if all the coefficients for the learning models in the school years between 2020-2023 are statistically different from the baseline year. Firstly, when analyzing whether there is a statistically significant long-term effect of virtual schooling on the college enrollment rate, the assumptions posed above must coincide with the data found for virtual schooling. In terms of the data in the graph, it must be true that yl8virtual = yl7virtual = yl6virtual = yl5virtual = yl4virtual = yl3virtual = yl2virtual = 0 and ylea0virtual ≠ 0 & ylea1virtual ≠ 0 & ylea2virtual ≠ 0. Just visualizing the data above, it seems that it might be possible that these assumptions hold for counties that used virtual schooling during 2020-2021. This looks possible due to the decently small t-values for the interaction terms before the target year. The t-values for these interaction terms for the school years from 2012-2019 were 1.04, 1.31, 1.02, 1.53, 3.55, 1.07, and 0.84 in order. Excluding the outlier of the 2016-2017 school year, these t-values are small enough to claim there is a possibility of 13-40% or higher that each one of them could have the same effect as the baseline year. This obviously isn't statistically significant enough to claim that each interaction variable is individually different from the baseline, but the assumption above is only true if all values of the pre-event coefficients are equal to 0 at some confidence level. To test this multivariable hypothesis, an F-test will be used to see whether or not yl8virtual = yl7virtual = yl6virtual = yl5virtual = yl4virtual = yl3virtual = yl2virtual = 0. This test is displayed below.

```
. test yl8virtual yl7virtual yl6virtual yl5virtual yl4virtual yl3virtual yl2virtual

 ( 1)  yl8virtual = 0
 ( 2)  yl7virtual = 0
 ( 3)  yl6virtual = 0
 ( 4)  yl5virtual = 0
 ( 5)  yl4virtual = 0
 ( 6)  yl3virtual = 0
 ( 7)  yl2virtual = 0

       F(  7,223696) =    1.91
            Prob > F =    0.0638
```

This F-test states that the probability that all of these coefficients are equal to each other and the baseline coefficient of 0, is 6.38%. This means that this condition for the long-term effect of virtual schooling holding is extremely significantly unlikely. This means the analysis of the level of the coefficients for these interaction terms is not necessary, as they wouldn't reveal any key data due to the significance condition failing. The secondary condition, ylea0virtual ≠ 0 & ylea1virtual ≠ 0 & ylea2virtual ≠ 0 although does, in fact, hold and is shown in the F-test of these lead variables below.

```
. test ylea0virtual ylea1virtual ylea2virtual

 ( 1)  ylea0virtual = 0
 ( 2)  ylea1virtual = 0
 ( 3)  ylea2virtual = 0

       F(  3,223696) =    7.66
            Prob > F =    0.0000
```

As can be seen in the visualization, the coefficients between the interactions of school learning models and after-event years were indeed different from the baseline. Despite this finding, the long-term effect that virtual schooling in 2020-2021 had on long-term changes in the enrollment rate controlling for previous years' trends is null due to the failed first condition.

Similar results are found when looking at the changes in enrollment rates for the demographic due to counties being hybrid during the target year. For there to be long-term effects on hybrid schooling on the chance someone attends university, yl8hybrid = yl7hybrid yl6hybrid = yl5hybrid = yl4hybrid = yl3hybrid = yl2hybrid = 0 and ylea0virtual ≠ 0 & ylea1virtual ≠ 0 & ylea2virtual ≠ 0 must hold. As visualized in the regression data, one could see how most of the t-values for the school learning model and year interaction terms have quite high values. These values being 0.58, 2.24, 3.84, 2.48, 2.16, 3.41, 0.52, 3.69 reveal that most of these coefficients are significantly different from the baseline on their own. These coefficients are tested in Stata to see whether or not they all are statistically the same as the baseline, which is seen below.

```
. test yl8hybrid yl7hybrid yl6hybrid yl5hybrid yl4hybrid yl3hybrid yl2hybrid

 ( 1)  yl8hybrid = 0
 ( 2)  yl7hybrid = 0
 ( 3)  yl6hybrid = 0
 ( 4)  yl5hybrid = 0
 ( 5)  yl4hybrid = 0
 ( 6)  yl3hybrid = 0
 ( 7)  yl2hybrid = 0

       F(  7,223696) =    4.37
            Prob > F =    0.0001
```

This F-test reveals that these variables are, in reality, significantly different from the baseline, down to the 99.9% level. This completely fails the assumption needed to identify whether the hybrid schooling model affected college enrollment in the long term. Also, although not shown, Stata ran the F-test to test whether or not the after-event coefficients were statistically significant. This test found that these values were quite significantly different from the baseline year, meaning this secondary condition once again held. The coefficient levels for hybrid schooling are not impactful to the analysis due to the long-term significance conditions failing. These results for the hybrid model are quite similar to that of the virtual model which would be expected. As explained above in the short-term model, due to these learning models being quite similar in nature, the results between the two should also be quite similar. Thus it makes sense that in the long term, it is found that the effects of hybrid and virtual schooling on college enrollment for the demographic of interest are null. This means that both hybrid and virtual schooling were observed and calculated to not affect enrollment in the long term.

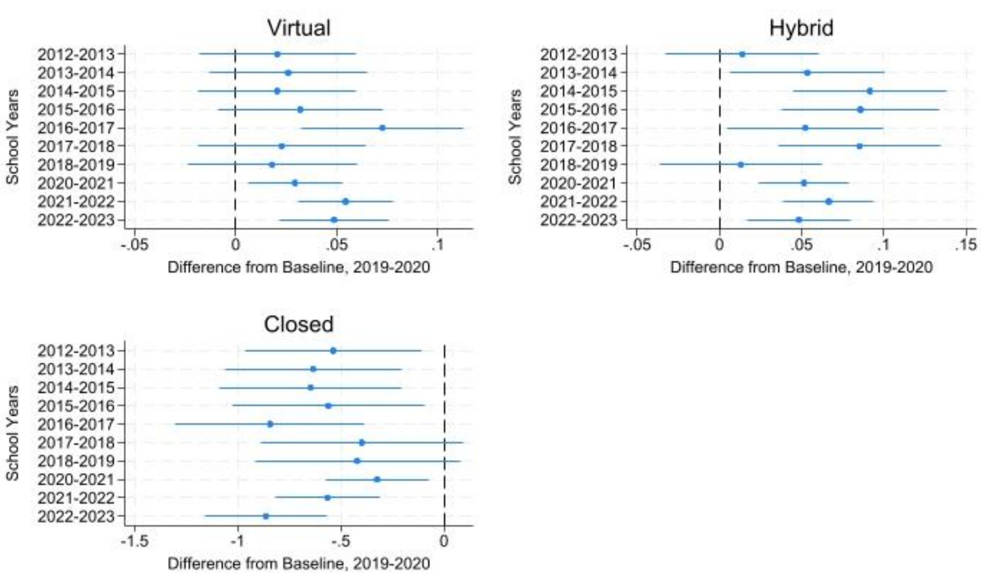
The last interesting factor to study would be the effect of closed schools on college enrollment in the long term. To analyze the significance of this effect it is important to make sure both significance conditions hold. For closed schools this condition is, yl8close = yl7close = yl6close = yl5close = yl4close = yl3close = yl2close = 0 and ylea0close ≠ 0 & ylea1close ≠ 0 & ylea2close ≠ 0. The extremely negative t-values represent similar findings as for hybrid and virtual schools. The t-values for the pre-event interaction terms are -2.46, -2.90, -2.87, -2.37, -3.60, -1.59, -1.66. With the exclusion of the last two school years, 2017-2018 and 2018-2019, the individual impact that closed schools have on the enrollment rate for 18-34-year-old high school graduates each year separately is calculated to be extremely high for years before the baseline school year 2019-2020. And this result is further solidified through the findings of a multivariate F-test run through Stata, shown below.

```
. test yl8close yl7close yl6close yl5close yl4close yl3close yl2close

 ( 1)  yl8close = 0
 ( 2)  yl7close = 0
 ( 3)  yl6close = 0
 ( 4)  yl5close = 0
 ( 5)  yl4close = 0
 ( 6)  yl3close = 0
 ( 7)  yl2close = 0

       F(  7,223696) =    3.37
            Prob > F =    0.0013
```

With a calculated F-value of 0.0013, this analysis suggests that the combination of pre-event years' interaction effect between the percentage of the school year that schools were closed in 2020-2021 and the year of study was quite significantly different from the baseline interaction effect found during the base year. This means that closed schools would also have no observed long-term effect on the percentage chance of enrollment for the focused demographic. Again, an

F-test was run for the years following the event year and found extremely similar results to the other two school teaching styles. This F-test found that in the years after the baseline school year of 2019-2020, the interaction between closed schools in COVID-19 and the years was extremely significantly different from the baseline. This was deducted through the extremely large F value of 12.50. Despite this significant value, the failure to meet the first event study significance condition means that the long-term effect of closed schools during the 2020-2021 school year on the enrollment rate to university would be nonexistent.

The event study's regression recognizes the lack of an existence in the long-term effects of changes in school learning models. Through its calculations, it finds that school learning models' effect on the changes in the percentage chance of enrollment for individuals of the key demographic were not significant in the long run due to the existence of these trends in enrollment for the areas of study. For example, a county with schools that were virtual or hybrid during the 2020-2021 school year had higher enrollment rates before and after this period when compared to areas where in-person instruction was the main model during this year. This can be best visualized using the coefficient difference model shown below.



This visualization shows the effects that these trends have on the long-term effects of enrollment. As seen in each of these graphs, the difference between the baseline school year of 2019-2020 and the interaction terms' coefficients were quite significant. One can notice that these differences between the baseline and the pre-event and post-event data are largely the same. This means that although there was an observed difference between these learning models, the long-term findings suggest that these effects existed before the event happened as well. This signifies that the difference in school learning models used during the 2020-2021 school year did not affect the chances of enrollment for 18-34-year-old high school graduates in the long term.

Limitations

The major source of limitations comes in the availability and consistency of the data found to run this study. As explained in the data-building portion of the piece, the data was found from two main datasets and combined with a third to get the data to cohesively combine. A main issue that came from these extensive data mergers was the availability of data that would fit the experiment's hard restrictions. For a county to be considered for this analysis it must fit the proper specifications to be able to get a balanced data set to regress from. To balance this data, all counties included in the data must have had COVID-19 learning model data from 2020-2021 and also must have college enrollment data from 2013-2023. These are quite hefty restrictions considering that these metrics come from two different data sets and may not efficiently merge. After keeping only the counties that fit these conditions, only 188 counties remained. However, because enrollment data was found at the individual level, there will still be over 200,000 observations. But even with these large amounts of observations, 188 counties is still a fairly small share of all the counties that exist within the United States. This limitation makes it hard to know whether the findings in this study would necessarily apply to other counties across America or just the counties found in the study.

Conclusion

The COVID-19 Pandemic and the school closures it caused led to the creation of a huge rift in education for many K-12 students in America, and by analyzing how the drastic shift in school learning models affected America, one could understand the possible positive and negative effects that each model had on the population. This study dives deep into the effects of school closures by looking at how these changes in learning models could have affected college and university enrollment in the short and long term. The piece, by using a short-term fixed effects model, found that there was a large and significant increase in the percentage chance some 18-34-year-old high school graduate was enrolled in university the next year. This same model found a possible negative effect that closing schools had on university enrollment for the demographic of focus in the next year. However, by using a long-run event study model, this paper found that the change in school learning models had no long-term effects on college enrollment for the demographic-fitting individuals of that county.

References

Aucejo, E. M., French, J., & Zafar, B. (2023). Estimating students' valuation for college experiences. *Journal of Public Economics*, *224*, 104926. doi:10.1016/j.jpubeco.2023.104926

COVID-19 School Data Hub. 2023. "All District Learning Model Data". *Data Resources* (Version 2/28/23). Accessed at https://www.covidschooldatahub.com/data-resources Fall 2023.

COVID-19 School Data Hub. 2023. "All School Learning Model Data". *Data Resources* (Version 3/8/23). Accessed at https://www.covidschooldatahub.com/data-resources Fall 2023.

Bulman, G., & Fairlie, R. (10 2022). The Impact of COVID-19 on Community College Enrollment and Student Success: Evidence from California Administrative Data. *Education Finance and Policy*, *17*(4), 745–764. doi:10.1162/edfp_a_00384

Fall 2020 Undergraduate Enrollment Down 4.4%; Graduate Enrollment Up 2.9%. (2020, November 12). Retrieved Spring 2023

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. IPUMS CPS: Version 11.0. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D030.V11.0

Thomas, C. L., & Allen, K. (2021). Investigating the Influence of COVID-Related Worry on University Enrollment Intentions: An Application of the Reasoned Action Model. *Journal of College Student Retention : Research, Theory & Practice*, 152102512110148–. https://doi.org/10.1177/15210251211014812

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Spring 2021, Fall Enrollment component. See Digest of Education Statistics 2022, table 303.50

U.S. Department of Education, National Center for Education Statistics, Education Demographic and Geographic Estimates, 2023, "School Locations & Geoassignments". *Data Resources* (Version 2022-2023). Accessed at https://nces.ed.gov/programs/edge/Geographic/SchoolLocations Fall 2023.