Description of the DC Master School File

Chelsea Coffin
DC POLICY CENTER

October 2019

The District of Columbia collects and publishes a lot of data regarding its public schools.[1] But school-level data currently available to the public are not always ready for easy use, especially across years. Education data are siloed at different agencies, especially before the release of the 2017–18 DC School Report Card, which now links core data on school information, enrollment, student characteristics, outcomes, school environment, and geography. Still, some data are missing from the DC School Report Card (e.g., waiting list length, enrollment by ward, and program offerings).

The DC Policy Center, in collaboration with the Urban Institute, created the longitudinal DC Master School File, which compiles publicly available data for unique **DC Public Schools (DCPS) and public charter schools over time** (school years 2014–15 through 2017–18) from several sources. The file is a resource for stakeholders who would like to better understand DC's public schools over time. In addition, the file provides a preliminary inventory of which historical data may be available for research purposes.

What Data Does the Master School File Include?

To be included in the DC Master School File, data must be publicly available, aggregated at the school level, calculated consistently for both DCPS and public charter schools, and available for multiple years.[2] By topic, the following data are included:

- **School information.** Local education agency (LEA) code, LEA name, school code, school name, school year, school address, number of facilities, program offerings (e.g., International Baccalaureate program, Montessori, dual language, interscholastic sports), sector

- **Enrollment.** Lowest grade offered at the school, highest grade offered at the school, enrollment by grade, students by ward of residence,[3] in-boundary share (DCPS only), boundary participation rate (DCPS only)

- **Student characteristics.** American Indian or Alaska Native students, Asian students, black or African American students, Hispanic or Latino students of any race, Native Hawaiian or other Pacific Islander students, students identifying as two or more races, white students, at-risk students, English learners, special education students, students with level 1 special education needs, students with level 2 special education needs, students with level 3 special education needs, students with level 4 special education needs

- **Outcomes.** School Transparency and Reporting (STAR) rating, STAR score, math Partnership for Assessment of Readiness for College and Careers (PARCC) 4+, English language arts PARCC 4+, math PARCC 4+ by subgroup, English language arts PARCC 4+ by subgroup, median growth percentile (MGP) in math, MGP in English language arts, four-year graduation rate, five-year graduation rate

- **School environment.** In-seat attendance, reenrollment, expulsion rate, suspension rate, midyear entry, midyear withdrawal, truancy, waiting list length

- **Geography.** Census tract in 2010, neighborhood cluster in 2000, Advisory Neighborhood Commission in 2012, council ward in 2012, zip code in 2002, x coordinate, y coordinate, latitude, longitude

A few additional data points are not currently available publicly, though they likely exist. As available, future iterations of the Master School File could add the following information:

- **Seats offered and waiting list offers for all four school years**. For previous years, the data are missing schools that are not currently open.

- **Share of students who are at risk based on the highest level of need for 2014–15, 2015–16, and 2016–17.** The DC School Report Card data count at-risk students at their highest level of need, not the time of the audit, that is available for previous years. These numbers are currently shared only at the LEA level.

- **Share of students with special education needs (levels 1–4) for 2017–18**. The STAR report shares the percentage of students who have any special education needs but not differentiated needs.

- **Number of students on DCPS school waiting lists for 2016–17.** This dataset should be available later in 2019.

- **Some student characteristics[4] and school environment data for 2014–15: enrollment by special population, in-seat attendance, midyear entry, expulsion rate, suspension rate, and midyear withdrawal**. The data exist in individual school Equity Reports, but a flat file is not publicly available.

How Should the Master School File Be Maintained in the Future?

The Master School File should be updated annually for public use. Future versions should be updated with the historical data points mentioned above that were not available when the file was created. In addition to the currently included information, the file should incorporate metrics now available in DC's School Report Card (e.g., school leader information, teacher experience, eighth-to ninth-grade promotion, Advanced Placement and International Baccalaureate participation and performance, Classroom Assessment Scoring System results for schools with earlier grades, reenrollment, and SAT results) as these datasets grow. The file should also include a crosswalk between Office of the State Superintendent of Education (OSSE) school IDs and National Center for Education Statistics IDs so stakeholders can easily link to federally available data, such as Title I status or number of teachers.

DC could also collect and publicly report additional school-level information that other districts include in their master school datasets. Texas publishes information on teachers and finance, such as total full-time equivalent staff, total full-time equivalent teachers, average teacher salary, teachers by race or ethnicity, teacher turnover rate, and revenue per student.[5] New York City's file includes the average number of students per class.[6] California provides additional school background, such as the opening and closing year, as well as detailed school type (e.g., application, adult, alternative, or special education).[7]

Procedures and Data Sources

The DC Master School File presents key education data that are publicly available for school years 2014–15 through 2017–18. Some data sources do not include all schools if the data do not apply to certain school types (e.g., boundary participation is not relevant for public charter schools). The procedures for creating the file and descriptions of data sources are below.

PROCEDURES

1. For each year, download the annual enrollment audit from the OSSE. Use this list of schools as the universe for each year with the OSSE-assigned school code as the unique identifier.

2. For each data source below, download the raw data and create comma-separated-value files with the relevant variables that include the variable, school year, and unique school code.

3. Beginning with the enrollment audit file, merge each variable file using the school code to create complete datasets for each year.

4. Merge datasets for all years to create the draft longitudinal file.

5. Validate the data for each variable to ensure consistent representation across years.

6. Check 5 percent of data against the original sources to ensure data quality.

Data have been merged from the following key sources.[8] Data suppression rules of the original dataset were also applied to this file (fewer than 10 students or fewer than 25 students, depending on the source and year). The Urban Institute and DC Policy Center are not responsible for any errors in the underlying publicly available data.

- **OSSE enrollment audits.** Annually, the OSSE conducts an enrollment audit to determine the number of students at each public school in the District. These files contain the audited enrollment by school and grade.

- **DC School Report Card.** First published for the 2017–18 school year, the OSSE DC School Report Cards give a look into all DC public schools. It provides an overall rating and more than 150 data points to communicate how a school is doing with all its students, in addition to helpful information about the school itself. These files include DC School Report Card and STAR framework metrics at the school, LEA, and state levels for all metrics and student groups on the report card and the within the STAR framework by student group and framework for the same reporting levels. STAR ratings and scores are not available before 2017–18. At this time, two other metrics included in the STAR report (at-risk share and reenrollment share) are not available in a comparable way for DCPS and charter schools before 2017–18.

- **PARCC.** The PARCC assessment in English language arts and mathematics measures the knowledge and skills that matter most for students (e.g., writing, problem solving, and other fundamental skills that lead to confidence and success in mathematics and English language arts). This dataset contains the official assessment performance results for the PARCC and Multi-State Alternate Assessment tests in English language arts and mathematics and detailed information, showing multiple levels of results for specific groups of students, for all grades within a school, and for individual grades.

- **My School DC.** The common lottery's School Finder tool provides information on school characteristics, including program offerings.

- **DME school facilities.** The Office of the Deputy Mayor for Education's (DME's) Master Facilities Plans and supplements provide detailed facility information, including addresses.

- **DME boundary participation data.** This source provides the number of grade-appropriate students that live in each DCPS boundary, the names of schools attended by students living in each boundary, and the number of students from the boundary who attend that school in 2017–18. In-boundary share and boundary participation for DCPS schools are not available before 2016–17.

- **MAR Geocoder.** The Master Address Repository (MAR) Geocoder is a desktop application that employs DC's MAR Web Services to assign a location in the form of geographic coordinates and returns other geospatial information.

- **OSSE graduation rates.** The OSSE publishes statewide graduation rates by school for DCPS and public charter schools. Five-year graduation rates are not available before 2015–16.

- **OSSE Equity Reports.** In the years before the publication of the DC School Report Card, the OSSE published annual Equity Reports. Equity in education refers to all students receiving the same caliber of education, regardless of their race, ethnicity, economic status, special education status, or other factors. Equity Reports gave schools, families, and communities transparent and comparable equity-related information across all DC schools. For 2014–15, Equity Report data are available only in individual school Equity Reports and were not included in this file, aside from student race or ethnicity.

- **OSSE Performance Oversight.** The OSSE, as a government agency, answers questions from the DC Council annually, including how many students live in each ward by school. These data have been suppressed if fewer than 10 students lived in a ward. Data for 2014–15 were not based on the audit and have not been included, as they are not comparable with later years.

- **OSSE attendance report.** This report focuses on absenteeism and truancy, primarily among compulsory-aged students. Truancy data are not available before 2016–17.

- **DCPS and PCSB waiting lists.** DCPS and the Public Charter School Board (PCSB) post their waiting lists as of the release of lottery results.

See the accompanying companion file for additional information on the individual variables in the file and the appendix for an overview of data characteristics.

Appendix. Data Characteristics

Name in dataset	Definition	Variable type	Sources, all years
lea_code	Unique, OSSE-assigned identifier for LEAs, three or four digits	Integer	OSSE enrollment audit
lea_name	Name of the LEA	String	OSSE enrollment audit
school_code	Unique, OSSE-assigned identifier for schools, three or four digits	Integer	OSSE enrollment audit
school_name	Name of the school	String	OSSE enrollment audit
sector	Whether a school is traditional public (DCPS) or public charter (charter)	String	OSSE enrollment audit
year	School year	String	N/A
address	Primary address	String	DME School Facilities Data
number_of_facilities	Number of facilities	Numeric	DME School Facilities Data
lowest_grade_offered	Lowest grade offered at the school	String	OSSE enrollment audit
highest_grade_offered	Highest grade offered at the school	String	OSSE enrollment audit
enroll_grade_pk3	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_pk4	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_kg	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_01	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_02	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_03	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_04	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_05	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_06	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_07	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_08	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_09	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_10	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_11	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_12	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_alt	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_adult	Number of students enrolled in each grade	Integer	OSSE enrollment audit
enroll_grade_sped	Number of students enrolled in each grade	Integer	OSSE enrollment audit
students_ward_1	Number of students living in each ward	String	OSSE Performance Oversight

Name in dataset	Definition	Variable type	Sources, all years
students_ward_2	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_3	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_4	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_5	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_6	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_7	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_8	Number of students living in each ward	String	OSSE Performance Oversight
students_ward_unknown	Number of students living in each ward	String	OSSE Performance Oversight
in_boundary_percent	Share of the school enrollment that is living in boundary (the numerator is the number of in-boundary students enrolled at the school, and the denominator is the number of students enrolled at the school)	Numeric	DME public school enrollments per DCPS boundary
boundary_participation_percent	Share of public school students living in each school's boundary who attend the school (the numerator is the number of in-boundary students enrolled at the school, and the denominator is all grade-specific public school students living in the boundary)	Numeric	DME public school enrollments per DCPS boundary
subgroup_african_american_percent	Share of students who are black or African American	Numeric	STAR Report Card and Equity Reports
subgroup_american_indian_alaskan	Share of students who are American Indian or Alaska Native	Numeric	STAR Report Card and Equity Reports
subgroup_asian_percent	Share of students who are Asian	Numeric	STAR Report Card and Equity Reports
subgroup_hawaiian_pacific _islander_percent	Share of students who are Native Hawaiian or other Pacific Islander	Numeric	STAR Report Card and Equity Reports
subgroup_latino_percent	Share of students who are Hispanic or Latino	Numeric	STAR Report Card and Equity Reports
subgroup_white_percent	Share of students who are white	Numeric	STAR Report Card and Equity Reports
subgroup_two_or_more_races _percent	Share of students who identify as two or more races	Numeric	STAR Report Card and Equity Reports
subgroup_english_learners_percent	Share of students who are identified as an English learner during the given school year and is between the ages of 3 and 21 as of the LEA's cutoff date	Numeric	STAR Report Card and Equity Reports
subgroup_at_risk_percent	Share of students who are at risk	Numeric	STAR Report Card
subgroup_special_ed_percent	Share of students with disabilities for the given school year, at any time throughout the school year and not just at the time of the enrollment audit	Numeric	STAR Report Card and Equity Reports
subgroup_level_1_percent	Share of students who are identified as level 1 special education for the given school year[a]	Numeric	Equity Reports

Name in dataset	Definition	Variable type	Sources, all years
subgroup_level_2_percent	Share of students who are identified as level 2 special education for the given school year[a]	Numeric	Equity Reports
subgroup_level_3_percent	Share of students who are identified as level 3 special education for the given school year[a]	Numeric	Equity Reports
subgroup_level_4_percent	Share of students who are identified as level 4 special education for the given school year[a]	Numeric	Equity Reports
star_score	The sum of the student group scores using all applicable STAR framework metrics (0 to 100 points)	Numeric	STAR Report Card
star_rating	A rating from one to five stars based on the star scale equivalency to the school's STAR score (e.g., a five-star school has a STAR score between 80 and 100)	Numeric	STAR Report Card
parcc_ela_4_plus_percent	In English language arts, share of students scoring at a level 4+	String	PARCC data
parcc_math_4_plus_percent	In math, share of students scoring at a level 4+	String	PARCC data
parcc_ela_4_plus_percent_african _american	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_american _indian_alaskan	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_asian	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_hawaiian _pacific_islander	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_latino	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_two_or _more_races	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_white	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_special_ed	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_english _learner	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_ela_4_plus_percent_at_risk	In English language arts, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent_african _american	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data

Name in dataset	Definition	Variable type	Sources, all years
parcc_math_4_plus_percent _american_indian_alaskan	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent_asian	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent _hawaiian_pacific_islander	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent_latino	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent_two _or_more_races	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent_white	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent _special_ed	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent _english_learner	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_math_4_plus_percent _at_risk	In math, share of students scoring at a level 4+ by subgroup	String	PARCC data
parcc_mgp_ela_percent	English language arts MGP, or a measure of the median academic growth of students at a school compared with students at other DC schools; MGP is a norm-referenced, school-level growth measure	String	STAR Report Card and Equity Reports
parcc_mgp_math_percent	Math MGP, or a measure of the median academic growth of students at a school compared with students at other DC schools; MGP is a norm-referenced, school-level growth measure	String	STAR Report Card and Equity Reports
graduation_four_year_percent	The four-year adjusted cohort graduation rate is the number of students who graduate in four years with a regular high school diploma divided by the number of students who form the adjusted cohort for the graduating class[b]	String	STAR Report Card, Equity Reports, and OSSE graduation rate
graduation_five_year_percent	The five-year adjusted cohort graduation rate is the number of students who graduate in five years with a regular high school diploma divided by the number of students who form the adjusted cohort for the graduating class[b]	String	STAR Report Card and Equity Reports
in_seat_attendance_percent	A student's attendance rate is calculated by dividing the number of days a student is present by the number of days the student is enrolled	Numeric	STAR Report Card and Equity Reports
reenrollment_percent	The share of students who can and do return to school each fall	Numeric	STAR Report Card

Name in dataset	Definition	Variable type	Sources, all years
midyear_entry_percent	The share of students entering a school from the October audit through June of the given school year as a share of the audited enrollment for the given school year	Numeric	STAR Report Card and Equity Reports
midyear_withdrawal_percent	The share of students exiting a school from the October audit through June of the given school year as a share of the audited enrollment for the given school year	Numeric	STAR Report Card and Equity Reports
expulsion	The share of students enrolled as of the enrollment audit for the given school year that were expelled from their audited school	Numeric	STAR Report Card and Equity Reports
suspension	Rates of students receiving out-of-school suspensions	Numeric	STAR Report Card and Equity Reports
truancy_percent	Share of students who are truant[c]	Double	OSSE attendance report
waitlist_total	Sum of students on waiting lists in all grades	Double	DCPS and PCSB websites
prog_dual_language	Indicator of whether a school offers dual language	String	My School DC
prog_ib	Indicator of whether a school offers the International Baccalaureate program	String	My School DC
prog_sports	Indicator of whether a school offers interscholastic sports	String	My School DC
prog_montessori	Indicator of whether a school offers Montessori	String	My School DC
census_tract	Census tract, 2010	String	MAR Geocoder
neighborhood_cluster	Neighborhood cluster, 2000	String	MAR Geocoder
advisory_neighborhood_commission	Advisory Neighborhood Commission, 2012	String	MAR Geocoder
council_ward_2012	Council ward, 2012	String	MAR Geocoder
zip_code	Zip code, 2002	Integer	MAR Geocoder
x_coordinate	X coordinate	Double	MAR Geocoder
y_coordinate	Y coordinate	Double	MAR Geocoder
latitude	Latitude	Double	MAR Geocoder
longitude	Longitude	Double	MAR Geocoder

Source:

Notes: DCPS = DC Public Schools; DME = Office of the Deputy Mayor for Education; LEA = local education agency; MAR = Master Address Repository; MGP = median growth percentile; OSSE = Office of the State Superintendent of Education; PARCC = Partnership for Assessment of Readiness for College and Careers; PCSB = Public Charter School Board; STAR = School Transparency and Reporting. "At risk," in DC, defines a student who possesses one of the following characteristics at any point during the given school year: Temporary Assistance for Needy Families enrollment, Supplemental Nutrition Assistance Program enrollment, identification as homeless by the student's school or other community partners, under the care of the Child and Family Services Agency (foster care), or overage (high school only). A high school student is overage if he or she is at least one year older than the expected age for his or her grade. "Level 4+" indicates the student is on track for the next grade level or is ready for college or a career. "Students with disabilities" are those receiving special education services under the Individuals with Disabilities Education Act and are ages 3 to 21 as of the date of the annual enrollment audit.

[a] Students are considered based on their highest level of need at any time during the school year, not at the time of the audit. For analysis by special education level, the denominator is the number of students with disabilities based on highest level of need and highest level of special education services during the school year.

[b] For any given cohort, students who are entering ninth grade for the first time form a cohort that is subsequently "adjusted" by adding any students who transfer into the cohort during the next three years and subtracting any students who transfer out, emigrate to another country, or die during that same period.

[c] Truancy is defined as the accumulation of 10 or more unexcused absences across all schools and sectors in a given school year for students of compulsory age. Compulsory age refers to students ages 5 to 18 as of September 30 of the school year.

Notes

1   Here and throughout, *public schools* refers to both traditional public and public charter schools.

2   School Transparency and Reporting report scores and ratings have been included as they will be available in future years.

3   Data for students by ward of residence have been suppressed if fewer than 10 students live in a particular ward. Data for 2014–15 were not based on the audit and are not included, as they are not comparable with other years.

4   Data on race or ethnicity are included as they were available from a previous project.

5   "Snapshot 2017: Item Definitions," Texas Education Agency, accessed October 9, 2019, https://rptsvr1.tea.texas.gov/perfreport/snapshot/2017/itemdef.html.

6   "The School-Level Master File," Research Alliance for New York City Schools, accessed October 9, 2019, https://research.steinhardt.nyu.edu/research_alliance/research/data_sets.

7   "Public Schools and Districts Data Files," California Department of Education, last updated May 28, 2019, https://www.cde.ca.gov/ds/si/ds/pubschls.asp.

8   Descriptions have been adapted from the sources themselves.

About the Author

Chelsea Coffin joined the DC Policy Center in 2017. She has more than 10 years' experience in early, K–12, and higher education at government agencies, foundations, nonprofits, multilateral institutions, and schools. Most recently, Coffin supported planning for DC's public charter sector with analyses of supply, demand, and other critical topics at the DC Public Charter School Board. She has also produced education case studies, evaluations, research, and monitoring tools for the World Bank, the Inter-American Development Bank, the Children's Investment Fund Foundation, and Educate! in Uganda. Coffin began her career in education as a secondary school teacher with the Peace Corps in Mozambique. She holds a bachelor's degree from Middlebury College and a master's degree in international economics and development from the Johns Hopkins University School of Advanced International Studies.

Acknowledgments

ABOUT THE DC POLICY CENTER

The DC Policy Center is a 501(c)(3) nonprofit think tank advancing policies for a strong and vibrant economy in the District of Columbia. The organization provides objective, high-quality data analyses to support wide-ranging and productive policy debate in the District of Columbia. The DC Policy Center has a singular focus on DC and keeps track of local policy priorities to ensure that its research is relevant, is meaningful, and speaks directly to the current policy environment. Its data and analyses have been used by DC government agencies, the DC Council, and a wide array of research and advocacy groups across the District.

1310 L Street NW
Washington, DC 20005

www.dcpolicycenter.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org