$$(\text{In})\textbf{SCHOOL} = b_0 + b_1\textbf{State}_i + b_2\textbf{LaborMarket}_{ij} + b_3\textbf{CWI}_{ij} + b_4\textbf{FINANCE}_{ij} + b_5\textbf{PopulationDensity}_{ij} + b_6\ \textbf{Enrollment}_{ij} + b_7\textbf{INDICATORS}_{ij} + b_8\textbf{Scale}_{ij} + b_9\textbf{Poverty}_{ij} + b_{10}\textbf{SchlType}_{ij} + b_{11}\textbf{DATABASE}_{ij} + e$$

DISTRICT COST DATABASE

USER'S GUIDE

VERSION 5.0 (RELEASED 2025)

This School Finance Indicators Database (SFID) is a collection of data and analysis measuring the adequacy and fairness of K-12 education finance and resources. Its purpose is to provide a single source of data for policymakers, the public, and researchers working in the fields of education finance and economics.

This **District Cost Database (DCD)** is one of two primary public datasets released annually as part of the SFID (this is the second release of the DCD). It allows users to assess the adequacy of K-12 public education spending for roughly 12,000 individual school districts between the 2008-09 and 2021-22 school years by comparing these districts' actual spending levels in each year to estimates of spending levels that would be required to achieve a common student outcome goal (i.e., national average test scores). The latter (required spending levels), which can be interpreted as imperfect but reasonable and policy-relevant adequate spending targets, are based on models, described below, that are designed to account for a host of educational and non-educational factors that affect the relationship between funding and outcomes. The database also includes a small group of "contextual" variables, such as district child poverty rates (U.S. Census) and districts' racial and ethnic composition, so that users can assess the relationship between spending adequacy and these other characteristics. This accompanying documentation is written to be accessible to all stakeholders, regardless of their background knowledge levels.

A similar set of NECM-based adequacy measures is also available for states (by district poverty quintile); these estimates are part of the State Indicators Database (SID), the SFID's other primary public dataset. Both the DCD and SID datasets, as well as online data visualization tools, reports and briefs using the data, and other resources, are available at: http://schoolfinancedata.org.

SECTIONS IN THIS GUIDE
1. **Data use agreement**
2. **Introduction to the database**: a brief non-technical description of methods; some caveats about interpreting and using the data; additional variables included in the full dataset; and how to access the DCD.
3. **List of variables**: a list of all variables in the database, with descriptions and notes.
4. **Changes to the dataset**: a record of significant changes since the original 2021 release

DATA USE AGREEMENT

The School Finance Indicators Database (SFID), as well as the contents of this guide, are the sole property of the authors. Public use of the datasets and results is encouraged, with proper attribution. Any alternative use of the data, models, or methods of the SFID must be approved by the authors.

You agree not to use the data sets for commercial advantage, or in the course of for-profit activities (with the exception of journalists). Commercial entities wishing to use the data should contact the Albert Shanker Institute at info@ashankerinst.org.

You agree that you will not use these data to identify or to otherwise infringe the privacy or confidentiality rights of individuals.

THE DATA SETS ARE PROVIDED "AS IS" AND THE AUTHORS, THE ALBERT SHANKER INSTITUTE, UNIVERSITY OF MIAMI, RUTGERS UNIVERSITY, THE WILLIAM T. GRANT FOUNDATION, THE BILL AND MELINDA GATES FOUNDATION, AND ALL OTHER ASSOCIATED PARTIES MAKE NO REPRESENTATIONS AND EXTEND NO WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED. THE ABOVE PARTIES SHALL NOT BE LIABLE FOR ANY CLAIMS OR DAMAGES WITH RESPECT TO ANY LOSS OR OTHER CLAIM BY YOU OR ANY THIRD PARTY ON ACCOUNT OF, OR ARISING FROM THE USE OF THE DATA SETS.

You agree that this Agreement and any dispute arising under it is governed by the laws of the District of Columbia of the United States of America, applicable to agreements negotiated, executed, and performed within the District of Columbia.

You agree to acknowledge the authors as the source of these data. In publications, please cite the dataset as:

Baker, Bruce D., Di Carlo, Matthew, Srikanth, Ajay, and Weber, Mark A. 2025. *School Finance Indicators Database: District Cost Database 2025 (5th Release).* Washington, DC: Albert Shanker Institute. Retrieved from: http://www.schoolfinancedata.org.

Subject to your compliance with the terms and conditions set forth in this Agreement, the Albert Shanker Institute, University of Miami, and Rutgers University grant you a revocable, non-exclusive, non- transferable right to access and make use of the Data Sets.

INTRODUCTION TO THE DATABASE

The District Cost Database (DCD) is a dataset (20 variables) of K-12 school funding adequacy and related measures for approximately 12,000 individual U.S. public school districts in each year between 2009 and 2022. The DCD is part of the School Finance Indicators Database (SFID), a collection of education funding and resource allocation data and analysis published annually by researchers from the Albert Shanker Institute, the University of Miami School of Education and Human Development, and the Rutgers University Graduate School of Education.

A quick overview of the DCD and the National Education Cost Model

In this section, we briefly describe our methods in non-technical terms. For a more thorough discussion of the data and models, as well as other issues addressed below, see Baker et al. (2021) and Baker et al. (2018).

The most important measure included in this district database is what we call "required spending," which is the estimated amount of spending (per-pupil) each district would have to spend in order to achieve the common outcome goal of national average math and reading scores. This variable is central, of course, because it is the common standard against which we assess the adequacy of *actual* district spending (i.e., total spending, direct to elementary and secondary education). We interchangeably refer to required spending as "adequate spending," "predicted cost," or "cost target."

Our required spending estimates are derived from the National Education Cost Model (NECM), which is part of the SFID. The NECM has been used to calculate *state-level* adequacy measures for the SFID since the first SFID release in 2019 (also see Baker et al. 2018).

This *district-level* DCD dataset also presents estimates from the NECM, which is continually updated and improved, to produce the same basic set of measures, but for individual school districts. The NECM estimates are, to our knowledge, the first input-/output-based measures of state and district spending adequacy that can be (carefully) compared across states. This release of the DCD includes estimates for each year between 2009 (the 2008-09 school year) and 2022 (the 2021-22 school year).[1]

The NECM uses a dataset of district test scores, funding, and numerous other variables between 2009 and 2022.[2] The DCD is published every year with an additional year of data. However, the variables used in the NECM are often corrected in prior years by the federal agencies that publish them, and we are constantly updating and improving the model. In addition, new student outcome data are released periodically. As a result, district estimates for a given year can and do vary between DCD releases.

---

[1] Note that the release of our State Indicators Database and products also includes NECM estimates between 2009-22, but they are provided for entire states (and for states by poverty quintile). These state estimates may differ slightly from aggregations of the estimates in DCD releases (the state measures are aggregated district measures) due, for instance, to improvements to the model implemented between the time of the SID and DCD releases, and districts excluded based on small sample sizes.

[2] In addition to school finance data from the U.S. Census Bureau (U.S. Census Bureau 2021), school and district characteristics data from the National Center for Education Statistics (NCES 2021), and several other data sources discussed in Baker et al. (2021), the NECM relies heavily on three additional datasets. The first is the Comparable Wage Index for Teachers (Cornman et al. 2019), an index of regional wage and salary variation developed by researchers at the NCES in collaboration with Dr. Lori Taylor of Texas A&M, who worked with NCES to develop the original version of the index in 2006. The second is the EDGE School Neighborhood Poverty Index, also published by the NCES, which is specifically designed to measure poverty surrounding schools and districts (Geverdt 2019). The third and perhaps most important NECM data source is the Stanford Education Data Archive (SEDA), a groundbreaking database of nationally-normed test scores going back to 2009 (Reardon et al. 2024). The SEDA allows for a better comparison of individual districts' test results across all states, a crucial tool for producing cost model estimates that are comparable across the U.S.

The core purpose of the NECM is to account for the fact that the cost of providing a given level of education is not uniform across districts. Perhaps most importantly, districts that serve larger shares of high-needs students (e.g., higher Census child poverty rates) will have higher costs. In addition, other factors, such as labor costs (e.g., districts in areas with higher costs of living will need to pay their employees more) size (economies of scale), and population density, all affect the "value of the education dollar." The model, therefore, first estimates the relationships between district spending and these important factors, including testing outcomes.

Importantly, the model the model accounts for the fact that school funding both affects and is affected by testing outcomes. For example, a district with higher test scores will tend to have higher property values than a district with lower scores. This high valuation allows the former district to collect more property tax revenues, which, in turn, boosts spending and positively affects testing outcomes. The NECM uses econometric methods to account for this endogeneity and tease out the relationship between spending and outcomes.

This initial model yields a kind of "relationship inventory" of how each factor is related to spending. We then use the "inventory" to predict the cost (spending levels) of achieving a common outcome level (e.g., national average math and reading test scores) for each individual district, based on that district's configuration of characteristics (in a sense, by comparing each district to similar districts). These "required spending" estimates can then be compared with *actual* spending levels in each district. The comparison between actual and required spending produces measures of adequacy relative to the common goal of national average scores.

Adequacy can be expressed in different ways, such as per-pupil funding gaps (actual minus required spending), actual spending as a percentage of required spending, or the percentage difference between actual and required spending.

A note on missing and excluded data. Estimates are not available for every single U.S. school district (i.e., the database does not include all districts), and not all districts are available in every year. Some of these districts are excluded due to missing finance and/or testing data. This includes but is not limited to fiscally-independent charter schools or other types of special schools or service centers. Wherever feasible, data are imputed to maximize our non-missing sample. We have also decided to exclude from the final database estimates for districts that serve fewer than 100 students, as results based on these small samples tend to be less reliable. As a result, there are a small number of districts for which data are available in some years but not others (i.e., if their enrollments fluctuate above and below 100 students). There are no estimates available in any year for Hawaii, as the state consists of a single geographically-isolated government-run school district, or for Vermont, due to data irregularities. We also exclude estimates for Alaska in all years in all SFID publications. This is due to the uniqueness of the state's climate, size, and other factors, which affect education costs in ways that we have decided that we cannot account for sufficiently in our model.

Limitations of the measures

It is important to interpret DCD estimates with caution. Even if we had a way to calculate perfect estimates of education costs, we would certainly never imply that these spending levels, if put into place in a given state or district, would quickly and certainly raise scores to the national average. This not only because that implication assumes efficient use of the additional funds, but also because real improvement is gradual and requires sustained investment.

More importantly, of course, our estimates are far from perfect. This is true of all cost models, but the NECM contends with particularly daunting challenges insofar as it is estimating education costs across the entire nation. Most basically, no model can control for everything (researchers call this "omitted variable bias"). The NECM includes numerous variables to help "capture" the (bi-directional) relationship between funding and student outcomes, but there are unobserved (i.e., unmeasured or unmeasurable) factors that we cannot include. And estimating costs across all states exacerbates this problem (e.g., comparing costs between, say, Connecticut and Mississippi).

Second, the variables that we *do* have are imprecise. Our dataset of test scores, for example, represent a truly groundbreaking effort to make all states' tests comparable across the nation (Reardon et al. 2024), but these methods, as well as the underlying state testing data, are necessarily subject to error (random and systematic). Similarly, on the funding side of the equation, our data may be biased by differences between states in how spending is tracked and reported to federal agencies (despite the best efforts of the latter). We have specific concerns about recent spending data from Vermont and New York, and about testing outcome data in western and upstate New York.

Third, it bears emphasizing that our cost estimates are based on common outcomes defined solely in terms of math and reading scores in grades 3-8. This is a very narrow picture of student performance. Districts may be spending money in ways that benefit students but do not necessarily affect these testing outcomes.

Fourth and finally, our required spending estimates are based on a "benchmark" student outcome goal of national average student outcomes. This is a very modest goal. In reality, all states set their own target outcome goals (e.g., state standards), and they vary between states. Our purpose, however, is to evaluate states' finance systems in a comparable manner, and so we do so based on an outcome goal that is the same across all states. We choose the modest goal of national average scores because it minimizes the chances of our vastly overstating costs in states with lower goals, and also because it is a performance level that people can conceptualize.

For all these reasons, we recommend that users of the DCD interpret each district's estimates not in absolute terms (e.g., "this district's funding is adequate or inadequate"), but rather in relative terms (e.g., "this district's funding is *more* adequate or *less* adequate, compared with that of other districts").

We believe the NECM produces reasonable cost estimates that are useful for assessing spending adequacy against a common standard and, ultimately, for improving state and federal school finance policy. We are constantly updating and improving the model to address the issues discussed above.

Additional variables in the DCD

In addition to required and actual spending, as well as basic information about each district, such as district name, state, and NCES local education agency identification number, the DCD includes a small group of supplemental district-level variables:

1. **Student outcome gaps.** In order to facilitate the comparison of spending adequacy with actual testing outcomes, the database includes a variable measuring, in standard deviations, the difference between each district's average score (math and reading combined) and the U.S. average (Reardon et al. 2024). These data are available for most districts between 2009-19.
2. **District contextual characteristics.** We include a small set of measures of district characteristics, including U.S. Census child poverty rates (U.S. Census Bureau 2021), and the following composition variables from the National Center for Educational Statistics (NCES 2022): total enrollment; percent special education students; percent American Indian/Alaska Native students; percent Asian students; percent Black students; percent Hispanic students; percent Pacific Islander/Hawaii Native students; percent multiracial ("two or more races") students; percent white students; and percent English language learners.[3]

Users who wish to incorporate additional variables can use the NCES unique district identification numbers included in the database to merge in district-level datasets.

Accessing the data

The full district dataset in Stata or Excel format, as well as a data visualization tool for viewing individual districts' estimates, are available at the SFID project website:
http://schoolfinancedata.org

---

[3] We use the terms "American Indian," "Black," and "Hispanic" in the database because these are the categories used by the NCES.

References

Baker, Bruce D., Weber, Mark, and Srikanth, Ajay. 2021. Informing Federal School Finance Policy with Empirical Evidence. *Journal of Education Finance* 47(1): 1-25*.*

Baker, Bruce D., Weber, Mark, Srikanth, Ajay, Kim, Robert, and Atzbi, Michael. 2018. *The Real Shame of the Nation: The Causes and Consequences of Interstate Inequity in Public School Investments*. New Brunswick, NY: Rutgers University. http://tinyurl.com/4xcswckc

Cornman, Stephen Q., Nixon, Laura C., Spence, Matthew J., and Taylor, Lori L., and Geverdt, Doug E. 2019. Education Demographic and Geographic Estimates (EDGE) Program: American Community Survey Comparable Wage Index for Teachers (ACS-CWIFT) (NCES 2018-130). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from: http://nces.ed.gov/pubsearch/

Geverdt, Doug. 2019. *Education Demographic and Geographic Estimates Program (EDGE): School Neighborhood Poverty Estimates, 2016-2017 (NCES 2019-112)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from: http://nces.ed.gov/pubsearch/.

National Center for Education Statistics (NCES). 2022. Common Core of Data (2008-09 to 2019-20). Washington, D.C.: National Center for Education Statistics. https://nces.ed.gov/ccd/

Reardon, Sean F., Ho, Andrew D., Shear, Benjamin R., Fahle, Erin M., Kalogrides, Demetra, and Saliba, Jim. 2024. Stanford Education Data Archive (Version 5.0). Palo Alto, CA: Stanford University. https://purl.stanford.edu/cs829jn7849

Taylor, Lori I. 2014. Extending the NCES Comparable Wage Index. College Station, TX: Texas A&M University. http://bush.tamu.edu/research/faculty/Taylor_CWI/

Taylor, Lori I., and Fowler, William J. Jr. 2006. *A Comparable Wage Approach to Geographic Cost Adjustment (NCES 2006-321)*. Washington, D.C.: National Center for Education Statistics. http://nces.ed.gov/pubsearch/

U.S. Census Bureau. 2023. Annual Survey of School System Finances (2009-21). Washington, D.C.: U.S. Census Bureau. https://www.census.gov/programs-surveys/school-finances.html

U.S. Census Bureau. 2022. Small Area Income and Poverty Estimates (2009-21). Washington, D.C.: U.S. Census Bureau. https://www.census.gov/programs-surveys/saipe.html

LIST OF VARIABLES

The table below provides a list of all variables in the District Cost Database, along with brief descriptions and, where applicable, additional notes. Only districts with non-missing actual/required spending estimates are included in the dataset in any given year. With the exception of state, state abbreviation, district identification number, and district name, all variables are continuous (numeric). With the exception of year, state, and state abbreviation, all variables are district-level estimates. This DCD release includes estimates for each year between 2009-19.

Variable name	Description	Notes
year	Year of data	Year refers to the spring semester of the school year (e.g., 2022 is the 2021-22 school year).
leaid	NCES district identification number	NCES unique district identifier (saved as a string variable).
district	District name	
state_name	State	
stabbr	State two-letter abbreviation	
ppcstot	Actual spending per pupil	Total state and local expenditures, direct to elementary and secondary education.
predcost	Required (adequate) spending per pupil	Based on national average test score benchmark.
fundinggap	Gap between actual and required spending per pupil	The difference ($) between *ppcstot* and *predcost* (negative values indicate spending below predicted cost targets).
outcomegap	Gap between district and U.S. average test scores (s.d.)	Expressed in standard deviations (negative values mean district scores below U.S. average). Outcome gaps are not available for all districts 2009-19, and are not available for any districts in 2020-2022.
enroll	Total student enrollment	
pov	Census child (5-17 year old) poverty rate	
iep	Percent special education students	Estimates not available for Colorado in any year, for two states (MS and OK) in 2009, and for a handful of non-CO districts in most years between 2010-2022.
ell	Percent English language learners	
amind	Percent American Indian/Alaska Native students	
asian	Percent Asian students	
black	Percent Black students	
hisp	Percent Hispanic students	
multi	Percent multiracial ("two or more races") students	Estimates not available (category not reported) in most states in 2009 and 2010.
pac	Percent Pacific Islander/Hawaii Native students	Estimates not available (category not reported) in most states in 2009 and 2010, and for New York in most years between 2011-16.
white	Percent white students	

CHANGES TO THE DATASET

This section provides a record of significant changes to the dataset since the initial release of the DCD in 2021.

Second release (February 2022)
- Addition of estimates for previous years (going back to 2009).
- Addition of all student race and ethnicity composition variables (previously only Black and Hispanic).

Third release (February 2023)
- No significant changes.

Fourth release (March 2024)
- Elimination of estimates in Alaska in all years.

Fifth release (May 2025)
- No significant changes.