National Longitudinal School Database (NLSD)
Data Description

Jamie M. Carroll, Douglas N. Harris, Anjana Nair, and
Emilia Nordgren

November 28, 2023

The dataset created here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C180025 to The Administrators of the Tulane Educational Fund. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

For questions, please email nlsd.reach.info@gmail.com

Table of Contents

Recommended Citation: Jamie M. Carroll, Douglas N. Harris, Anjana Nair, and Emilia Nordgren (2023). *National Longitudinal School Database: Data Description*. New Orleans, LA: Tulane University, National Center for Research on Education Access and Choice.

Overview

The National Longitudinal School Database (NLSD) comprises three files, making up a near-census of all schools and districts in the United States from school years 1990-91 to 2019-20. The three files are the Public School File, Private School File, and District File. As evident by the titles, the first two files report data at the school-level for both public and private schools. The District File reports data at the public school district-level. We have set up these files so that they can be easily merged together in a variety of ways.

This dataset is unique in that it allows researchers to examine various aspects of school choice across traditional public schools, charter schools, magnet schools, and private schools. These data have been used to examine changes in and effects of charter schools over time (Chen & Harris, 2022) as well as trends and predictors of school closures (Harris & Martinez-Pabon, 2023). Below are some of the key characteristics of this dataset:

- The NLSD merges together more than a dozen public data files, harmonizing variable names and coding procedures to present the first comprehensive universe of traditional public, charter, and private schools.
- The NLSD includes extensively cleaned private school data, which are historically difficult to standardize and compile. The private school data also include the release of the PSS-Universe files, NCES's full list of private schools. These data have not been previously publicly available.
- The NLSD includes a unique school identifier that allows for tracking schools over time, even if there are changes in the school's governance, authority, or type.
- The NLSD integrates proprietary data to improve upon existing public data sources to provide more accurate and complete data, especially with regards to schools' charter and closure status. For example, we use data from the National Alliance for Public Charter Schools (NAPCS) to refine charter school data reported by the Common Core of Data (CCD).
- The NLSD provides much more accurate and complete data about school closures, across all sectors, by creating new indicators for school closures and takeovers.
- The NLSD provides researchers with a way to construct a universe of all traditional public, charter, and private schools that exist within the geographic boundaries of traditional public school districts.[1]

This Codebook provides documentation on the sources and methods used to create the first release of the NLSD. It is organized into three sections, corresponding to the three NSLD files (Public, Private, District). Each of the three data files are in a "long" format, such that each row

[1] The School District File contains data at both the LEA and geographic public school district level, depending on the data source (this is described more in the District File Codebook).

observation provides data for a given school (or school district) in a given school year (1990-1991 through 2019-2020). The accompanying Appendix spreadsheet describes the variables in each dataset in more detail, including information on years available, number of observations, ranges, blanks and missing data, and data source.

In future releases, we expect to augment the NLSD with additional data beyond the 2019-20 school year and from other sources relevant to its users. Potential developments include integrating data from the Office of Civil Rights, incorporating state-level charter and school takeover policies, and exploring methodological refinements to the existing data.

How to Combine NLSD Files

Together, the two school-level NLSD files (Public School File and Private School File) attempt to present the evolving landscape of the nation's public and private schools over the last three decades. The NLSD School District File provides additional information about how these schools are organized and defined into Local Education Agencies (LEAs). In the data, there are two main definitions of a school district:

1. School districts that are defined by their geographic boundaries, such that any school physically located within a district is assigned to that district. This is analogous to how traditional public school districts have been defined, but our dataset allows researchers to locate charter schools and private schools in these districts.
2. School districts that are defined by their administrative and financial reporting obligations (LEA), which are less sensitive to the physical location of the school.

This distinction is crucial to interpreting data across the three NLSD files, and between datasets within each file. Every publicly-funded school and school district is assigned a Local Education Agency ID (LEA ID) by their State Education Agency (SEA) that identifies the district that a school is legally accountable to. In many cases with traditional public schools, this assigned LEA ID is perfectly aligned with the school's geographic district, and therefore there is no distinction between the two types of IDs. Private schools, however, are independently financed and operated, and therefore not assigned an LEA ID or local governing authority. Finally, charter schools are often assigned a unique LEA ID that is neither indicative of the geographic district it belongs to, nor shared with other schools in the proximity. In that case, an LEA is defined as a single charter school or charter management organization (CMO).

Government agencies determine funding allocation and program administration through the assigned LEA ID, rather than the geographic district ID. However, LEA IDs are subject to change if there are administrative or operational changes to the district, or if the district splits or merges with another district. Given that charter schools are often their own LEA, any significant change to a charter's operational status, size, or location could trigger a change in LEA ID. In contrast, geographic district boundaries have remained relatively stable over time, and those changes are reported yearly by SEAs to the National Center for Education Statistics (NCES) and the U.S. Census Bureau. The relevant ID depends on the type of analysis being conducted.

In the NLSD Public School File, we have included a school's administrative LEA ID, *leaid,* and its geographically assigned district, *geodistid_1.* The NLSD Private School File identifies schools based on the geographic district they would be assigned to if they were part of the public school system, using *geodistid_1*. In the NLSD School District File, data is organized by *leaid*. To translate between the two types of districts, we recommend the following:

I. If the object of research is tied to the administrative aspects of a publicly funded school or district, such as school finance or test scores, then merge the District File and Public School File using *leaid* by year to uniquely identify the data.

II. If the research question only necessitates the use of geographic districts, then the District File is not necessary. Instead, merge the Public and Private School Files and aggregate up to the geographic district level using *geodistid_1* as the unique identifier.

III. If researchers are curious about how LEAs are organized within geographic districts, or would like information only available in the district file (i.e., finance information) at the geographic district level, they can merge the Public School File to the District File using *leaid* and year and use the *geodistid_1* indicator to connect LEAs and geographic districts.

Public School File

I. Introduction

The National Longitudinal School Database (NLSD) is an annual near-census of all schools and districts in the country. The Public School File includes demographic, school inputs, school type, school characteristics, student performance, and outcome data from school years 1990-91 to 2019-20 on traditional public schools, charter schools, and magnet schools. It integrates data from the National Center for Education Statistics (NCES), including the Common Core of Data (CCD), data on student achievement from the Stanford Education Data Archive (SEDA), and other school-level data. The observations are unique by school and school year.

The Public School File can be combined and analyzed with the Private School File and the School District File. The Public and Private School Files can be appended together and analyzed at both the school-level and geographic district-level. The Public School File can also be merged with the School District File by administrative district.

Section II of this Public School File codebook describes each data source and its use. Section III explains the data cleaning and the process of combining each dataset to create the NLSD Public School File. Section IV provides a detailed description of each variable created for the NLSD Public School File. The Appendix spreadsheet includes tables detailing the remaining variables in the dataset.

II. Data Sources

i. Common Core of Data (CCD)

The CCD is released by NCES. It is "a national statistical program that collects and compiles administrative data from state education agencies covering the universe of all public elementary and secondary schools and school districts in the United States."[2] It contains school-level data on school location, demographics, enrollment, federal lunch programs, teachers employed, status, and classifications, among other data going back to school year 1986-1987. The CCD files are organized by year and topic.

[2] (2021). *CCD Overview.* National Center for Education Statistics. Retrieved September 19, 2023 from https://nces.ed.gov/ccd/online_documentation.asp

The CCD is the starting point for the NLSD Public School File creation. It provides the list of schools included in the NLSD, and we keep all relevant data regarding school location and characteristics.

ii. ED*Facts*

ED*Facts* is a Department of Education initiative that collects and reports student outcome data. "ED*Facts* centralizes data provided by the state education agencies (SEAs) at the SEA, local education agency (LEA), and school levels, and provides the Department with the ability to easily analyze and report the data."[3] ED*Facts* publishes multiple datasets, and only their Assessment Proficiency datasets are included in the NLSD. They are organized by year and subject, covering schools years 2008-09 to 2018-19 and subjects math and Reading Language Arts (RLA). The data are further broken up by grade (ranging from grade 3 to high school) and subgroup (e.g. gender, race). Note that ED*Facts* is not publishing 2019-20 data because reporting requirements were suspended due to COVID-19.

Because of privacy protections, certain data need to be suppressed because of small student subgroups. The suppressed data are either completely taken out of the dataset or reported as a range. The NLSD Public School File leaves the data as reported by ED*Facts*. Table 1, from ED*Facts*' 2018-19 Data Documentation, provides details on the ranges ED*Facts'* reports based on the population size (if the population is less than 6 students, the data is not reported as a range and is instead completely suppressed):

[3] (2020). *State Assessments in Reading/Language Arts and Mathematics School Year 2018-19 EDFacts Data Documentation (Version 1.0).* U.S. Department of Education. Retrieved October 5, 2023 from https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html

Table 1

Number of Students Reported in the Cell	**Ranges Used for Reporting the Percent Proficient and Percent Participation for that Group**[4]
6-15	<50%, ≥50%
16-30	≤20%, 21-39%, 40-59%, 60-79% ≥80%
31-60	≤10%, 11-19%, 20-29%, 30-39%, 40-49%, 50-59%, 60-69%, 70-79%, 80-89%, ≥90%
61-300	≤5%, 6-9%, 10-14%, 15-19%, 20-24%, 24-29%, 30-34%, 35-39%, 40-44%, 45-49%, 50-54%, 55-59%, 60-64%, 65-69%, 70-74%, 75-79%, 80-84%, 85-89%, 90-94%, ≥95%
More than 300	≤1%, 2%, 3%, . . . , 98%, ≥99%

ED*Facts* provides information on outcomes that are not in the CCD, so we have included it in the NLSD. However, because states conduct their own assessments, the ED*Facts* data are not necessarily comparable across states. As stated in the ED*Facts* documentation, "both the content on the tests and achievement standards students must meet to be considered 'proficient' vary widely across states. Specific proficiency rates for schools in different states should not be considered comparable."[5] The Stanford Education Data Archive (SEDA) (see below) standardizes these data to be comparable across the country and is also included in the NLSD Public School File.

iii. Stanford Education Data Archive (SEDA)

"SEDA includes a range of detailed data on educational conditions, contexts, and outcomes in schools, school districts, counties, commuting zones, and metropolitan statistical areas across the United States."[6] SEDA uses the National Assessment of Educational Progress (NAEP) to standardize achievement nationally, so that student scores can be compared across states. The school-level files are organized by the type of metric used to calculate the outcome data: cohort standardized (CS) and grade cohort standardized (GCS). SEDA also reports data on school demographics, enrollment, and composition. Although some of these data are similar to those

[4] In the data, "greater than" is coded as "GT," "greater than or equal to" is coded as "GE," "less than" is coded as "LT," and "less than or equal to" is coded as "LT."

[5] Ibid.

[6] Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2021). *Stanford Education Data Archive: Technical Documentation (Version 4.1)*. Stanford University. Retrieved September 19, 2023 from http://purl.stanford.edu/db586ns4974

provided by the CCD, there are additional indicators constructed by SEDA that may be of use to researchers.

While some of the SEDA data overlap with the CCD, the outcome data do not. They are matched to each school from the CCD in the process of creating the NLSD Public School File. The most recent SEDA version at the time of the NLSD release (version 4.1) is used in the NLSD. Specifically, SEDA provides data on average academic achievement using standardized test scores administered in 3rd through 8th grade in math and Reading and Language Arts (RLA). SEDA relies on ED*Facts* data which, at the time of the SEDA 4.1 release, contained data from 2008-09 to 2017-18. SEDA differs from ED*Facts* in that it can be used to compare schools across different states. States report their test score data to ED*Facts*, but states have the freedom to choose their own tests and benchmarks for proficiency. For more information on SEDA's standardization methods, see their "Technical Documentation (Version 4.1)."

iv. National Adequate Yearly Progress and Identification Database (NAYPI)

NAYPI was created by the American Institutes for Research (AIR) for two U.S. Department of Education studies. "The NAYPI database contains detailed information on whether each school met each of up to 37 standard AYP targets including reading proficiency, math proficiency, reading test participation, and math test participation."[7] The datafiles each correspond to a different school year.

The data are no longer available to download online. When downloaded for the NLSD, the data were available only for the school years 2003-04 to 2005-06. Variables regarding the AYP targets for each school are included in the NLSD Public School File.

v. Education Demographic and Geographic Estimates Program (EDGE) National Center for Education Statistics (NCES)

NCES annually releases shape files containing the geographic boundaries of each school district, using data collected and updated by the Census Bureau. As stated by NCES, "The U.S. has more than 13,000 geographically defined public school districts. These include districts that are administratively and fiscally independent of any other government, as well as public school

[7] *National AYP and Identification Database*. American Institutes for Research. Retrieved September 19, 2023 from https://www.air.org/project/national-ayp-and-identification-database

systems that lack sufficient autonomy to be counted as separate governments and are classified as a dependent agency of some other government—a county, municipality, township, or state."[8]

We match each shape file to the corresponding school year to assign each school in the data to a geographic district. We describe this process more in depth in Section IV.

vi. Census

U.S. Census Bureau data were downloaded through Social Explorer for the years 1990, 2000, 2010, and 2020.[9] Data are aggregated by block group in the Public and Private School Files. For another aggregation, see the School District File for the district-level data.

In 2000, the U.S. Census Bureau began using the American Community Survey (ACS) as its primary method of collecting social, economic, housing, population, and demographic data. Therefore, the ACS 5-year estimates were downloaded for Census years 2010 and 2020, while the decennial Census was used for years 1990 and 2000.

For the block group level, data were extracted for the years 2010 and 2020 from the ACS 2010-2014 and ACS 2016-2020 5-year estimates, respectively. Additional information about the ACS, including design, collection, production, and data release are available through their technical documentation.[10]

vii. National Alliance for Public Charter Schools (NAPCS)

NAPCS developed a database of traditional public schools and charter schools across the country using data from the CCD and state data files from 2005-06 through 2018-19. They update charter school status using their detailed records of charter schools gathered through their work with state partners, authorizers, and state education agencies.[11]

While we do not include the NAPCS data as standalone files or variables, the NLSD Public School File uses the NAPCS data to update and confirm our indicators of charter status, described more in depth in Section IV.

[8] Geverdt, D. (2019). *Education Demographic and Geographic Estimates Program (EDGE): Composite School District Boundaries File Documentation, 2018* (NCES 2017-035). National Center for Education Statistics. Retrieved September 19, 2023 from http://nces.ed.gov/pubsearch

[9] (2023). *Social Explorer.* Retrieved September 19, 2023 from https://www.socialexplorer.com/explore-maps

[10] The ACS Technical Documentation can be found at the following link: https://www.census.gov/programs-surveys/acs/technical-documentation.html

[11] White, Jamison. (2019). *Modified Count Report*. National Alliance for Public Charter Schools.

III. Database Creation Methodology

i. Clean CCD Files and Compile with Other Data Sources

The creation of the NLSD Public School File started with the CCD files, using data from school years 1989-90 to 2019-20. Although the NLSD Public School File begins at year 1990-91, data from 1989-90 was needed in order to track schools that closed in 1990 (see Section IV for more detail on the creation of the NLSD school closure variable). Because the CCD files are organized by year and can vary slightly year-to-year, each file was cleaned for consistency and then combined into one dataset. The Appendix spreadsheet provides the list of CCD variables included in the database. After cleaning and combining each file, the following groups of schools were dropped:

- Schools in U.S. territories.
- Schools housed in correctional institutions, detention centers, or hospitals (i.e. reportable programs, schools with "correctional", "hospital", "administrative services" or "central office" in the school or LEA name or "detention center", "juvenile shelter", "JCC", "hospital", "clinic" or "medical" in the school name). These schools are often dropped by researchers because their student populations are institutionalized and, therefore, impacted by school policies in different ways than schools with non-institutionalized populations. The nature of these schools makes them less comparable to the others in the dataset.
- Schools reported as opening in the future, only in the years they are listed as such.
- Schools reported as closed that were previously reported as future but never opened. This only eliminates the observations in the years they are listed as closed.
- Schools reported as closed that were previously reported as inactive but never opened. This only eliminates the observations in the years they are listed as closed.
- Schools reported as reopened that close or disappear immediately after being reported as reopened. This only eliminates the observations in the years they are listed as reopened.
- Schools that were already reported as closed in the years immediately before. This only eliminates the observations from the second year they are listed as closed.[12]
- School with a blank or not defined operational status.
- Schools listed under orphanages, central offices, or administrative services (i.e. schools with "central office," "ctrl office," "administrative services," or "admin svc," in the school or LEA name or "orphan" in the school name).
- Schools that are duplicates in terms of their NLSD ID (see variable description in section IV) and school year, or schools that do not have an NLSD ID.

[12] Harris, D. N., Martinez-Pabon, V. (2023). Appen*dix B of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*. *Education Finance and Policy*; doi: https://doi.org/10.1162/edfp_a_00386

After the CCD data were cleaned and compiled, the data from the remaining data sources were merged onto it by NCES school ID and school year. Since the SEDA data reflect an average over the 2008-09 to 2017-18 school years, it is just matched to the NLSD Public School File by school ID. School observations in the remaining data sources that were not merged onto the cleaned CCD data were dropped from the dataset.

For the Census data, we merged indicators by block group level for 1990, 2000, 2010, and 2020 to each NLSD ID. To locate the block group of each school, we used the 2020 block group shape files available from the Census and connected schools to their corresponding block group by latitude and longitude.[13] Census variable names and labels were changed such that each variable can be identified by the geographic level (block group) and year of the Census dataset to which it belongs.

ii. Other General Cleaning Steps

We took several steps to standardize variable values and names across years. For example, variable names were sometimes updated in the source data over the years (e.g. CCD's school ID changed from *SCHNO* to *schid*). Variable names in the NLSD reflect the most recent year of the source data.

For the variables *lcity* and *location*, if the variable was blank, but the next year's value was not blank, then the current year's blank was replaced with the next year's value. *Lcity*, *charter*, *chartauth*, and *sch_name_stn* were also changed to the next year's value if the previous year and the next year had the same value and they differed from the current year's value.

The variable *lzip* also had a number of blank values, particularly for the years 1998-2000. A similar approach as described in the paragraph above was taken to fill in these blank *lzip* observations. First, blanks were filled in using the last previous available zip code for that school if the next available zip code was the same as the previous, for a maximum of three years of blanks in a row. Second, if the blank zip codes appeared in the first three years of the data, then the first available zip code was filled in for those first three years (this second step applied to 951 observations). Last, if the blank zip codes appeared in the last three years of the data, then the last available zip code was filled in for those last three years (this applied to 92 observations). Note that *lzip*, the school location zip code, is a CCD variable that wasn't introduced by the CCD until 1998. Prior to that, CCD only included *mzip*, the school mailing address. The NLSD includes only *lzip*, however *lzip* is replaced with *mzip* for the first years of the data before it was reported.

[13] Census TIGER/Line Shapefiles are available at the following link: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2020.html#list-tab-790442341

Additionally, variables for closed schools (according to variable *close_preferred*, described more in depth in Section IV) were all replaced with blanks in the first year they appear as closed except for the following variables: *ncessch, ncessch_st, nlsdsch, flag_id, sy_fall, sy_spring, sy_text, schid, st_schid, sch_name, leaid, lea_name, lcity, fipst, lstate, cnty, nmcnty, lat, lon, flag_geo, lstreet, lzip, charter, charter2, status, close_preferred, takeover_preferred, takeover_alternate*. These variables are filled in for closed schools to make tracking of school closures across years, locations, and school types easier. For example, a school that closed in 2000-01 would show *close_preferred* equal to "1" and all other variables, except those just listed, as blank for the year 2000-01. In 2001-02 and all following years, that school is no longer in the database.

iii. Coding of Blanks and Missing Data

The blank and missing codes can change over time and vary between variables in the source data. While many data sources are used in the NLSD Public School File, only the CCD includes more detail and documentation on why values are blank. The labels the CCD uses were standardized in the NLSD into the values in Table 2.

Table 2

Variable Type	NLSD Value	Meaning
Numeric	.m	Missing
Numeric	.n	Not applicable
Numeric	.r	Not reported
Numeric	.s	Suppressed
Numeric	.c	Calculation error[14]
Text	M	Missing
Text	NA	Not applicable
Text	NR	Not reported
Text	S	Suppressed

[14] This missing code is not taken from the data sources. It has been added for variables created for the NSLD if they cannot be calculated with the given data. For example, a calculated percentage where the denominator's value is blank is coded as a calculation error (".c").

The CCD did not start making the distinction between “missing” and “not reported” until 2016. They define “missing” values to be those reported by the State Education Agency (SEA) as missing. “Not reported” is assigned to values if nothing was reported by the SEA.[15]

The remaining blanks reflect variables that were not available for given years or missing data that had no missing code in the source documentation.

IV. Variables

In this section, we describe the variables created for the NLSD Public School File. The remaining variables that are mainly unchanged from the original data sources are described in the Appendix spreadsheet to this codebook including the variable name, source, number of observations, and other descriptive information.

i. NLSD Identifier

To allow for tracking a school longitudinally, the NLSD Public School File includes the NLSD identifier (*nlsdsch*), which is the first NCES identifier (*ncessch*) assigned to a school and does not change even if *ncessch* does. The *ncessch* can change over time if the school changed the Local Education Agency (LEA) to which it was affiliated. Since the school NCES identifier is a twelve-digit code that combines information from the state (2 digits), the LEA (5 digits), and the school (5 digits), any change in LEA implies a new NCES identifier.[16] Other cases that lead to changes in the NCES identifier include when: (a) the grade span of the school changes by more than three grades (not including pre-kindergarten/kindergarten as grades), (b) the school’s physical location changed and the attendance area changed significantly, (c) two schools of about the same size, or with different grade spans, merge.[17] We determined whether a school with a new NCES identifier was the same school from the prior school year through a few steps. First, we isolated the 5-digit school identifier from the state and LEA identifiers to have a school identifier that is unaffected by changes in the LEA. By matching this identifier with the school’s state and city, we were able to track schools through changes in their NCES identifier. Second, we used the CCD indicator for school status, which indicates whether a school has changed its LEA affiliation (available starting in 1998-99). If the school was ever indicated to have a change

[15] The CCD documentation can be found at the following link: https://nces.ed.gov/ccd/online_documentation.asp

[16] White, J. (2019). *2019 NCES ID report*. National Alliance for Public Charter Schools.

[17] Harris, D. N., Martinez-Pabon, V. (2023). *Appendix of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*, pp. 3-4. Education Finance and Policy; doi: https://doi.org/10.1162/edfp_a_00386

in its LEA affiliation, the NLSD identifier is the NCES identifier for the first year that school was in operation. The variable *flag_id* was created for the NLSD and indicates if there was a change in the school's NCES identifier.

The NCES identifier is still helpful in the data as it is an ID used widely in other data sources, including SEDA and Ed*Facts*, for example. While the NCES identifier allows for connecting schools across datasets, the NLSD identifier was constructed to better track schools longitudinally within datasets in cases where the NCES identifier may change.

ii. Charter School Indicator

The NLSD Public School File contains an indicator for charter school status, *charter*, based on the CCD indicator (*charter*) with several additional steps taken to confirm charter status, following the approach of Harris & Martinez-Pabon (2023). As stated in their paper, "first, since the CCD does not include the charter indicator from 1991 to 1998, we imputed the charter status backwards in time from when the charter indicators first became available and using the years when state charter laws begin and the school name (e.g., whether the word "charter" or "academy" is used). Second, we checked for coding errors during the entire period of data and assumed that a school reported as a charter (TPS) in one year but TPS (charter) in both adjacent years is a coding error and we recode the middle year with the adjacent year value."[18] We made additional updates to this variable based on information in the NAPCS database (available for school years 2005-06 to 2017-18). If the CCD charter status was missing or not indicated as a charter, we updated it to the NAPCS indicator.

iii. School Closure Indicator

The NLSD Public School File school closure indicator (*close_preferred*) was created using the CCD operational status indicator (*status*), along with additional steps described below. The CCD currently has eight operational statuses they use to classify schools: Open, Closed, New, Added, Changed, Inactive, Future, and Reopened. However, not all of these categories are available for every year. The first four categories have been available since the 1990-91 school year, the "changed" status was added in 1998-99, "inactive" and "future" were added in 2002-03, and "reopened" was added in 2005-06.[19] *Status* is a categorical variable with the following values:

[18] Harris, D. N., Martinez-Pabon, V. (2023). *Appendix A3 of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*. Education Finance and Policy; doi: https://doi.org/10.1162/edfp_a_00386

[19] See Appendix A2 of Harris and Martinez-Pabon (2023) for more detail.

- 1 - open
- 2 - closed
- 3 - new
- 4 - added
- 5 - changed
- 6 - inactive
- 7 - future
- 8 - reopened

To create a consistent measure of school closure across all years of the NLSD Public School File, we used the following steps:

- Define as open those schools reported as open, new, added or reopened. This is because all schools reported within these categories are operational.
- Define as closed those schools reported as closed or inactive, and drop schools from the NLSD listed as closed (under the new aggrupation) that repeat that operational status in consecutive years. This only eliminated the observations from the second year they are listed as closed.
- In years where schools are reported as closed, and in previous and subsequent years they are reported as open, we recoded the school as open.
- For the years from 1991 to 1994, when schools disappeared from one year to the next, we assumed they were closed even if there was no closure flag. This process allowed us to identify all schools that closed in this period, which were removed from the CCD data files.
- For schools identified as closed, we redefined the year of the closure as the first year without reported enrollment or enrollment equal to zero.
- Define as closed those schools that change their location (address) in the same year as a change in city, change in name, or a decline in enrollment of more than 25 percent relative to the previous year.[20] We found that even after using standardized variables, many address changes seem to occur because of typos or slight changes in the way addresses are reported. By requiring other changes (e.g., in city) at the same time, we minimized the possibility that typos lead to false closure indications.[21]

One limitation to our definition of school closure is that it focuses on situations where buildings cease to function as schools, following Harris and Martinez-Pabon (2023). In some cases, school personnel and leadership may have moved to another location, but the building where they were

[20] Our methodology here differs slightly from the followed Harris & Martinez-Pabon (2023) methodology. While we flagged schools as having a potential location change if their enrollment *declined* by more than 25%, Harris & Martinez-Pabon flagged schools if their enrollment *changed* (in either direction) by more than 25%.

[21] Harris, D. N., Martinez-Pabon, V. (2023). *Appendix B1 of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*. Education Finance and Policy; doi: https://doi.org/10.1162/edfp_a_00386

housed in was boarded up, torn down, used for other educational purposes, or used for non-educational purposes. These types of changes in building are challenging to distinguish from other types of closures, which is why we pair a change in address with other school features, as described above. However, we may still be over-identifying some closures.

iv. School Takeover Indicator

The NLSD Public School File has two school takeover variables – *takeover_preferred* and *takeover_alternate* – that indicate school restructuring, or where significant and involuntary changes occur in personnel, management, and/or governance. *Takeover_alternate* uses fewer steps to identify restructured schools, while *takeover_preferred* incorporates the same steps as takeover_alternate plus additional rules.

The alternate takeover variable was created as follows:

- Define as restructured those schools that report a change in their charter status.
- Define as restructured those charter schools that report both a change in LEA and a change in name.[22]

The preferred takeover variable is created as follows:

- Define as restructured those schools that report a change in their charter status.
- Define as restructured those schools that report they were reconstituted (starting in 2011).
- Define as restructured those charter schools that report both a change in LEA and a change in name.
- Define as restructured those charter schools that report a change in the charter authorizer (starting in 2014).
- Re-define as restructured instead of closed cases of schools reported as closed when a new school is reported in the exact location during the same or following school year.[23]

Our main purpose for these indicators is to reflect changes in the decision-making authority in schools, following Harris and Martinez-Pabon (2023). These changes include restarts, reconstitutions, conversions, turnarounds, and takeovers. We do not include school transformations, which generally include only changes in curriculum and programs, or changes in school principals, which generally leave the school otherwise unchanged.

[22] Ibid.

[23] Ibid.

v. Geographic District Indicator

For most public schools, their LEA Identifier (*leaid*) supplied by the CCD corresponds to the geographic area from which they draw their student enrollment. However, schools may have a different LEA Identifier if they are not governed by their local school district or have other political designations for their funding status. For example, schools associated with military bases, schools undergoing state takeover, and charter schools may have an LEA Identifier that does not correspond to a geographic district area. We used the NCES EDGE shape files to locate NLSD schools in their correct geographic district area.

NCES combines the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) school district layers with district-level demographic data from the American Community Survey (ACS) to produce the EDGE shape files. The school district layers are updated regularly to reflect any changes in school district boundaries, names, LEA identifiers, grade ranges, and school district levels that have been reported to the Census Bureau.

The following table indicates the EDGE shape files that were chosen to geolocate schools within districts for each year of the NLSD. This matching process was guided by the NCES documentation.[24]

Table 3

NCES EDGE Shapefile Year	NLSD School Years
2021-2022	2018-2022
2017-2018	2014-2018
2013-2014	2010-2014
2009-2010	2006-2010
2006-2007	2004-2006
2004-2005	2002-2004
2002-2003	2001-2002
1999-2000	1998-2000
1997-1998	1996-1998
1996-1997	1991-1996
1994-1995	1990-1991

We began by geolocating all NLSD schools to produce latitude and longitude coordinates. Then, we used Stata code *geoinpoly* to determine which geographic district the NLSD school is located in. In some cases, the school was geolocated to more than one geographic district (up to 4 in

[24] Geverdt, Douglas E. (2019). *Education Demographic and Geographic Estimates Program (EDGE): Composite School District Boundaries File Documentation, 2018* (NCES 2017-035). National Center for Education Statistics. Retrieved September 19, 2023 from http://nces.ed.gov/pubsearch

1990, up to 3 from 1991-2002, and up to 2 in the remaining years). This geolocating process accounts for the creation of four district ID variables – *geodistid_1, geodistid_2, geodistid_3,* and *geodistid_4* – as well as their associated name variables – *geodistnm_1, geodistnm_2, geodistnm_3*, and *geodistnm_4*. Before 1998, NCES EDGE shape files did not include the district names that corresponded with district IDs. If pre-1998 observations were geolocated to a district ID that matched the administrative LEA ID, the LEA name was assigned as the district name.

We determined which geographic district is the preferred district (*geodistid_1*) using the following criteria:

- If any of the geographic district identifiers matched the school's LEA Identifier, we used their LEA Identifier as their preferred geographic district. This was the case for 92% of school by year records.
- For remaining schools matched to multiple geographic districts, if one of the districts was described as "consolidated" or "unified,"[25] we used that district as the preferred geographic district.
- For remaining schools matched to multiple geographic districts, we matched the district name to the school level. For example, for a high school matched to one district described as a "high school district" and another district described as an "elementary school district," we made the preferred geographic district the high school district.
- For schools not assigned a preferred geographic district using the steps above, we made the first district the school was matched to the preferred geographic district.
- We kept any remaining districts in the data file (*geodistid_2*, *geodistid_3*, and *geodistid_4*).
- Given such few observations for *geodistnm_4*, we dropped it from the final dataset.

As a general cleaning step, we recoded any district names labeled as "school district not defined" to "M" for "missing."

vi. Latitude and Longitude

Latitude and longitude data from CCD were used when available. If a school's latitude or longitude was blank, then it was filled in, when possible, using the Census geocoding tool.[26]

[25] The geographic district data files do not include the name of the geographic district from 1990 through 1998.

[26] The United States Census Bureau's Geocoder "Find Geographies" tool was used to obtain geographic coordinates for school addresses. The tool can be found at this link: https://geocoding.geo.census.gov/geocoder/

vii. Locale Description

We created a combined indicator of the type of location the school is in (*locale_combined*), based off of the variables *ulocal* and *locale* from CCD. Before 2006, CCD used *locale* to indicate 7 (or 8, depending on the year) categories of school location. From 2006 to 2013, the CCD added *ulocal*, which has 12 categories, and from 2014 to 2019 returned to using locale with the same 12 categories as *ulocal*. We combined this information into one 7-category variable (*locale_combined*) to align changes in the definition of school location across the years.

Locale_combined is a categorical variable with the following values:

- 1 - large city
- 2 - mid-size city
- 3 - urban fringe of large city & large suburb
- 4 - urban fringe of midsize city & midsize suburb
- 5 - large town & fringe
- 6 - small town & remote/distinct
- 7 - rural

viii. Title I Indicators

The Title I indicator, *title_one*, combines and standardizes the CCD indicators of Title I status, which change over the years (*titlei, titlei_status_text, titlei_text*) to flag whether a school was Title I-eligible in a given year based on their LEA. A school is eligible if 1) the percentage of children from low-income families in the school is at least as high as the percentage of children from low-income families served by the LEA as a whole or 2) 35% or more of the children in the school are from low-income families.

The schoolwide Title I indicator, *stitle_one*, combines and standardizes the CCD variables *stitli*, *stitlei*, and *titlei_status_text* to flag schoolwide Title I-eligible schools. These are schools that are Title I-eligible and at least 40% of the children in the school are from low-income families.

We keep the original CCD indicators in the dataset (*titlei, titlei_status, titlei_status_text, titlei_text, stitli, stitlei*), as they provide more detail about the different kinds of eligibility and how that has changed over the years.

ix. Proportion of Students by Race/Ethnicity

There are seven possible race/ethnicity categories defined in the CCD: Black, Hispanic, white, Native American, Asian, Native Hawaiian / Pacific Islander, and two or more races. All are included in the NLSD Public School File, however Native Hawaiian / Pacific Islander and two or more races are only available starting in the 2008-09 school year (the other categories are available from 1990). CCD reports the number of students of each race/ethnicity by school. For the NLSD, we included the proportion of students by race/ethnicity, rather than the count, using the variable *member* in the denominator. For example, the percent of students who are Black was calculated as the number of Black students divided by *member*. These calculations produced the variables *p_bl, p_hp, p_wh, p_am, p_as, p_pc,* and *p_tr*.

We added an eighth category for other students who don't fall into one of the seven listed above. The number of students of another race was calculated as *member* minus the sum of the number of students identified by the other race/ethnicity categories. Note that there are seven race/ethnicity categories, but not all are available for every year. So the other number of students can represent a different group of race/ethnicities depending on the year of the data. This number of students divided by *member* produced *p_ot*. If *member* minus the sum of the students in the identified race/ethnicity categories equaled a negative number, then *p_ot* was marked as a calculation error (".c"). If any of the available race/ethnicity student counts were blank or missing, then *p_ot* was also marked as a calculation error.

x. National School Lunch Program (NSLP)

The CCD provides two kinds of data on eligibility for the NSLP. First, the CCD includes the number of students receiving free or reduced price lunch. Similar to the proportion of students by race/ethnicity, we calculated the proportion of students eligible for free or reduced price lunch (*frpl*) as the number of students eligible divided by *member*. Starting in 2017, in addition to the number of students who are qualified for free or reduced price lunch, the CCD also reports the number of students who received "direct certification" of their eligibility through their participation in government programs. We used the former in our calculations to align with prior year data reporting, since the students who qualify under direct certification are a subgroup of the total student eligible for free or reduced-price lunch.

The Healthy, Hunger-Free Kids Act was amended in 2010, which changed the way students can apply and qualify for free and reduced-price lunch and, therefore, impacted the number of students marked as eligible in the CCD data. In 2010, the Community Eligibility Option (CEO) was added, which made entire schools NSLP eligible if at least 40% of their students were identified as eligible through direct certification in a qualifying year. Schools qualifying under the CEO do not collect student-level applications, impacting, therefore, the student-level data. Since all students are eligible for receiving a free or reduced-price lunch under the CEO, schools may report all students in the school as eligible, regardless of their individual status.[27] However, starting in 2014-15, Ed*Facts* (who collects the data from schools) began advising schools that qualify under the CEO to "report current headcounts of free and reduced price students, when possible. If the data are not available due to schools implementing the NSLP provisions, estimate the count of students by multiplying current year membership by the percentage of eligible students in the most recent year for which the school collected that information."[28]

The second CCD indicator, *nslpstatus*, provides more information on the school-level participation in the NSLP, starting in the 2013-2014 school year. CCD changed the name of this variable to *nslpstatus_code* for 2014-15 and 2015-16, then *nslp_status* moving forward. This indicator provides information about how students' eligibility for free or reduced price lunch is determined. To accommodate changes in data reporting over the years, we combined the CCD indicators (*nslpstatus, nslp_status,* and *nslpstatus_code*) to create one indicator of whether the school participated in NSLP at all (*nslp*). We keep the other indicators in the dataset for researchers interested in the kind of school eligibility (Provisions 1-3 [PR1-PR3], Community Eligibility Option [CEO]).[29]

xi. Indicators for Grades Offered

There is a variable for each grade that indicates whether or not that grade is offered in the school (*off_g01, off_g02, off_g03, off_g04, off_g05, off_g06, off_g07, off_g08, off_g09, off_g10, off_g11, off_g12*). The CCD has this variable in the raw data for some years, but not all. When available, the CCD variable value was used. However, the CCD value was changed from "0" or blank to "1" if the number of students for the grade (variable *g01*, for example) was greater than zero and not blank/missing. When the CCD variable was not available, the flag was marked as "1" if the number of students in that grade was greater than zero and not missing/blank.

[27] For more detail on the Healthy, Hunger-Free Kids Act and the CEO, see *Free and Reduced-Price Lunch Eligibility Data in EDFacts: A White Paper on Current Status And Potential Changes,* available at https://eric.ed.gov/?id=ED556048

[28] (2015). *C033 – Free and Reduced Price Lunch File Specifications – V11.1* (SY 2014-15). U.S. Department of Education, Washington, DC: ED*Facts*. Retrieved December 1, 2023 from http://www.ed.gov/edfacts.

[29] For more information on the CCD's coding of these variables, see their *File 129 – CCD School File Specifications*, available by year at this link: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html

xii. Charter Law

The variable *charter_law* was created to track the states' charter legislation. The values vary only by state and represent the year that that state passed their charter law.

xiii. Standard Abbreviations

Unless otherwise noted, the NLSD follows variable values, including abbreviations, as used in the source data. Some frequently used abbreviations and their meanings are listed below.

- PK - Pre-Kindergarten
- KG - Kindergarten
- AE - Adult Education
- UG - Ungraded
- GT - Greater than
- GE - Greater than or equal to
- LT - Less than
- LE - Less than or equal to

Private School File

I. Introduction

The National Longitudinal School Database (NLSD) is a near-census of all schools and districts in the country. The Private School File includes demographic, school inputs, school type, and school characteristics data for private schools in the U.S., covering every other school year from 1991-92 to 2019-20. It expands on data from the National Center for Education Statistics (NCES) Private School Universe Survey (PSS) with information from HuffPost on school choice programs as well as manual searches to determine school status. The observations are unique by school and school year.

The Private School File can be combined with the Public School File. They can be appended together and analyzed at both the school-level and geographic district-level. The School District File is organized by administrative district, which private schools are naturally not assigned to. The District File also largely includes only data that is applicable to public schools. However, it is possible to map the administrative districts in the School District File to the geographic districts in the Private School File.

Section II of this Private School File codebook describes each data source and its use. Section III explains the data cleaning and the process of combining each dataset to create the NLSD Private School File. Section IV provides a detailed description of each variable created for the NLSD. The Appendix spreadsheet includes tables detailing the remaining variables in the dataset.

II. Data Sources

i. Private School Universe Survey (PSS)

The PSS is released by NCES in order to track and report data on private elementary and secondary schools. "The target population for PSS is all schools in the United States that are not supported primarily by public funds, provide classroom instruction for one or more of grades kindergarten through 12 (or comparable ungraded levels), and have one or more teachers."[30] It contains school-level data on school location, demographics, enrollment, federal lunch programs,

[30] Broughman, S.P., Kincel, B., and Peterson, J. (2021). *Private School Universe Survey (PSS): Public-Use Data File User's Manual for School Year 2019–20* (NCES 2022-021). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved October 9, 2023 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022021.

and teachers employed, among other data. The PSS files cover every other year, going back to school year 1989-90.

Unlike the NCES Common Core of Data (CCD) used in the Public School File, the PSS is a sample, as private schools are not required to respond to the survey. For example, the 2019-20 PSS had a response rate of 74.5%.[31] Following the approach of Harris & Martinez-Pabon (2023), we chose to use just the PSS data rather than the PSS-*Universe* data obtained from the U.S. Department of Education. The PSS-Universe includes the schools that responded to the survey as well as the schools that did not. However, since the data on the non-response schools includes only their name, identifier, and address, we are not able to apply the same cleaning methods, described in Section III, as we do to the response schools. This would impact the creation of our NLSD variables, namely the closure variables. In order to maintain the accuracy of the closure variables and keep them comparable to those in the Public School File, we chose to focus on the PSS data. However, because the PSS-Universe data can still be helpful to researchers, we release it as a separate file. The PSS-Universe data is described more below.

The PSS is the starting point for the NLSD Private School File creation. It provides the list of schools included in the NLSD. All relevant data regarding school location and characteristics are included in the NSLD.

ii. Private School Universe Files (PSS-Universe)

The PSS-Universe files are produced by NCES but are not publicly available on their website. They were sent to the NLSD research team, and NCES has given us permission to release certain data from these files. The files list the names and addresses of the private schools that NCES sends the PSS to each year, regardless of whether those schools responded to the survey or not. The files are organized by year, from 1991-92 to 2015-16, and have been combined for release.

As mentioned above, we release the PSS-Universe data as its own file, as the limited information in the PSS-Universe prevents the same types of cleaning steps taken in creating the Private School File. The variables in the PSS-Universe data file are largely unchanged from the raw data, although some general cleaning steps were taken and additional variables were created. These are described more in Sections III and IV.

[31] Ibid. See Table 3 for more detail.

iii. HuffPost School Choice List

In 2017, HuffPost published a list of private schools in the US that were participating in a private school choice scholarship program at the time of the research. “To do this, we looked for the most up to date list we could find on either a state’s department of education or department of revenue website. If a state did not maintain a public list of schools participating in these programs, we reached out to representatives from the state. If the representative did not have a list, we looked on the website of all the individual scholarship granting organizations in the state. If the scholarship granting organization did not post a list, we reached out to them for help. If the scholarship granting organizations did not respond, we were not able to include its schools in our database.”[32]

HuffPost’s list is used to flag schools in the NLSD Private School File that participate in a choice program. However, note that the list was created according to the situation in 2017 and may not be applicable to other years.

iv. Education Demographic and Geographic Estimates Program (EDGE) National Center for Education Statistics (NCES)

The same NCES shape files are used in the Private School File as are used in the Public School File. NCES annually releases shape files containing the geographic boundaries of each school district, using data collected and updated by the Census Bureau. As stated by NCES, “The U.S. has more than 13,000 geographically defined public school districts. These include districts that are administratively and fiscally independent of any other government, as well as public school systems that lack sufficient autonomy to be counted as separate governments and are classified as a dependent agency of some other government—a county, municipality, township, or state.”[33]

We match each shape file to the corresponding school year to assign each school in the data to a geographic district. We describe this process more in depth in Section IV.

[32] See the “Methodology” tab of HuffPost’s database. Retrieved October 9, 2023 from http://big.assets.huffingtonpost.com/HuffPostPrivateSchoolChoice121317.xlsx.

[33] Geverdt, D. (2019). *Education Demographic and Geographic Estimates Program (EDGE): Composite School District Boundaries File Documentation, 2018* (NCES 2017-035). National Center for Education Statistics. Retrieved September 19, 2023 from http://nces.ed.gov/pubsearch.

v. Google Maps Application Programming Interface (API)

Google API is a tool that can return information on places, including schools' operational status, based off of their geographic location.[34] It is used in the NLSD Private School File to track private schools that do not appear in the PSS from one year to the next. Because the PSS data includes only those schools which responded to the survey for that year, and responses are not required, a school may not appear in the next year's data for two reasons: 1) the school closed, or 2) the school chose not to respond to the PSS. Google API is used to verify whether or not the schools that disappear from the PSS data have closed in those years.

vi. Census

The same Census data are used in the Private School File as are used in the Public School File. U.S. Census Bureau data were downloaded through Social Explorer for the years 1990, 2000, 2010, and 2020.[35] Data are aggregated by block group.

In 2000, the U.S. Census Bureau began using the American Community Survey (ACS) as its primary method of collecting social, economic, housing, population, and demographic data. Therefore, the ACS 5-year estimates were downloaded for Census years 2010 and 2020, while the decennial Census was used for years 1990 and 2000.

For the block group level, data were extracted for the years 2010 and 2020 from the ACS 2010-2014 and ACS 2016-2020 5-year estimates, respectively. Additional information about the American Community Survey, including design, collection, production, and data release are available through their technical documentation.[36]

III. Database Creation Methodology

i. Clean PSS files and Compile with Other Data Sources

The creation of the NLSD Private School File started with the PSS data. The NLSD Private School File starts at school year 1991-92, so PSS data from 1989-90 is downloaded in order to

[34] For more information, see the Google API overview, found at this link: https://developers.google.com/maps/documentation/places/web-service/overview

[35] (2023). *Social Explorer.* Retrieved September 19, 2023 from https://www.socialexplorer.com/explore-maps

[36] The ACS Technical Documentation can be found at the following link: https://www.census.gov/programs-surveys/acs/technical-documentation.html

track schools that closed in 1990 and 1991 (since the PSS is collected every other year). Each year's file was cleaned for consistency and combined into one dataset. The PSS variables were renamed to match the corresponding variable name in the Public School File, making it easier to analyze the two files together. The PSS flag variables were also recoded such that a value of "1" indicates a "yes" or positive flag, and "0" indicates a "no" or negative flag. The PSS default flag values are "1" and "2." After cleaning and combining each file, the following groups of schools were dropped:

- Schools offering no grade higher than kindergarten.
- Schools that are not and have never been designated as "regular" (i.e. alternative, technical, early childhood, montessori, and special education schools).
- Schools that have a yearly average of less than 20 students enrolled.

These schools were dropped largely because of data availability and consistency issues. A main feature of the NLSD is tracking schools and school closures over time. The data for the dropped schools make this feature difficult, as information about these schools is generally not reported as reliably or consistently as those that were included.

After the PSS files were cleaned and combined, the data from the remaining data sources were merged onto the NLSD Private School File by school ID and school year. Observations from the remaining data sources that were not merged onto the NLSD were dropped from the database. The HuffPost data is only available for one year (2017), so it was only merged onto the NLSD by school ID. The Google API data was used to aid in the creation of the closure variable. This process is described more in Section IV.

For the census data, we merged indicators by block group level for 1990, 2000, 2010, and 2020 to each NLSD ID. To locate schools within block groups, we used the 2020 block group shape files available from the Census and connected schools to their corresponding block group by latitude and longitude.[37] Census variable names and labels were changed such that each variable can be identified by the geographic level and year of the Census dataset to which it belongs.

The PSS-Universe, released separately from the Private School File, remains largely unchanged from the raw data received from NCES and approved for release. Variable names were changed to be consistent with the corresponding variables in the Private School File. Nine-digit zip codes were also cleaned and separated into their five- and four-digit parts. Additionally, depending on the year, a second set of address variables is given in the raw data for each school. The second address variable was replaced with the first if its raw value was "SAME." No schools or observations were dropped or added to the PSS-Universe.

[37] Census TIGER/Line Shapefiles are available at the following link:
https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2020.html#list-tab-790442341

ii. Other General Cleaning Steps

General cleaning steps were taken to standardize variables across years and to keep them consistent and comparable to their corresponding version in the NLSD Public School File. Additionally, steps were taken to fill in blank variable observations. For *sch_name* and *lstreet*, if one year's value was blank but either of the following two years' values were not, then the blank was filled in with the future value. A blank value was also filled in with the previous year's value if the previous year's value was not blank for *sch_name* and *lstreet*, as well as *fipst, cnty, lat, lon, flag_geo, lzip, lcity, lstate, nmcnty, geodistid_1, geodistid_2, geodistid_3, geodistnm_1, geodistnm_2,* and *geodistnm_3*.

IV. Variables

i. NLSD Identifier

The NLSD identifier (*nlsdsch*) is the same as the NCES school identifier (*ncessch_st*) in both the Private School File and PSS-Universe. It was created in the Public School File so that schools can be tracked longitudinally, since public schools' NCES identifiers can change over time (e.g., because of changes in LEA affiliation or grade span offered). However, the private school identifiers do not have the same issue. *Nlsdsch* was added to the Private School File and PSS-Universe to make combining the school-level files easier.

ii. School Closure Indicator

There are two school closure indicators created for the NLSD Private School File: *close_preferred* and *close_alternate*. For both variables, manual searches were considered for determining whether or not a school was closed. The variables differ where manual searches couldn't verify school status.

Because schools are not required to respond to the PSS, a school that appears in the PSS data one year but not the next is not necessarily closed. Therefore, additional cleaning was needed in order to separate the schools that actually closed from the schools that chose not to respond to the PSS. We looked up schools, using their names and addresses, in the Google Maps Application Programming Interface (API) tool to verify their status. However, since data on schools' street addresses wasn't added to the PSS until the 2007-08 school year, this verification process could only be done starting in 2007-08. Additionally, even though the 2019-20 PSS data is included in the NLSD, it has not yet been run through our verification process using Google

API. Therefore, the following process applies only to the 2007-08 to 2017-18 data (what we will refer to as the PSS-sample):

- Define as open those schools that always reported information to the PSS-sample.
- Identify as possibly closed those schools that appeared in the PSS-sample and never reappeared in later years.
- Define as closed the schools in the group of possibly closed that were reported as closed through (programmed and manual) online searches:
    - For the programmed online search, we used the Google Maps Application Programming Interface (API) to request the operational status (operational, temporarily/permanently closed) and web address for each school in the group of possibly closed.
    - If the Google Maps' operational status of the school was temporarily/permanently closed, we coded the school as closed.
    - If the Google Maps' operational status of the school was open or missing, we did additional manual online searches. In these steps, we received help from undergraduate coders who checked the accuracy of the linked website and searched for relevant information regarding the operational status of the schools using their name and address. Some information in this case included news, State Department of Education websites, and GreatSchools.org, among others.
- For the preferred measure, define as closed those schools without a verified operational status through online searches.
- For the alternate measure, define as open those schools without verified operational status through online searches.[38]

Close_alternate is available for every year of the NLSD. *Close_preferred* is left blank for the years outside of the PSS-sample in which the cleaning process could not or has not been completed.

One limitation to our definition of school closure is that it focuses on situations where buildings cease to function as schools, following Harris and Martinez-Pabon (2023). In some cases, school personnel and leadership may have moved to another location, but the building where they were housed in was boarded up, torn down, used for other educational purposes, or used for non-educational purposes. These types of changes in building are challenging to distinguish from other types of closures, thus we may still be over-identifying some closures.

[38] Harris, D. N., Martinez-Pabon, V. (2023). *Appendix B2 of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*. Education Finance and Policy; doi: https://doi.org/10.1162/edfp_a_00386

iii. Proportion of Students by Race/Ethnicity

The proportion of students by race/ethnicity are variables already calculated in the PSS. There are seven race/ethnicity categories: Black, Hispanic, white, Native American, Asian, Native Hawaiian / Pacific Islander, and two or more races. However, note that Native Hawaiian / Pacific Islander and two or more races were not added as categories in the PSS until 2009-10. The variable *p_ot*, the proportion of students who are another race/ethnicity, was created for the NLSD Private School File. It represents the remaining percentage of students such that the sum of *p_ot* and all the other proportions add up to one. If the sum of the defined race/ethnicity proportions was greater than one, then *p_ot* was coded as ".c," a calculation error.

iv. School Choice Program Indicator

The school choice program indicator (*choice*) was created using data released by HuffPost (see Section II.ii). HuffPost in 2017 released a list of private schools in the country that, at the time, participated in school choice scholarship programs. A school in the NLSD Private School File was flagged using the choice variable if it appears in the HuffPost list, and it was flagged for every year it appears in the data. However, note that the list was created in 2017, and participation in choice programs can change over time.

v. Geographic District Indicator

The geographic district indicators for the Private School File were constructed using the same methodology as the Public School File. To enable connection between private schools and the public schools in the same area, we used the NCES EDGE shape files to locate NLSD private schools in the geographic public school district area.

NCES combines the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) school district layers with district-level demographic data from the American Community Survey (ACS) to produce the EDGE shape files. The school district layers are updated regularly to reflect any changes in school district boundaries, names, LEA identifiers, grade ranges, and school district levels that have been reported to the Census Bureau.

The following table indicates the EDGE shape files that were chosen to geolocate private schools within geographic districts for each year of the NLSD. This matching process was guided by the NCES documentation.[39]

Table 1

NCES EDGE Shapefile Year	NLSD School Years
2021-2022	2019
2017-2018	2015, 2017
2013-2014	2011, 2013
2009-2010	2007, 2009
2006-2007	2005
2004-2005	2003
2002-2003	2001
1999-2000	1999
1997-1998	1997
1996-1997	1991, 1993, 1995

We began by geolocating all NLSD private schools to produce latitude and longitude coordinates. Then, we used Stata code *geoinpoly* to determine which geographic district the NLSD private school is located in. This geolocating process accounts for the creation of three district ID variables – *geodistid_1, geodistid_2,* and *geodistid_3* – as well as their associated name variables – *geodistnm_1, geodistnm_2,* and *geodistnm_3*. Before 1998, NCES EDGE shape files did not include the district names that corresponded with district IDs.

We determined which geographic district was the preferred district (*geodistid_1*) using the following criteria:

- If one of the matched geographic districts was described as "consolidated" or "unified,"[40] we used that district as the preferred geographic district.
- For remaining schools matched to multiple geographic districts, we matched the district name to the school level. For example, for a high school matched to one district described as a "high school district" and another district described as a "elementary school district," we made the preferred geographic district the high school district.
- For schools not assigned a preferred geographic district using the steps above, we made the first district the school was matched to the preferred geographic district.
- We kept any remaining districts in the data file (*geodistid_2*, and *geodistid_3*).

As a general cleaning step, we recoded any district names labeled as "school district not defined" to "M" for "missing."

[39] Geverdt, Douglas E. (2019). *Education Demographic and Geographic Estimates Program (EDGE): Composite School District Boundaries File Documentation, 2018* (NCES 2017-035). National Center for Education Statistics. Retrieved September 19, 2023 from http://nces.ed.gov/pubsearch

[40] The geographic district data files do not include the name of the geographic district from 1990 through 1998.

vi. Locale Description

We created a combined indicator of the type of location the school is in (*locale_combined*), based off the variables *ulocale* and *locale* from PSS. Through 2003, PSS used locale to indicate 7 (or 8, depending on the year) categories of school location. In 2003, the PSS added ulocale, which has 12 categories. They kept both locale and ulocale for 2003 and 2005, then dropped locale from the dataset in 2007. We combined this information into one 7-category variable (locale_combined) to align changes in the definition of school location across the years.

Locale_combined is a categorical variable with the following values:

- 1 - large city
- 2 - mid-size city
- 3 - urban fringe of large city & large suburb
- 4 - urban fringe of midsize city & midsize suburb
- 5 - large town & fringe
- 6 - small town & remote/distinct
- 7 - rural

vii. PSS-Universe Variables

There are five variables created and added to the PSS-Universe for the NLSD release: *nlsdsch, sy_fall, sy_spring, sy_text,* and *response*. *Nlsdsch* is the unique NLSD school identifier. For the PSS-Universe, it is the same as the NCES school identifer (*ncessch_st*). It is added to the PSS-Universe to be consistent with the other NLSD files, making analyses across files easier. The three school year variables – *sy_fall, sy_spring, sy_text* – reflect the school year of each individual file sent by NCES.

The last variable, *response*, is a flag for if the school responded to the PSS. It was created by merging the raw PSS-Universe with the raw PSS by school and year. If a school from the PSS-Universe appears in the PSS, then it is marked as a “responder” for that year. There is a small percentage of schools in the raw PSS data that are not listed in the PSS-Universe. These observations are mostly concentrated in school years 1991-92 and 1993-94. We did not add these schools to the PSS-Universe. However, we note this here since we would expect all schools that respond to the PSS to be listed in the PSS-Universe.

School District File

I. Introduction

The National Longitudinal School Database (NLSD) is an annual near-census of all schools and districts in the country. The School District File includes demographic, financial, and student outcome data from traditional public and charter school districts for the school years 1990-91 to 2019-20. In this file, districts are defined by their Local Education Agency IDs (LEA IDs), rather than their physical geographic boundaries, and include all of the LEAs in the NLSD Public School File.[41]

The School District File integrates data from the National Center for Education Statistics (NCES), including the Common Core of Data (CCD), data on student achievement from the Stanford Education Data Archive (SEDA), and other district-level datasets.

Section II of this School District codebook describes each data source and its use. Section III explains the data cleaning and the process of combining each dataset to create the NLSD School District File. Section IV provides a detailed description of each variable created for the NLSD School District File. The Appendix spreadsheet includes tables detailing the remaining variables in the dataset.

II. Data Sources

i. Common Core of Data (CCD) Directory Files

The CCD is released by NCES. It is "a national statistical program that collects and compiles administrative data from state education agencies covering the universe of all public elementary and secondary schools and school districts in the United States."[42] The CCD files are organized by year and topic. For the NLSD School District File, the CCD Directory, Finance, and Elementary and Secondary Information System (ELSI) were used.[43]

[41] With the exception of the SEDA data, which is presented by geographic district ID.

[42] *CCD Overview.* (2021). National Center for Education Statistics. Retrieved September 19, 2023 from https://nces.ed.gov/ccd/online_documentation.asp

[43] Note that as of the 2019-20 school year, the CCD no longer reports data on English language learners. Instead, those data are reported through the Ed*Facts* Data Express. For more information: *Ed Data Express.* U.S. Department of Education. Retrieved November 9, 2023 from https://eddataexpress.ed.gov/download

Along with the LEA universe derived from the NLSD Public School File, the CCD Directory file forms the foundation of the NLSD School District File. The Directory File contains data on district type and location, operational changes from the prior year, full-time staff, and enrollment, among other data going back to the school year 1986-87. [44]

ii. CCD Finance Files

The CCD Finance data come from the School District Finance Survey (F-33), which is administered yearly by the National Center for Education Statistics (NCES). State Education Agencies (SEAs) in all 50 states and the District of Columbia submit the F-33 to NCES, which reports the finance data for "all local education agencies (LEAs) that provide free public elementary and secondary (prekindergarten through grade 12) education in the United States."[45]

The CCD Finance Files include data on revenues by source, expenditures, debts, assets, and student membership counts, among others. National and state totals are excluded. Data are presented at the school district level in positive, whole dollar amounts, with the exception of a few variables that take on at least one negative value in the raw data.[46]

Data was downloaded for all school years available: 1991-92, and 1994-95 through 2019-20.[47]

iii. CCD Elementary/Secondary Information System (ELSI) Files

The ELSI is a data tool published by the NCES.[48] It compiles researchers' most used variables across the CCD's public and private school universes, and allows researchers to generate custom reports at the state, district, and school level. We downloaded district-level public school data on the number of diploma recipients, which was available from the 1986-87 school year to 2009-10,

[44] Data was downloaded through the Urban Institute: (2023). *Common Core of Data Directory.* Education Data Portal (Version 0.19.0), Urban Institute. Retrieved October 20, 2023 from https://educationdata.urban.org/documentation/. Made available under the ODC Attribution License.

[45]Cornman, S.Q., Ampadu, O., Hanak, K.S. (2022). *Documentation for the NCES Common Core of Data School District Finance Survey (F-33), School Year 2019–20 (Fiscal Year 2020), Provisional File Version 1a (NCES 2022-304)*. NCES, Institute of Education Sciences, U.S. Department of Education. Retrieved October 20, 2023 from https://nces.ed.gov/ccd/pdf/2022304_FY20F33_Documentation.pdf

[46] Variables with negative values in the raw data include: *debt_interest, debt_longterm_outstand_beg_fy, exp_current_student_transport, rev_fed_state_title_i, rev_local_misc, rev_state_sch_lunch,* and *rev_state_transportation.*

[47] (2023). *Common Core of Data Finance.* Education Data Portal (Version 0.19.0), Urban Institute. Retrieved October 20, 2023 from https://educationdata.urban.org/documentation/. Made available under the ODC Attribution License.

[48] (2023). *Elementary/Secondary Information System.* NCES, Institute of Education Sciences, U.S. Department of Education. Retrieved October 20, 2023 from https://nces.ed.gov/ccd/elsi/

as well as student enrollment numbers during that time period for grades 8, 9, and 10. These data are used to construct estimates of the Average Freshman Graduation Rate (NLSD variables *afgreb* and *afrgr*) which researchers can use as an alternative outcome metric to test scores and other student achievement data. See Section IV for more details.

iv. Small Area Income and Poverty Estimates (SAIPE)

The Small Area Income and Poverty Estimates (SAIPE) dataset is released annually by the U.S. Census Bureau.[49] It estimates income and poverty statistics at the school district, county, and state levels. The school district poverty estimates were constructed from the county-based estimates, federal tax information, and multi-year surveys. The data are primarily used for administering federal programs to local districts, including the distribution and management of federal funds.

SAIPE's district-level estimates are only available for school years beginning in 1995, 1997, and 1999-2019.[50] Only state and county estimates are available in 1996 and 1998. The NLSD School District File integrates SAIPE data on the sizes of the district's total population, school-aged population, and the school-aged population in poverty.

v. Stanford Education Data Archive (SEDA)

The Stanford Education Data Archive (SEDA) publishes data on academic achievement and growth for schools, geographically defined school districts, counties, commuting zones, metropolitan statistical areas, and states. SEDA uses the National Assessment of Educational Progress (NAEP) to standardize achievement nationally, so that student scores can be compared across states. The most recent SEDA version at the time of the NLSD release (version 4.1) is used in the NLSD District File. Average academic achievement is measured through standardized test scores for 3rd through 8th grade mathematics and Reading Language Arts (RLA), spanning school years 2008-09 to 2017-18.

The SEDA school-level data, which were used in the NLSD Public School File, are only available as pooled overall estimates of outcome data by metric—cohort standardized (CS), or grade cohort standardized (GCS). However, the school-district data include estimates in three

[49] *Small Area Income and Poverty Estimates Program*, U.S. Census Bureau, Retrieved October 20, 2023 from https://www.census.gov/programs-surveys/saipe.html
[50] Data was downloaded through the Urban Institute: (2023). *Small Area Income and Poverty Estimates (SAIPE).* Education Data Portal (Version 0.19.0), Urban Institute. Retrieved October 20, 2023 from https://educationdata.urban.org/documentation/. Made available under the ODC Attribution License.

pooling levels. Data files are either "long," containing estimates for each grade and year separately; "pooled by subject," where estimates are averaged across grades and years within subjects; and "pooled overall," which averages estimates across grades, years, and subjects. Data are reported for all students by demographic subgroup, with the exception of special education students. SEDA does not report these numbers at the district level, but instead reports them separately as statewide "SEDA Special Education Districts." These districts were not included in the NLSD School District File.

SEDA 4.1 also provides estimates of socioeconomic, demographic, and segregation characteristics of school districts from the CCD and ACS in a separate file. These data include raw and computed measures, such as the composite socioeconomic status measure, and are available in the form of covariate files. At the geographic district-level, there are three types of covariate files offered for the set of variables reported, differing based on the level of pooling across grades and years.

The NLSD District File uses the CS "long" data file with covariates pooled by year. Note that the SEDA variables' names are identical across the NLSD Public School and School District Files, and should be differentiated before merging the two files. Further, SEDA data is reflective of geographic district estimates and not tied to administrative LEA ID, diverging from the other datasets in the NLSD School District File. SEDA uses the 2019 Elementary and Unified School District Boundaries File published by the NCES to geolocate schools to their corresponding physical school district. For more information on SEDA's standardization methods, see their "Technical Documentation (Version 4.1)."[51]

vi. Census

U.S. Census Bureau data used in the NLSD District, Public, and Private School Files were downloaded through Social Explorer for the years 1990, 2000, 2010, and 2020.[52]

In 2000, the U.S. Census Bureau began using the American Community Survey (ACS) as its primary method of collecting social, economic, housing, population, and demographic data. In the District file, ACS 5-year estimates were downloaded for Census years 2010 and 2020 at the school district level. Data at the school district level were not available in 1990 and 2000, so we manually aggregated block group level data to the geographically-defined school districts. We describe this process more in depth in Section IV.

[51] Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2021). *Stanford Education Data Archive: Technical Documentation (Version 4.1)*. Stanford University. Retrieved September 19, 2023 from http://purl.stanford.edu/db586ns4974

[52] (2023). *Social Explorer*. Retrieved September 19, 2023 from https://www.socialexplorer.com/explore-maps

vii. ED*Facts*

As in the NLSD Public School File, data from ED*Facts*, a Department of Education initiative that collects and reports student outcome data, are included in the NLSD District File. District-level Ed*Facts* variables are identical to the school-level version, except data are reported by LEA ID.

"ED*Facts* centralizes data provided by the state education agencies (SEAs) at the SEA, local education agency (LEA), and school levels, and provides the Department with the ability to easily analyze and report the data."[53] ED*Facts* publishes multiple datasets, and only their Assessment Proficiency datasets are included in the NLSD. They are organized by year and subject, covering schools years 2008-09 to 2018-19 and subjects math and Reading Language Arts (RLA). The data are further broken up by grade (ranging from grade 3 to high school) and subgroup (e.g. gender, race). Note that ED*Facts* is not publishing 2019-20 data because reporting requirements were suspended due to COVID-19.

Because of privacy protections, certain data need to be suppressed because of small student subgroups. The suppressed data are either completely taken out of the dataset or reported as a range.

The NLSD School District File leaves the data as reported by ED*Facts*. The table below, from ED*Facts*' 2018-19 Data Documentation, provides details on the ranges ED*Facts'* reports based on the population size (if the population is less than 6 students, the data is not reported as a range and is instead completely suppressed):

[53] (2020). *State Assessments in Reading/Language Arts and Mathematics School Year 2018-19 EDFacts Data Documentation (Version 1.0).* U.S. Department of Education. Retrieved October 5, 2023 from https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html

Table 1

Number of Students Reported in the Cell	**Ranges Used for Reporting the Percent Proficient and Percent Participation for that Group**[54]
6-15	<50%, ≥50%
16-30	≤20%, 21-39%, 40-59%, 60-79% ≥80%
31-60	≤10%, 11-19%, 20-29%, 30-39%, 40-49%, 50-59%, 60-69%, 70-79%, 80-89%, ≥90%
61-300	≤5%, 6-9%, 10-14%, 15-19%, 20-24%, 24-29%, 30-34%, 35-39%, 40-44%, 45-49%, 50-54%, 55-59%, 60-64%, 65-69%, 70-74%, 75-79%, 80-84%, 85-89%, 90-94%, ≥95%
More than 300	≤1%, 2%, 3%, . . . , 98%, ≥99%

ED*Facts* provides information on outcomes that are not in the CCD. It is therefore added to the NLSD School District File for the years it is available. However, because states conduct their own assessments, the ED*Facts* data are not necessarily comparable across states. The Stanford Education Data Archive (SEDA) standardizes test score data to be comparable across the country and is also included in the NLSD.

III. Database Creation Methodology

i. Clean CCD Files and Compile with Other Data Sources

The creation of the NLSD School District File started with the NLSD Public School File, which contains data for school years 1990-91 through 2019-20.

We begin with a census of LEA IDs from the Public School File. Next, we merge in the CCD Directory, Finance, and ELSI data for the same time period by LEA ID and year. Because CCD files are organized by year and can vary slightly year-to-year, each file was cleaned for consistency and then combined into one dataset. The Appendix spreadsheet provides the list of CCD variables included in the database.

The NLSD School District File contains only the universe of LEA IDs that appear in the NLSD Public School File. Therefore, if certain school types were excluded from the Public School File,

[54] In the data, "greater than" is coded as "GT," "greater than or equal to" is coded as "GE," "less than" is coded as "LT," and "less than or equal to" is coded as "LT."

their associated LEA IDs would only appear in the District File if that school district contained other school types that meet our inclusion criteria. As noted in the school-level codebook, cleaning steps taken in the Public School File included dropping the following types of schools:

- Schools in U.S. territories.
- Schools housed in correctional institutions, detention centers, or hospitals (i.e. reportable programs, schools with “correctional”, “hospital”, “administrative services” or “central office” in the school or LEA name or “detention center”, “juvenile shelter”, “JCC”, “hospital”, “clinic” or “medical” in the school name). These schools are often dropped by researchers because their student populations are institutionalized and, therefore, impacted by school policies in different ways than schools with non-institutionalized populations. The nature of these schools makes them less comparable to the others in the dataset.
- Schools reported as opening in the future, only in the years they are listed as such.
- Schools reported as closed that were previously reported as future but never opened. This only eliminates the observations in the years they are listed as closed.
- Schools reported as closed that were previously reported as inactive but never opened. This only eliminates the observations in the years they are listed as closed.
- Schools reported as reopen that close or disappear immediately after being reported as reopen. This only eliminates the observations in the years they are listed as reopen.
- Schools that were already reported as closed in the years immediately before. This only eliminates the observations from the second year they are listed as closed.[55]
- School with a blank or not defined operational status.
- Schools listed under orphanages, central offices, or administrative services (i.e. schools with “central office,” “ctrl office,” “administrative services,” or “admin svc,” in the school or LEA name or “orphan” in the school name).
- Schools that are duplicates in terms of their NLSD ID (see variable description in section IV) and school year, or schools that do not have an NLSD ID.

After the CCD data were compiled, the data from the remaining data sources were merged onto it by LEA ID and school year, with the exception of Census data. Census variables in the NLSD School District File were renamed to reflect their geographic level– school district– and year. For this reason, Census data were merged into the NLSD using *leaid* only.

Across all data sources, observations that did not merge onto our universe of LEA IDs were dropped from the dataset. Note also that general cleaning steps were taken to standardize variable values and names across years. The NLSD reflects the most recent version of a variable

[55] Harris, D. N., Martinez-Pabon, V. (2023). Appen*dix B of Extreme Measures: A National Descriptive Analysis of Closure and Restructuring of Traditional Public, Charter, and Private Schools*. *Education Finance and Policy*; doi: https://doi.org/10.1162/edfp_a_00386

available in the source data. In other instances, such as in the Ed*Facts* data, variable labels were abbreviated to comply with Stata character limits.

ii. Coding of Blanks and Missing Data

The blank and missing codes can change over time and vary between variables. While many data sources are used in the NLSD School District File, only the CCD includes more detail and documentation on why values are blank. The labels the CCD uses were standardized in the NLSD into the values in Table 2.

Table 2

Variable Type	NLSD Value	Meaning
Numeric	.n	Not applicable
Numeric	.r	Not reported
Numeric	.s	Suppressed
Text	NA	Not applicable
Text	NR	Not reported
Text	S	Suppressed

“Not reported” is assigned to values if nothing was reported by the SEA.[56] The remaining blanks reflect variables that were not available for given years or missing data that had no missing code in the source documentation.

IV. Variables

In this section, we describe the variables created or otherwise altered for the NLSD School District File. The remaining variables that are mainly unchanged from the original data sources are described in the Appendix spreadsheet to this codebook including the variable name, source, number of observations, and other descriptive information.

[56] (2021). *CCD Overview.* National Center for Education Statistics. Retrieved September 19, 2023 from https://nces.ed.gov/ccd/online_documentation.asp

i. Locale Description

We created a combined indicator of the degree of urbanization that the school district is in (*locale_combined*), based on the CCD Directory variable *urban_centric_locale*. This variable measures a local education agency's (LEA) physical location relative to populous areas.

Before 2005, CCD used *urban_centric_locale* to describe 7 locale categories based on a school district's metro status. This variable was originally coded from 1-7, with 8 categories in some years. After 2005, the CCD updated *urban_centric_locale* to include 12 types of locale descriptions that fall into four main groups: city, suburb, town, and rural. We combined this information into one 7-category variable (*locale_combined*) to align changes in the definition of district urbanicity across all years. For more information on *urban_centric_locale,* see the CCD Directory File Documentation.[57]

Locale_combined is a categorical variable with the following values:

- 1 - large city
- 2 - mid-size city
- 3 - urban fringe of large city & large suburb
- 4 - urban fringe of midsize city & midsize suburb
- 5 - large town & fringe
- 6 - small town & remote/distinct
- 7 - rural

ii. Census Variables

Census variables were only available at the school district level in 2010 and 2020. To create the 1990 and 2000 Census variables at the school district level, we use the NCES School District Geographic Relationship Files (GRF)[58] for 2013 (the oldest GRF record published). This file connects each school district to the block group(s) included within its boundaries and the land area within the district these block groups encompass.

Using the cleaned Census block group data used in the NLSD Public School File for 1990 and 2000, we calculate district-level measures by weighting the average of each block group level measure by the land area they encompass within the district.

[57] Ibid.

[58] For documentation related to the 2013-14 school year, see the *2013 School District Geographic Reference Files: Technical Documentation* at this link: https://nces.ed.gov/programs/edge/docs/EDGE_SDGRF_2013_UserDoc.pdf

iii. ELSI Variables

The NLSD District File constructs an estimate of the Average Freshman Graduation Rate (*afgr*) for school years 1990-91 through 2009-10 using the CCD ELSI variable *afgr_ccd.* The CCD's version is only available for the years 2005-06 to 2009-10. To estimate the AFGR prior to 2005, we calculate the weighted average of student enrollment in grades 8-10 (the Average Freshman Graduation Rate enrollment base for each year, *afgreb)* and divide the number of diplomas awarded in a district per year by the AFGR enrollment base. Note that all values of *afgr* that are greater than one (and not missing) have been recoded to one.

iv. Other Categorical Variables

The NLSD School District File includes several categorical variables originating from the CCD Directory Files.

a) agency_type

This variable describes the type of agency that an LEA ID belongs to.

- 1 - regular local school district
- 2 - local school district that is a component of a supervisory union
- 3 - supervisory union
- 4 - regional education service agency
- 5 - state-operated agency
- 6 - federally-operated agency
- 7 - charter agency
- 8 - other education agency
- 9 - specialized public school district

b) boundary_change_indicator

This variable classifies changes made to a LEA's boundaries since the last time they reported to NCES.

- 1 - no change/operational
- 2 - closed
- 3 - new
- 4 - added

- 5 - significant change in geographic boundaries or instructional responsibility
- 6 - temporarily closed
- 7 - future
- 8 - reopened

c) agency_charter_indicator

This variable indicates the makeup of charter schools within an LEA.

- 1 - all charter schools
- 2 - some charter schools
- 3 - no charter schools

d) lowest_grade_offered & highest_grade_offered

The remaining two variables listed below are CCD grade level indicators, recoded from numeric values [-1, 15] to string values for the prekindergarten, kindergarten, adult education, ungraded, and grade 13 levels. These variables describe the span of grades offered by a given LEA, and are constructed using data reported by SEAs annually to the EdFacts EMAPS reporting system.

Since the 2015-16 school year, the CCD has included grade 13 and adult education data, where grade 13 is defined as, "High school students enrolled in programs to earn college credit in an extended high school environment, or career and technical education (CTE) students in a high school program continuing past grade 12."[59] The adult education grade category includes those enrolled in adult education programs, but also those students who previously dropped out of school and have re-enrolled. Despite being outside of public elementary and secondary school systems and excluded from the CCD Finance Files, this category is "now treated by the CCD as another grade level for LEAs that provide primarily adult education."[60]

The ungraded category is assigned to students taking courses that have not been assigned a particular grade level by the state. Ungraded and grade 13 counts are reported for a state only if the state indicates in its annual report that it offers these grades. Beginning in the 2015-16 school year, the CCD made adjustments to the reporting requirements to allow states that do not have official ungraded offerings to still report their ungraded student enrollment.[61]

[59] Cornman, S.Q., Ampadu, O., Hanak, K.S. (2022). *Documentation for the NCES Common Core of Data School District Finance Survey (F-33), School Year 2019–20 (Fiscal Year 2020), Provisional File Version 1a (NCES 2022-304)*. NCES, Institute of Education Sciences, U.S. Department of Education. Retrieved October 20, 2023 from https://nces.ed.gov/ccd/pdf/2022304_FY20F33_Documentation.pdf

[60] Ibid.

[61] (2021). *Non-fiscal Data Reviews and Edits.* National Center for Education Statistics. Retrieved September 19, 2023 from https://nces.ed.gov/ccd/online_documentation.asp

This variable describes the grade span offered by an LEA.

- 1 - grade 1
- 2 - grade 2
- 3 - grade 3
- 4 - grade 4
- 5 - grade 5
- 6 - grade 6
- 7 - grade 7
- 8 - grade 8
- 9 - grade 9
- 10 - grade 10
- 11 - grade 11
- 12 - grade 12
- 13 - grade 13
- PK - prekindergarten
- KG - kindergarten
- AE - adult education
- UG - ungraded

e) agency_level

This variable describes the level of instruction that characterizes a school district.

Beginning in 2017-18, two level categories were added– prekindergarten and secondary. "Secondary" describes districts that offer mainly grades 9, 10, and 11, but not grade 12. The same year, the CCD made changes to the way level categories are derived. Previously, the categories were based solely on the grade variables, *lowest_grade_offered* and *highest_grade_offered.*[62] The new approach, which relies on the full breadth of grade data reported by states, aims to recategorize the "other" category to make it more meaningful for researchers. "Other" now describes any district that offers either 1) both elementary and secondary grades, or 2) grades from each of the elementary, secondary, or high school levels. Also, LEAs that offered only grades 9, 10, 11, and/or 12, and that also offered prekindergarten

[62] These variables are named *gslo* and *gshi* in the CCD Directory File, but were renamed for the NLSD to match the Urban Institute's labeling system and documentation. For more information, see the following: (2023). *Common Core of Data Directory.* Education Data Portal (Version 0.19.0), Urban Institute. Retrieved October 20, 2023 from https://educationdata.urban.org/documentation/. Made available under the ODC Attribution License.

and kindergarten, were classified as “high” before 2016-17. Since 2017-18, these schools have been classified as either “secondary” or “high.”[63]

- 0 - prekindergarten
- 1 - primary
- 2 - middle
- 3 - high
- 4 - other
- 5 - ungraded
- 6 - adult education
- 7 - secondary

[63] For CCD documentation on these changes, see their file *Changes to CCD-assigned school and LEA levels,* available at this link: https://nces.ed.gov/ccd/reference_library.asp