STANFORD EDUCATION DATA ARCHIVE

TECHNICAL DOCUMENTATION

Version 2.0
4 December 2017

Erin M. Fahle, Benjamin R. Shear, Demetra Kalogrides, Sean F. Reardon,
Richard DiSalvo, and Andrew D. Ho

CONTENTS

WHAT IS SEDA?

The Stanford Education Data Archive (SEDA; seda.stanford.edu)[1] is an initiative aimed at harnessing data to help scholars, policymakers, educators, and parents learn how to improve educational opportunity for all children. SEDA includes a range of detailed data on educational conditions, contexts, and outcomes in school districts and counties across the United States. It includes measures of academic achievement and achievement gaps for school districts and counties, as well as district-level measures of racial and socioeconomic composition, racial and socioeconomic segregation patterns, and other features of schooling systems.

By making the data files available to the public, we hope that anyone who is interested can obtain detailed information about American schools, communities, and student success. We hope that researchers will use these data to generate evidence about what policies and contexts are most effective at increasing educational opportunity, and that such evidence will inform educational policy and practices.

The construction of SEDA has been supported by grants from the Institute of Education Sciences (R305D110018), the Spencer Foundation, the William T. Grant Foundation, the Bill and Melinda Gates Foundation, the Overdeck Family Foundation, and by a visiting scholar fellowship from the Russell Sage Foundation. Some of the data used in constructing the SEDA files were provided by the National Center for Education Statistics (NCES). The findings and opinions expressed in the research reported here are those of the authors and do not represent views of NCES, the Institute of Education Sciences, the Spencer Foundation, the William T. Grant Foundation, the Bill and Melinda Gates Foundation, the Overdeck Family Foundation, the Russell Sage Foundation, or the U.S. Department of Education.

The remainder of this document describes the source data and procedures used to prepare the 16 test score data files, 3 covariate data files, and geographic crosswalk file contained in SEDA 2.0.[2] Note that in order to access and use the data files, users must enter their email address and agree to the Data Use Agreement on the SEDA website.

TEST SCORE DATA

SEDA 2.0 contains 16 test score data files – eight files at the geographic school district-level and eight files at the county-level. Each file contains information about the distribution of academic achievement (means and standard deviations) as measured by standardized test scores administered in 3rd – 8th grade in mathematics and English/Language Arts (ELA). Estimates are reported for all students and by demographic subgroups for geographic school districts and counties. Each file contains either estimates

[1] Suggested citation for data: Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & DiSalvo, R. (2017). Stanford Education Data Archive (Version 2.0). Retrieved from http://purl.stanford.edu/db586ns4974.
Suggested citation for technical documentation: Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., DiSalvo, R., & Ho, A. D. (2017). Stanford Education Data Archive: Technical Documentation (Version 2.0). Retrieved from http://purl.stanford.edu/db586ns4974.

[2] See section “Versioning and Publication” for a summary of prior versions of SEDA and changes to SEDA 2.0.

for each grade and year separately ("long files") or contains estimates that are averaged across grades, years, and subjects ("pooled files"). In each data file there are variables corresponding to test score means, standard deviations, and their respective standard errors. In the pooled data, there is an additional variable describing the average increase in test scores each year. Table 1 lists the files and file structures. A complete list of variables can be found in Appendix D and in the codebook that accompanies this documentation.

Table 1. Test Score Estimates: Means, Standard Deviations, and Achievement Gaps.

Test Score Estimates: Means, Standard Deviations, and Achievement Gaps														
			Disaggregated by…					Estimated for…						
								Group				Gaps		
File Name	Form	Metric	Geographic District	County	Year	Grade	Subject	All	Race	Gender	ECD	Race	Gender	ECD
SEDA_geodist_long_CS_v20	Long	CS	X		X	X	X	X	X			X		
SEDA_geodist_long_GCS_v20	Long	GCS	X		X	X	X	X	X			X		
SEDA_geodist_long_NAEP_v20	Long	NAEP	X		X	X	X	X	X			X		
SEDA_geodist_long_State_v20	Long	State	X		X	X	X	X	X			X		
SEDA_geodist_poolsub_CS_v20	Pooled	CS	X				X	X	X			X		
SEDA_geodist_poolsub_GCS_v20	Pooled	GCS	X				X	X	X			X		
SEDA_geodist_pool_GCS_v20	Pooled	CS	X					X	X			X		
SEDA_geodist_pool_CS_v20	Pooled	GCS	X					X	X	To Be Released in Future Updates		X	To Be Released in Future Updates	
SEDA_county_long_CS_v20	Long	CS		X	X	X	X	X	X			X		
SEDA_county_long_GCS_v20	Long	GCS		X	X	X	X	X	X			X		
SEDA_county_long_NAEP_v20	Long	NAEP		X	X	X	X	X	X			X		
SEDA_county_long_State_v20	Long	State		X	X	X	X	X	X			X		
SEDA_county_poolsub_CS_v20	Pooled	CS		X			X	X	X			X		
SEDA_county_poolsub_GCS_v20	Pooled	GCS		X			X	X	X			X		
SEDA_county_pool_CS_v20	Pooled	CS		X				X	X			X		
SEDA_county_pool_GCS_v20	Pooled	GCS		X				X	X			X		

Metric: CS = Cohort Scale; GCS = Grade Scale; NAEP = NAEP Scale; State = State-referenced Scale
Academic Years: 2008/09 – 2014/15
Grades: 3 – 8
Subjects: Math, ELA
Race: white, black, Hispanic, and Asian
Race Gaps: white-black, white-Hispanic, white-Asian
ECD: economically disadvantaged (as defined by states)

COVARIATE DATA

SEDA 2.0 also provides estimates of socioeconomic, demographic and segregation characteristics of geographic school districts. The measures included in the covariates files come primarily from two sources: the 2006-2010 Education Demographic and Geographic Estimates (EDGE) and the Common Core of Data (CCD).[3] EDGE is a special school district-level tabulation of American Community Survey (ACS) data. It includes tabulations of demographic and socioeconomic characteristics of families who live in each school district in the U.S. and who have children enrolled in public school. Thus, it provides detailed data on the family characteristics of children enrolled in each school district. The CCD is an annual survey

[3] The EDGE raw data can be accessed at https://nces.ed.gov/programs/edge/. The CCD raw data can be accessed at https://nces.ed.gov/ccd/.

of all public elementary and secondary schools and school districts in the United States. The data includes basic descriptive information on schools and school districts, including demographic characteristics.

Three files in SEDA 2.0 contain CCD and ACS that data have been curated for use with the geographic school district-level achievement data. These data include raw measures as well derived measures (e.g., a composite socioeconomic status measure, segregation measures), and CCD data are imputed to reduce missingness in some years. The composite construction and imputation are described in detail in the Covariate Data Construction section of the documentation. Each of the three files contain the same variables, but differ based on whether they report these variables separately for each grade and year or average across grades (hence providing a single value per district per year) or average across grades and years (hence providing a single value per district).

Table 2 shows the structure of the three covariate data files.

Table 2. Covariate Data Files.

File Name	Form	Disaggregated by		
		District	Year	Grade
SEDA_cov_geodist_long_v20	Long	X	X	X
SEDA_cov_geodist_poolyr_v20	Pooled	X	X	
SEDA_cov_geodist_pool_v20	Pooled	X		

ACHIEVEMENT DATA CONSTRUCTION

OVERVIEW

Source data. The SEDA 2.0 achievement data is constructed using data from the ED*Facts* data system at the U.S. Department of Education (USEd), which collects aggregated test score data from each state's standardized testing program as required by federal law. The data include assessment outcomes for seven consecutive school years from the 2008-09 school year to the 2014-15 school year in grades 3 to 8 in English/Language Arts (ELA) and Math.

Under federal legislation, each state is required to test every student in grades 3 through 8 and in one high school grade in Math and ELA each year (high school data are not currently included in SEDA 2.0 due to differences across states in what grade they are administered). States have the flexibility to select (or design) and administer a test of their choice that measures student achievement relative to the state's standards. States then each set their own standards regarding the level of performance considered "proficient" in each grade and subject. States are required to report the number of students scoring at the proficient level, both overall and disaggregated by certain demographic subgroups, for each school. More often, states report the number of students scoring at each of a small number (usually 3-5) of ordered performance levels, where one or more levels represent proficient achievement.

When states report this information to the USEd, they are compiled into the ED*Facts* database. The ED*Facts* database reports the number of students disaggregated by subgroup scoring in each of the ordered performance categories, for each grade, year and subject. The student subgroups include race/ethnicity, gender, socioeconomic disadvantage, among others. The raw data include no suppressed cells, nor do they have a minimum cell size. In other words, each row of data corresponds to a school-by-subject-by-grade-by-year-subgroup cell, and *no individual student-level data is reported.*

Table 3 illustrates the structure of the raw data from ED*Facts* prior to use in constructing SEDA 2.0.

Table 3. Example Data Structure Before Data Are Aggregated to Geographic Districts.

State	School	Group	Subject	Grade	Year	Number of students scoring at…			
						Level 1	Level 2	Level 3	Level 4
A	1	All Students	Math	3	2009	26	87	185	32
A	1	All Students	ELA	3	2009	13	102	195	20
A	2	All Students	Math	3	2009	35	238	192	7
A	2	All Students	ELA	3	2009	7	278	187	0

Construction. The construction process, which produces the estimated means and standard deviations of achievement on a scale that is comparable across states, occurs in a series of ten steps. These steps are outlined in Figure 1. We provide a brief description of each step here, with additional detail in the subsequent sections.

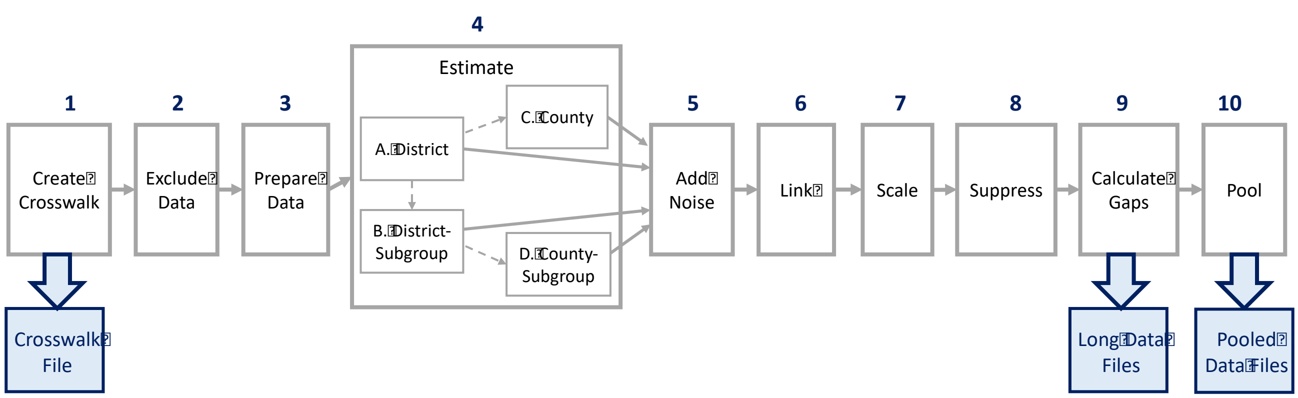

Figure 1. SEDA 2.0 Construction Process.

Throughout the documentation we use the term "group" to refer to the unit of analysis for which we are summarizing performance with a mean and standard deviation. Because we produce estimates at multiple units of analysis, the term "group" may refer to all students within a geographic school district, students of a particular demographic subgroup within a geographic school district, all students within a county, or students of a particular subgroup within a county.

1. Creating the Crosswalk and Defining Geographic School Districts. Each public school in the U.S. can be thought of as belonging to both an "administrative school district" (the local education agency that has administrative control over the school) and a "geographic school district" (GSD; a geographic catchment area defined by a traditional public school district). Each school's administrative district is defined as the school's local education agency (LEA) as reported by the ED*Facts* data and NCES Common Core of Data (CCD). Each school's GSD is defined as the traditional public school district geographic boundaries within which the school is physically located. Most traditional public schools have the same geographic and administrative district. There are a number of other types of schools (e.g. charter schools, virtual schools), however, that belong to an administrative district that does not have a corresponding geographic boundary. We assign each such school to a GSD (using a set of decision rules described in detail below). There are also a small number of cases where we do not use the schools' administrative NCES school district ID.

SEDA 2.0 contains estimates for GSDs because they allow linking of achievement data to demographic and economic information from EDGE/ACS, which is reported for students living in GSD boundaries regardless of where they attend school.

The crosswalk also links each GSD uniquely to a county, such that the data can be aggregated from GSDs to counties. The assignment to counties is taken from the CCD data.

2. Excluding Data Prior to Estimation. Estimation is performed simultaneously for all GSDs within a state-subject-grade-year case. In some cases we are not able to carry out the estimation: (1) students took different tests within the state-subject-grade-year; (2) the rate of participation was lower than 95% in the state-subject-grade-year; or (3) data for the state-subject-grade-year was not reported to ED*Facts*. No estimates are reported for any GSD or county for these state-subject-grade-year cases. In addition to the exclusion of entire state-subject-grade-year cases, we also made idiosyncratic eliminations of individual GSDs due to identified data errors.

3. Preparing Data for Estimation. There are three practical challenges that limit our ability to estimate test score distributions from proficiency data. First, per our data agreement with NCES, we cannot report estimates for groups that are based on test data for fewer than 20 students. Second, there are some groups with vectors of proficiency counts for which maximum likelihood estimates are not defined. In some of these cases we can estimate a unique mean for a group by placing additional constraints on the model, but cannot estimate a unique standard deviation; in others we can estimate neither a unique mean nor a unique standard deviation. Third, when groups have data for fewer than 50 students, estimates of standard deviations can be biased and very imprecise.

To address these issues, we take the following two steps to prepare each state-subject-grade-year case for estimation: 1) rearrange groups with fewer than 20 students or data that do not allow estimation of either a mean or standard deviation into "overflow" groups and 2) place additional constraints on the HETOP model.

4. Estimating Means and Standard Deviations. In this step, we use various forms of ordered probit models to estimate the mean and standard deviation of achievement for groups at different levels of geographic aggregation (GSD, GSD-subgroups, counties, and county-subgroups) from the proficiency count data. For each group we estimate the mean, standard error (SE) of the mean, standard deviation, and SE of the standard deviation of academic achievement from the frequency counts. We use either Heteroskedastic Ordered Probit (HETOP), Partially Heteroskedastic Ordered Probit (PHOP), or Homoskedastic Ordered Probit (HOMOP) models to estimate these parameters, as described in Reardon, Shear, Castellano and Ho (2017). The use of either HETOP, PHOP, or HOMOP models depends primarily on the nature of the data in each state-subject-grade-year case and is described below. Each of these models estimates the means and standard deviations of normal distributions that share a common scale and are most likely to have produced the observed proficiency counts. That is, the estimated means and standard deviations summarize the achievement represented by the observed counts. We fit these ordered probit models at the GSD level and use the results to estimate GSD-subgroup, country and county-subgroup means and standard deviations so that all share a common scale. All estimates are obtained using maximum likelihood estimation.

A. GSD Estimates. Estimation of GSD test score distribution parameters is performed simultaneously for each GSD within a state-subject-grade-year case. Depending on the number of proficiency categories, we use either a HETOP model or a HOMOP model. The HETOP model is used for all state-subject-grade-year cases that report performance in three or more proficiency categories. To reduce bias and sampling error, we place constraints on the standard deviation of any GSD with fewer than 50 students during estimation, and refer to this model as a PHOP model. The HOMOP model estimates a unique mean for each GSD but constrains all GSDs to have an equal standard deviation; the HOMOP model is used only in cases that report data in two proficiency categories, for which it is not possible to estimate a unique standard deviation for each group. The resulting estimates from all three models are scaled in units of state-grade-year-subject student-level test score standard deviations. Each model also produces estimates of the location of the cut scores distinguishing between the ordered proficiency levels.

B. GSD-Subgroup Estimates. In order to estimate the GSD-subgroup means, standard deviations, and associated SEs, we use the cut scores estimated in 4A. Fixing the cut scores keeps the GSD-subgroup estimates on the same scale as the GSD estimates and has the additional benefit of simplifying the computation – specifically, with known cut scores the likelihood function is separable. We still conduct the GSD-subgroup estimation simultaneously for each subgroup type in order to keep the same model structure – either HETOP/PHOP or HOMOP. In the case of the PHOP model, we constrain the standard deviations for the subgroups relative to all other students of the same subgroup (rather than relative to all students).

C. County Estimates. In order to estimate county means, standard deviations, and associated SEs, we aggregate the GSD estimates from step 4A. Specifically, we estimate the mean of each county as an enrollment-weighted mean of all GSDs within a county and we estimate the standard deviation of a county as the total standard deviation across GSDs within a county, taking into account both variation within and between GSDs. This approach (and that used in 4D) ensures the county and county-subgroup estimates can be interpreted on the same scale as the GSD estimates in 4A.

D. County-Subgroup Estimates. In order to estimate county-subgroup means, standard deviations, and associated SEs, we aggregate the GSD-subgroup estimates from step **4B**. Again, we estimate the county-subgroup mean and standard deviation as a weighted average and estimate of the total variation among all GSD-subgroups within the county, separately for each subgroup type.

5. Adding Noise to the Estimates. Our agreement with the US Department of Education requires that a small amount of random noise is added to each estimate in proportion to the sampling variance of the respective estimate. This is done to ensure that the raw counts of students in each proficiency category cannot be recovered from published estimates. Imprecise estimates have greater noise added and more precise estimates have less noise added. The SEs of the means are adjusted to account for the additional error. The added noise is roughly equivalent to randomly removing one student's score from each GSD-subgroup-subject-grade-year estimate.

6. Linking Means and Standard Deviations. The estimated means and standard deviations produced in step **4** are standardized relative to their state-subject-year-grade-specific distributions. Using the National Assessment of Educational Progress (NAEP), we place the estimates for each state-subject-year-grade on the NAEP scale so that they are comparable across states, years, and grades. Reardon, Kalogrides, and Ho (2017) describe the method used to link the state-specific estimates to the NAEP scale and provide a set of validity checks for the method.

7. Scaling Means and Standard Deviations. In order to make these linked estimates useful and interpretable, they are standardized in two ways.

CS Scale. Standardized by dividing by the national grade-subject-specific standard deviation for a given cohort in our data (those who were in 4^{th} grade in 2009 and 8^{th} grade in 2013). This metric is interpretable as an effect size, relative to the grade-specific standard deviation of scores in a single, common cohort. This scale can be used to describe aggregated change over time in test scores.

GCS Scale. Standardized relative to the average difference in NAEP scores between students one grade level apart in a given cohort. A one-unit difference in this grade-equivalent unit scale is interpretable as equivalent to the average difference in skills between students one grade level apart in school.

GSD means reported on the CS scale have an overall average near 0 and tend to range from -1 to +1. GSD means reported on the GCS scale have an overall average near 5.5 and tend to range from 1 to 10. As examples, a GSD with a mean of 0.5 on the CS scale represents a GSD where the average student scored approximately one half of a standard deviation higher than the national reference cohort scored in that same grade. A GSD with a mean of 6 on the GCS scale represents a GSD where the average student scored at about the same level as the average 6^{th} grader in the national reference cohort.

The standardization and interpretation of the scales is described in more detail in the corresponding section of this documentation and in Reardon, Kalogrides and Ho (2017).

8. Suppressing Data Post-Estimation. We suppress (do not report) a small number of estimates in the publicly available versions of SEDA 2.0 for data quality reasons. We suppress estimates using four stages: (1) removing GSD- and county-subject-grade-year cases where the participation rate is less than 95%; (2)

removing GSD- and county-subgroup-subject-grade-year cases where the participation rate is less than 95%; (3) removing GSD- and county-subgroup-subject-grade-year cases where the number of test scores reported by subgroups is less than 95% of the number of reported test scores for all students (i.e., cases where subgroup information is missing for more than 5% of students); and (4) removing imprecise individual estimates. Participation rate data is available for all students as well as for subgroups for the 2012-13 through 2014-15 school years and we suppress estimates using all four stages; in the 2008-09 through 2011-12 school years, we suppress test score means using only the third and fourth stages because participation rate data are not available in these years. In stage (4) an imprecise estimate is defined as any estimate in the state-standardized scale (the estimates produced in Step **4** above) with a standard error greater than 2 standard deviations on the state-standardized scale.

9. Calculating Achievement Gaps. We calculate three types of achievement gaps for GSD-subject-grade-year cells where there is sufficient data (estimated means for both subgroups): white-black, white-Hispanic, and white-Asian. Gaps are always computed in the order in which the subgroups are labeled; for example, the white-Black gap is calculated as $mean(white) - mean(black)$. We calculate these by taking the differences of the (noisy, linked, scaled) estimates and calculate the SE of the gap as the square root of the sum of the squared SEs of the individual estimates. Note that the achievement gaps reported in SEDA 2.0 differ from those reported in earlier versions of SEDA (earlier versions used the V-statistic to measure gaps). Future versions of the data may include achievement gaps between additional subgroups.

10. Pooling Mean, Standard Deviation, and Gap Estimates. SEDA 2.0 provides subgroup-grade-year-subject specific estimates, as well as estimates pooled across grades and years (within GSDs, for each subject separately) and estimates pooled across grades, years and subjects (within GSD-subgroups). Pooling provides more precise estimates of GSD and county test score patterns than do individual GSD-subgroup-grade-year-subject or county-subgroup-grade-year-subject estimates. SEDA 2.0 provides pooled estimates based on random coefficient (multi-level) models. These models are based on up to 84 subject-grade-year estimates for a given GSD or county, adjusting for grade and cohort (and subject in the latter model). The models weight the estimates by the precision of each of the 84 estimates. They allow each GSD or county to have a subject-specific intercept (average score), a subject-specific linear grade slope (rate at which scores change across grades, within a cohort), and a subject-specific cohort trend (the rate at which scores change across student cohorts, within a grade). See Reardon, Kalogrides and Ho (2017) for details. We report both Empirical Bayes (EB) and ordinary least squares (OLS) estimates and their SEs for all model parameters. We do not report estimates when the reliability of the OLS parameter estimate is less than 0.70.

1. Creating the Crosswalk & Defining Geographic School Districts

Placing Schools in Geographic School Districts

Each public school in the U.S. can be thought of as belonging to both an “administrative school district” (the local education agency that has administrative control over the school) and a “geographic school district” (GSD; a geographic catchment area defined by traditional public school district). Each school’s administrative district is defined as the school's local education agency (LEA) as reported by the ED*Facts* data and NCES Common Core of Data (CCD). Each school's GSD is defined as the traditional public school district in whose geographic boundaries the school is physically located.

Most traditional public schools have the same geographic and administrative district. However, there are a number of other types of schools (e.g. charter schools, virtual schools) that do not belong to an administrative district that has a corresponding geographic boundary. We use the following decision rules to assign those schools to a GSD.

A. Charter schools. If a charter school is listed in the CCD as belonging to a traditional public school district that has a corresponding geographic boundary, its GSD is then the same as its administrative district regardless of where the charter school is located. If a charter school is listed in the CCD as belonging to an administrative district that only has charter schools or is authorized by a state-wide administrative agency, it is geo-located and assigned to a GSD based on location.[4]

B. Virtual schools. By their nature, most virtual schools do not draw students from within strict geographic boundaries. We identify schools as virtual using the CCD data from 2013-14 and 2014-15. The virtual school identifier did not exist in earlier years of data, so we use the 2013-14 indicator to identify schools as virtual in earlier years. In other words, schools flagged as virtual in 2013-14 are flagged as virtual schools in all previous years of our data.

Additionally, we identify virtual schools by searching school names for terms such as "virtual", "cyber", "online", "internet", "distance", and "extended". Since schools may change names, if we identify a school as virtual by this approach in one year, we flag the school as virtual in all years. Some naming or classification of schools was ambiguous. When the type of school was unclear, research staff consulted school and district websites for additional details. Schools whose primary mode of instruction was online but that required regular attendance at a computer lab or school building were coded as belonging to the GSD in which they are located. For purposes of estimating district test score means, virtual schools are retained in the estimation, but are assigned a unique GSD ID (separate from the "overflow" group described above), so that their students' scores are included in the estimation procedures, but are not included in any existing GSD's score distribution. Estimates for virtual schools are not reported in the SEDA 2.0 data.

C. Schools in high school districts. In the cases where schools in high school districts serve students in grades 7 and 8, elementary school district boundaries are used and the high schools with grades 7 and 8 are assigned to the elementary school district in which they are geographically located.

D. Schools in districts that cross state boundaries. A few school districts overlap state borders. In this case, schools on either side of the state border take different accountability tests. We treat these districts as two GSDs, each one coded as part of the state in which it resides.

---

[4] Geographic location is determined by the latitude and longitude coordinates of a school's physical address as listed in the CCD. The location of charter schools sometimes varies from year to year. This can result in the charter school being placed in different geographic districts in different years. Approximately 4% of the roughly 8,000 charter schools that are represented in SEDA 2.0 change geographic school districts at least once.

In addition to these cases, there are two incidences where we do not use the administrative districts as defined by NCES, as follows:

> **A. Schools in districts that restructure.** Some districts changed structure during the time period covered by SEDA 2.0 data. We have identified a small number of these cases. In California, two Santa Barbara districts (LEA IDs: 0635360, 0635370) joined to become the Santa Barbara Unified School District. In South Carolina, two districts joined to become the Sumter School District (LEA IDs: 4503720, 4503690). In Tennessee, Memphis Public Schools and Shelby County Public Schools (LEA IDs: 4702940, 4703810) merged. North Forest ISD merged with Houston ISD in Texas (LEA IDs: 4833060, 482364). In all cases, SEDA 2.0 contains estimated test score distributions for the two original GSDs in all years in order to link them to covariate data from the EDGE. A single estimate for the new combined district can be obtained by computing the weighted average of the means or standard deviations within each grade, year, and subject.
>
> **B. Schools in New York city.** The CCD assigns schools in New York City to one of thirty-two districts or one "special schools district." We aggregate all New York City Schools to the city level and give them all the same GSD code, creating one unified New York City GSD code.

Geographic Crosswalks and Shape Files

We provide a geographic crosswalk on the SEDA website that enables linking the GSD estimates to higher levels of aggregation: counties, Metropolitan Statistical Areas (MSAs), and Commuting Zones (CZs). MSAs are a group of contiguous counties defined by the Office of Management and Budget and used by the ACS for data collection. Commuting Zones are also divisions composed of contiguous counties. Commuting zones that cross one or more state boundaries are treated in the same way as MSAs. No additional geolocation is done in support of this crosswalk. GSDs are assigned to counties, MSAs, and CZs based on the county codes provided in CCD.

While every effort is made to ensure schools are placed in the proper GSD based on the decision rules described in the previous sections, if you believe that the crosswalk contains an error, please contact sedasupport@stanford.edu.

The shape files used to locate schools within each geographic unit are also available online. The county, MSA, and commuting zone shape files are original from the US Census Bureau. A district level shape file was created using the U.S. Census Bureau's 2010 TIGER/Line Files. These files were from the National Historical Geographic Information System (NHGIS). The Census Bureau provides three shape files: elementary district boundaries, high school district boundaries, and unified district boundaries. Research staff merged the elementary and unified shape files to conform to the decision rules outlined above.

2. Excluding Data Prior to Estimation

There are three general cases when we cannot estimate GSD test score means or standard deviations:

A. Students took different tests within the state-subject-grade-year. There are two common ways this appears within the data. First, cases where districts were permitted to administer locally-selected assessments. This occurred in Nebraska during SY 2008-2009 (ELA and Math) and SY 2009-2010 (Math). Second, students take end-of-course (rather than end-of-grade) assessments. This is the case in 7th and 8th grade math for California, Virginia and Texas. In both of these cases, assessments were scored on different scales and using different cut scores. Therefore, proficiency counts cannot be compared across districts or schools within these state-subject-grade-year cases.

B. The state had participation lower than 95% in the tested subject-grade-year. Using the ED*Facts* data, we are able to estimate a participation rate for all state-subject-grade-year cases in the 2012-13 through 2014-15 school years. We use the test score data, from which we have the number of reported test scores ($numscores_{dygb}$) for each GSD. We also use a separate participation data file, from which we have the total number of enrolled students reported by GSD-subject-grade-year ($numenrl_{dygb}$). This participation data file is not available prior to the 2012-13 school year, and therefore we cannot calculate participation rates prior to 2012-13. We aggregate the GSD level counts to get the number of test scores reported in a state-subject-grade-year ($numscores_{fygb}$) and the number of enrolled students reported in a state-subject-grade-year ($numenrl_{fygb}$). Participation is then the ratio of the number of scores to the number enrolled in a state-subject-grade-year:

$$\widehat{part}_{fygb} = \frac{numscores_{fygb}}{numenrl_{fygb}} \tag{2.1}$$

for each state f, year y, grade g, and subject b.

We suppress all GSD-level and county-level estimates for state-subject-grade-year cases where fewer than 95% of students participated in tests in the state-grade-year-subject. For example, in the 2014-15 school year, New York had less than 95% participation in math in grade 8. Therefore, we report no estimates for school districts in New York for 8th grade math in 2014-15.

This state-level suppression is important because both the quality of the estimates and the linkage process depends on having the population of student test scores for that state-subject-grade-year. State participation may be low due to a number of factors, including student opt out (e.g., NY 1415) or pilot testing (e.g. CA 1314). Note that we do not suppress any entire state-subject-grade-year cases prior to the 2012-13 school year as enrollment data is not available in ED*Facts*. However, opt out was low in 2012-13 (no state was excluded based on this threshold), which suggests states met 95% threshold in prior years when data is not available.

C. Insufficient data was reported to ED*Facts*. There are a small number of cases where state data was not reported to ED*Facts*. Estimates are not available for these years. For example, Wyoming did not report any assessment outcomes in SY 2009-2010. In Colorado in the 2009, 2010, and 2011 school years, we have data reported in only two proficiency categories and a large majority of the data (88% across subjects, grades, and years) fall into a single category. These data do not provide sufficient information to estimate means and standard deviations. Additionally, there are cases where states reported scores for only a small percentage of students. This is most common in the 2013-14 school year when states were transitioning between standardized tests; however, there were occasionally other reported issues, such as hacking, that affect the availability or quality of the data.

In addition to the exclusion of state-subject-grade-year cases, we also made idiosyncratic eliminations of individual GSDs due to identified data errors. For a single GSD, grade and year in Arkansas and Louisiana, respectively, the reported scores were implausible given the available data for other grades and years. In particular, the distribution of students across proficiency categories for the given cohort changed too abruptly in the given year compared with their performance in the prior and subsequent years, as well as compared with other cohorts in the GSD, to be believable change. These data were determined to be entry errors and were removed.

Complete lists of state-subject-grade-year estimates eliminated from the data and GSD-subject-grade-year cases removed are in reported in Appendix A.

3. Preparing Data for Estimation

In addition to the data cleaning and exclusions described above, we take additional data preparation steps to address an administrative requirement based on our data use agreement and two practical challenges that can arise with our HETOP estimation framework.

First, per our data agreement with NCES, we cannot report estimates for groups that are based on test data for fewer than 20 students. Second, there are some groups with vectors of proficiency counts for which maximum likelihood estimates are not defined. In some of these cases we can estimate a unique mean for a group by placing additional constraints on the model, but cannot estimate a unique standard deviation; in others we can estimate neither a unique mean nor a unique standard deviation. Third, when groups have data for fewer than 50 students, estimates of standard deviations can be biased and very imprecise.

To address these issues, we take the following two steps to prepare each state-subject-grade-year case for estimation: 1) rearrange groups with fewer than 20 students or data that do not allow estimation of either a mean or standard deviation into "overflow" groups and 2) place additional constraints on the HETOP model.

A. Rearrange groups based on reporting or estimation criteria*.* In this step we reconfigure groups within a state-subject-grade-year that have fewer than 20 students or vectors of counts that cannot support estimation of either a mean or a standard deviation into county-level "overflow GSDs." If GSD overflow groups still do not reach the 20-student threshold or their distributions cannot be estimated via maximum likelihood, they are then moved to a state-level overflow group. If the state-level overflow group does not meet the size threshold or cannot support estimation, it is removed from the data. Although we do not report estimates for the overflow groups (or the groups within them), this reconfiguration allows us to retain the maximal possible number of test scores in the estimation sample. This is important as the standardization and linking methods rely on having information about the full population of all students.

B. Designate which groups will be constrained in estimation. As described in more detail below, we can sometimes obtain better estimates (or can only obtain estimates) by placing additional constraints on the HETOP model used to estimate means and standard deviations. With small samples, attempting to estimate a unique standard deviation for each group can produce biased or extremely imprecise estimates. In other cases, the vector of counts for a particular group may not allow for estimation of both a unique mean and unique standard deviation, but it is possible to estimate a unique mean if the standard deviation is constrained. In our models, we constrain the standard deviation of some groups to be equal to a function of the freely estimated standard deviations of larger groups with sufficient data. The following types of groups require these constraints and are flagged in preparation for estimation:

- There are fewer than 50 student assessment outcomes in a group.
- All student assessment outcomes fall in only two adjacent performance level categories
- All student assessment outcomes fall in the top and bottom performance categories.
- All student assessment outcomes fall in a single performance level category.

This process is performed for all levels of aggregation (GSD, GSD-subgroup, county, and county-subgroup) within a state-subject-grade-year.

NOTATION

In the remainder of the document, we use the following notation:

- Mean estimates are denoted by $\hat{\mu}$ and standard deviation estimates by $\hat{\sigma}$.
- A subscript indicates the aggregation of the estimate. We use the following subscripts:

d = GSD	f = state	b = subject
c = county	y = year	r = subgroup
n = school	g = grade	

- A superscript indicates the scale of the estimate. The metric is generically designated as x. There are four scales:

  $state$ = state-referenced metric

  $naep$ = NAEP test score scale metric

  cs = Cohort scale metric

  gcs = Grade (within cohort) scale metric

4. ESTIMATING MEANS AND STANDARD DEVIATIONS

STATISTICAL FRAMEWORK

Formally, to estimate the mean and standard deviation of academic achievement we assume there is an unobserved continuous variable y^* that is normally distributed in each GSD.[5] The mean and standard deviation of y^* in GSD d are denoted μ_d^* and σ_d^*, respectively. Here, y^* represents an unobserved continuous measure of test performance not included in ED*Facts*. Instead, the ED*Facts* database includes a "coarsened" version of y^*. The coarsening divides y^* into K ordered proficiency categories that are defined by $K-1$ threshold values of y^*, denoted $c_1^*, \dots, c_{K-1}^*$, where $c_{k-1}^* < c_k^*$ for all k, and where we define $c_0^* \equiv -\infty$ and $c_K^* \equiv +\infty$. We do not observe the values of the c_k^*'s. For each group, we observe only the distribution of values of the coarsened variable, denoted $s \in \{1, \dots, K\}$, where $s \equiv k$ iff $c_{k-1}^* < y^* \leq c_k^*$. For most states, $2 \leq K \leq 5$.

Note that y^* is not necessarily directly equivalent to the continuous scale scores, y, in each GSD that would be reported on a state's test. Rather $y^*$is a monotonic increasing transformation of the scale score: $y^* = f(y)$ where f is a monotonic increasing function that renders the distributions of y^* normal in each GSD. There is no requirement that the distributions of the scale score y be normal in each GSD, only a requirement that some function f exists that would render them so. Moreover, our estimates are robust to this assumption (Ho & Reardon, 2012; Reardon et al., 2017).

Under this assumption, the model-implied proportion of students scoring in category k for GSD d is therefore

$$\pi_{dk} = \Phi\left(\frac{\mu_d^* - c_{k-1}^*}{\sigma_d^*}\right) - \Phi\left(\frac{\mu_d^* - c_k^*}{\sigma_d^*}\right) = \Pr(c_{k-1}^* < y_d^* \leq c_k^*) \equiv \Pr(c_{k-1} < y_d \leq c_k), \qquad (4.1)$$

where $\Phi(\bullet)$ is the standard normal cumulative distribution function. This is an instance of a heteroskedastic ordered probit (HETOP) model.

To formalize the model and introduce notation, let $\mathbf{N}$ be an observed $D \times K$ matrix with elements n_{dk} containing the counts of observations in GSD d for which $s = k$; let $\mathbf{P} = [p_1, \dots, p_D]$ be the $1 \times D$ vector of the GSD's proportions in the population (i.e., all students in the state-subject-year-grade); and let $\mathbf{n} = [n_1, \dots, n_D]$ be the $1 \times D$ vector of the observed sample sizes in each GSD, with $N = \sum_d n_d$.[6]

Our goal is to estimate the vectors $\mathbf{M}^* = [\mu_1^*, \dots, \mu_D^*]^t$, $\mathbf{\Sigma}^* = [\sigma_1^*, \dots, \sigma_D^*]^{\mathrm{t}}$ and $\mathbf{C}^* = [-\infty, c_1^*, \dots, c_{K-1}^*, +\infty]$. In practice, it is preferable to estimate $\mathbf{\Gamma}^* = [\gamma_1^*, \dots, \gamma_D^*]^t$, where $\gamma_d^* = \ln(\sigma_d^*)$. This ensures that the estimates of σ_d^* will all be positive. Following estimation of $\mathbf{\Gamma}^*$, we have $\widehat{\mathbf{\Sigma}}^* = \left[e^{\hat{\gamma}_1^*}, \dots, e^{\hat{\gamma}_D^*}\right]^t$. Given $\mathbf{M}^*$, $\mathbf{\Gamma}^*$, and $\mathbf{C}^*$, and under the assumption of conditional independence of scores within GSDs, the log likelihood of drawing a sample with observed counts $\mathbf{N}$ is

[5] Here we describe the HETOP framework for estimating district-level distributions, but the same general methodology can apply to the data grouped at other levels of aggregation (e.g., for district-subgroups, counties, etc.).

[6] Here we use $p_d = n_d/N$, although this is not necessarily required.

$$L = \ln[P(\mathbf{N}|\mathbf{M}^*, \mathbf{\Gamma}^*, \mathbf{C}^*)] = \sum_{d=1}^{D} \left\{ \ln(n_d!) + \sum_{k=1}^{K} [n_{dk} \ln(\pi_{dk}) - \ln(n_{dk}!)] \right\}$$

$$= A + \sum_{d=1}^{D} \sum_{k=1}^{K} n_{dk} \ln \left[\Phi\left(\frac{\mu_d^* - c_{k-1}^*}{e^{\gamma_d^*}}\right) - \Phi\left(\frac{\mu_d^* - c_k^*}{e^{\gamma_d^*}}\right) \right], \quad (4.2)$$

where $A = \ln\left(\frac{\prod_{d=1}^{D} n_d!}{\prod_{d=1}^{D} \prod_{k=1}^{K} n_{dk}!}\right)$ is a constant based on the observed counts in $\mathbf{N}$.

The parameters in $\mathbf{M}^*$, $\mathbf{\Sigma}^*$ (using $\mathbf{\Gamma}^*$), and $\mathbf{C}^*$, as well as $\mathbf{V}^* = Cov(\widehat{\mathbf{M}}^*, \widehat{\mathbf{M}}^*)$, $\mathbf{W}^* = Cov(\widehat{\mathbf{\Sigma}}^*, \widehat{\mathbf{\Sigma}}^*)$, and $\mathbf{Z}^* = Cov(\widehat{\mathbf{M}}^*, \widehat{\mathbf{\Sigma}}^*)$ can be estimated via maximum likelihood methods provided the following necessary assumptions are met (Reardon et al., 2017):

1. All observed frequency counts must be based on the administration of a common test, with common cut scores. This assumption is satisfied within any given state-subject-grade-year, and is the primary reason for carrying out the estimation for each state-subject-grade-year separately.[7]
2. Two linear constraints are needed to identify the scale of the estimates. We use a set of constraints and linear transformations that produce a scale such that the implied distribution of the underlying variable has a population mean of 0 and population standard deviation of 1 (Reardon et al., 2017).
3. There must be sufficient data in each GSD to estimate both μ_d^* and σ_d^*. When a state has only two proficiency categories, for example, or when all students in a GSD score in only two of the possible proficiency categories, there is not enough information to estimate both the μ_d^* and σ_d^*. Similarly, when a GSD has a very small number of students, estimates can be particularly noisy.

With respect to the last point, we can use two basic strategies to deal with this assumption being unmet in practice. First, during the data preparation stage (Step **3**), we combine GSD that have very small sample sizes or do not have sufficient data to estimate both parameters. Second, we can place additional constraints on the HETOP model. When a state has only two proficiency categories, we estimate a HOMOP model, in which the standard deviation for all GSDs is set to a single, fixed constant. That is, $\sigma_d^* = \sigma^*$ for all GSDs. In states with three or more proficiency categories, we can estimate a partially heteroskedastic ordered probit (PHOP) model. The PHOP model freely estimates μ_d^* and σ_d^* for GSDs with sufficient data and large sample sizes. For GSDs without sufficient data, or with small sample sizes (sample sizes below $n = 50$), a single pooled standard deviation parameter is estimated. Specifically, we

[7] As noted elsewhere, this requirement is not met in some state-grade-subject-year cases. For example, in 8th grade mathematics in CA, students take different tests depending upon the math course they are enrolled in. For these state-subject-grade-year cases, we do not estimate or report μ_d^* and σ_d^*. With the adoption of the Common Core State Standards (CCSS) and associated tests, some states administer common tests with common cut scores. Although it would thus be possible to combine states using a common assessment, the construction of SEDA conducts the estimation separately for each state.

constrain $\hat{\gamma}_d^* = \ln(\sigma_d^*)$ for the GSDs with small samples and/or insufficient data to be equal to the unweighted average $\hat{\gamma}_d^*$ of the remaining groups.[8]

A. GSD Estimates

We first estimate $\hat{\mu}_{dygb}^{state}$ and $\hat{\sigma}_{dygb}^{state}$, the mean and standard deviation of achievement in GSD d, year y, grade g, and subject b for each state. The estimates $\hat{\mu}_{dygb}^{state}$ and $\hat{\sigma}_{dygb}^{state}$ are estimated on a standardized scale in which the marginal distribution across all GSD in a given state-subject-grade-year case has a marginal mean of 0 and standard deviation of 1. This estimation is carried out separately for each state-subject-grade-year case by applying the HETOP model described above to the proficiency counts for a given state-subject-grade-year case aggregated to the GSD level. For states reporting only two proficiency categories, we use a HOMOP model. For states with three or more categories we use a HETOP or PHOP model, placing constraints as described above in Step **3**.

In what follows we will sometimes refer to $\widehat{\mathbf{M}}_{ygb}^{state}$ and $\widehat{\mathbf{\Sigma}}_{ygb}^{state}$, the vectors of $\hat{\mu}_{dygb}^{state}$ and $\hat{\sigma}_{dygb}^{state}$ estimates, respectively, and $\widehat{\mathbf{V}}_{ygb}^{state} = Cov\left(\widehat{\mathbf{M}}_{ygb}^{state}, \widehat{\mathbf{M}}_{ygb}^{state}\right)$, $\widehat{\mathbf{W}}_{ygb}^{state} = Cov\left(\widehat{\mathbf{\Sigma}}_{ygb}^{state}, \widehat{\mathbf{\Sigma}}_{ygb}^{state}\right)$, and $\widehat{\mathbf{Z}}_{ygb}^{state} = Cov\left(\widehat{\mathbf{M}}_{ygb}^{state}, \widehat{\mathbf{\Sigma}}_{ygb}^{state}\right)$ the estimated sampling covariance matrices of the estimates.

We also estimate $\widehat{\mathbf{C}}_{ygb}^{state}$, a vector containing $\hat{c}_{1ygb}^{state}, \ldots, \hat{c}_{(k-1)ygb}^{state}$, the $(K-1)$ estimated cut scores in the standardized metric for each state-subject-grade-year case. These are used for the subgroup estimation below.

B. GSD-Subgroup Estimates

We estimate $\hat{\mu}_{drygb}^{state}$ and $\hat{\sigma}_{drygb}^{state}$, the mean and standard deviation of subgroup r in GSD d, year y, grade g, and subject b for each state. Although it would be possible to use the same estimation approach described above to estimate means and standard deviations for student subgroups within GSDs, we use a slightly different approach in practice. We follow the paradigm above (i.e., use a HOMOP model for states with only two proficiency levels, and a PHOP or HETOP model for others) with two differences:

1. The groups are now student subgroups within GSDs (rather than entire GSDs), and the estimation is performed separately for each state-subgroup-subject-grade-year case.
2. We fix the thresholds for each state-subgroup-subject-grade-year model to be equal to $\widehat{\mathbf{C}}_{ygb}^{state}$, the thresholds estimated for the associated state-subject-grade-year model using entire GSDs. This results in GSD-subgroup mean and standard deviation estimates that are on the same scale as the estimates for complete GSDs.

When the cut scores are set at fixed values, the estimates of $\hat{\mu}_{drygb}^{state}$ and $\hat{\sigma}_{drygb}^{state}$ from each group are independent and could be estimated separately using the likelihood function in Equation (4.2) and treating the cut scores as known. However, we continue to estimate the group parameters

[8] The constraint is placed on the natural logarithm, rather than directly on the SDs, due to the manner in which maximum likelihood estimates are obtained.

simultaneously for each subgroup type so that we can place the equality constraints implied by the PHOP and HOMOP models. In practice, this is accomplished by fitting the same PHOP and HOMOP models, but with constraints placed on $\mathbf{C}$.

One additional step is needed for cases with only two proficiency levels. These models have only a single cut score, and hence the scale of the parameters is not identified solely by fixing the cut score; a second constraint is needed to identify the model. Specifically, because we fit a HOMOP model separately for each subgroup type and obtain a common $\hat{\sigma}_{rygb}^{state}$ parameter for each subgroup type, this parameter must be set to a fixed constant. In order to fix this parameter for identification while also ensuring the estimates are on a scale that can be compared to the GSD estimates, we adopt the following approach. We carry out the estimation with $\hat{\sigma}_{rygb}^{state}$ set to 1 and then re-scale the estimates so that the ratio of the marginal standard deviation across all GSDs for subgroup r in our model is equal to $\tilde{\sigma}_r = \hat{\sigma}_r^{NAEP}/\hat{\sigma}_T^{NAEP}$, where $\hat{\sigma}_r^{NAEP}$ and $\hat{\sigma}_T^{NAEP}$ are the NAEP estimates of the population standard deviation for subgroup r and for the total population in a given state-grade-year-subject. We use linear interpolation to obtain values of $\hat{\sigma}_r^{NAEP}$ and $\hat{\sigma}_T^{NAEP}$ in non-NAEP years and grades. In cases where NAEP does not report a standard deviation by subgroup, we assume this ratio is 1. A detailed review of this procedure is provided in Appendix B-1.

C. County Estimates

We adopt a different approach to estimate $\hat{\mu}_{cygb}^{state}$ and $\hat{\sigma}_{cygb}^{state}$, the mean and standard deviation of achievement in county c, year y, grade g, and subject b for each state. The approach used to estimate these values produces estimates in a metric that is again comparable to the one in which GSD estimates are reported. In brief, this is a two-step procedure:

1. Estimate GSD-level parameters $\hat{\mu}_{dygb}^{state}$ and $\hat{\sigma}_{dygb}^{state}$ as described above in 4A.
2. Use the GSD-level estimates from 4A to estimate an overall mean and variance for a county based on all GSDs within that county.

Suppose there are a set of C counties, each of which contains one or more unique GSDs. These higher-level units are defined geographically and are non-overlapping. Hence, each GSD falls within exactly one county. The county mean is estimated as the weighted average of GSD means across all D_c GSDs in county c, computed as

$$\hat{\mu}_{cygb}^{state} = \sum_{d=1}^{D_c} p_{dc}\hat{\mu}_{dygb}^{state}, \tag{4.3}$$

where p_{dc} is the proportion of county c represented by GSD d. The estimated county standard deviation is estimated as the square root of the estimated total variance between and within GSDs within a county,

$$\hat{\sigma}_{cygb}^{state} = \sqrt{\hat{\sigma}_{B_c}^2 + \hat{\sigma}_{W_c}^2} \tag{4.4}$$

where $\hat{\sigma}^2_{B_c}$ is the estimated variance between GSDs in county c and $\hat{\sigma}^2_{W_c}$ is the estimated variance within GSDs in county c. The formulas used to estimate $\hat{\sigma}^2_{B_c}$ and $\hat{\sigma}^2_{W_c}$ are based on equations in Reardon et al. (2017). These formulas and formulas for estimating the standard errors of the county means and standard deviations, $\hat{\mu}^{state}_{cygb}$ and $\hat{\sigma}^{state}_{cygb}$, are included in Appendix B-2.

D. County-Subgroup Estimates

In order to estimate county-subgroup means, standard deviations, and associated SEs, we aggregate the GSD-subgroup estimates from step **4B**. Again, we estimate the county-subgroup mean and standard deviation as a weighted average and estimate of the total variation among all GSD-subgroups within the county (as in **4C**), separately for each subgroup type.

5. Adding Noise to the Estimates

In the raw ED*Facts* files, no data are suppressed; proficiency counts are reported in all cells no matter how small the cell population. However, our agreement with the USDoE restricts publication of means and standard deviations to GSD-subject-grade-year cells with at least 20 assessment outcomes (in each group reported). As described above, we do not estimate unique means or standard deviations for any group where there are fewer than 20 students (see Section **3** above).

Additionally, our agreement requires that a small amount of random noise is added to GSD, GSD-subgroup, county, and county-subgroup estimates in proportion to the sampling variance of the respective estimate. This is done to ensure that the raw counts of students in each proficiency category cannot be recovered from published estimates.

The random error added to each to GSD, GSD-subgroup, county, or county-subgroup estimate is drawn from a normal distribution $\mathcal{N}(0,(1/n) * \widehat{\omega^2})$ where $\widehat{\omega^2}$ is the squared estimated standard error of the estimate and n is the number of student assessment outcomes to which the estimate applies. Imprecise estimates have greater noise added, and more precise estimates have less noise added. SEs of the mean are adjusted to account for the additional error. The added noise is roughly equivalent to the amount of error that would be introduced by randomly removing one student's score from each GSD-grade-year estimate.

Note that all linked, scaled and pooled estimates are based on the long-form noisy estimates; no additional noise is added in later steps.

6. Linking the Estimates to the NAEP Scale

The estimated means and standard deviations produced by the ordered probit model are scaled relative to their state-, grade-, year-, and subject-specific student-level test score distributions. We want to place these test score distributions on a scale that is common across states, grades, and years. To do so, we use data from the National Assessment of Educational Progress (NAEP) tests. NAEP data provide estimates of each state's 4th and 8th grade test score means and standard deviations in on a common scale in all states.

Note that the NAEP scales are not comparable across math and reading, but they are comparable across grades and years within each subject. We use these state-specific NAEP estimates to place each GSD's test score distribution on the NAEP scale. SEDA 2.0 data include GSDs' test score distributions scaled on both the state-specific scale and the NAEP scale. The methods we use—as well as a set of empirical analyses demonstrating the validity of this approach—are described by Reardon, Kalogrides, and Ho (2017). We provide a brief summary of the methods here. The equations shown here are in terms of the GSD-subgroups; however, the methodology is equivalent for whole GSDs, counties, and county-subgroups.

As above, we denote the estimated GSD means and standard deviations as $\hat{\mu}_{drygb}^{\text{state}}$ and $\hat{\sigma}_{drygb}^{\text{state}}$, respectively, for GSD d, subgroup r (e.g., all students, white students, black students, etc.), year y, grade g, and subject b. These means and standard deviations are expressed in units of their respective state-year-grade-subject student-level standardized distribution. The HETOP model estimation procedure also provides standard errors of these estimates, denoted $se\left(\hat{\mu}_{drygb}^{\text{state}}\right)$ and $se\left(\hat{\sigma}_{drygb}^{\text{state}}\right)$, respectively (Reardon, Shear, Castellano, & Ho, 2016).

In order to convert $\hat{\mu}_{drygb}^{\text{state}}$ and $\hat{\sigma}_{drygb}^{\text{state}}$ to their estimated corresponding values on the NAEP math and reading scales, we require estimates of NAEP means and standard deviations at the state (denoted s) level, denoted $\hat{\mu}_{sygb}^{\text{naep}}$ and $\hat{\sigma}_{sygb}^{\text{naep}}$, respectively, as well as their standard errors. Because NAEP is administered only in 4^{th} and 8^{th} grades in odd-numbered years, we interpolate and extrapolate linearly to obtain estimates of these parameters in grades (3, 5, 6, and 7) and years (2010, 2012, and 2014) in which NAEP was not administered. First, within each NAEP-tested year, 2009, 2011, 2013, and 2015, we interpolate between grades 4 and 8 to grades 5, 6, and 7 and extrapolate to grade 3. Next, for all grades 3-8, we interpolate between the NAEP-tested years to estimate parameters in 2010, 2012, and 2014, using the interpolation/extrapolation formulas here:

$$\begin{aligned} \hat{\mu}_{sygb}^{naep} &= \hat{\mu}_{sy4b}^{naep} + \frac{g-4}{4}\left(\hat{\mu}_{sy8b}^{naep} - \hat{\mu}_{sy4b}^{naep}\right), \quad \text{for g} \in \{3,5,6,7\} \\ \hat{\mu}_{sygb}^{naep} &= \frac{1}{2}\left(\hat{\mu}_{s[y-1]gb}^{naep} + \hat{\mu}_{s[y+1]gb}^{naep}\right), \quad \text{for y} \in \{2010, 2012, 2014\} \end{aligned} \tag{6.1}$$

We do the same to interpolate/extrapolate the state NAEP standard deviations. The reported NAEP means and standard deviations, along with interpolated values, by year and grade, are reported in Table 4 below.

Table 4. NAEP Means and Standard Deviations by Year and Grade.

	Grade	Reading / English Language Arts 2008	2009	2010	2011	2012	2013	2014	2015
Means	8	259.1	260.1	260.9	261.7	263.3	264.8	263.9	263.0
	7	248.5	249.3	250.0	250.7	252.1	253.4	252.8	252.3
	6	237.9	238.6	239.2	239.8	240.9	242.0	241.7	241.5
	5	227.3	227.8	228.3	228.8	229.7	230.5	230.6	230.8
	4	216.7	217.0	217.4	217.8	218.5	219.1	219.6	220.0
	3	206.1	206.2	206.5	206.8	207.3	207.7	208.5	209.3
SDs	8	36.8	36.3	36.1	35.8	35.6	35.3	35.6	35.8
	7	37.2	36.7	36.5	36.3	36.2	36.1	36.2	36.4
	6	37.5	37.0	36.9	36.9	36.9	36.9	36.9	36.9
	5	37.9	37.4	37.4	37.4	37.5	37.6	37.5	37.5
	4	38.2	37.7	37.8	37.9	38.2	38.4	38.2	38.0
	3	38.6	38.1	38.2	38.4	38.8	39.2	38.9	38.6

	Grade	Math 2008	2009	2010	2011	2012	2013	2014	2015
Means	8	279.1	280.1	280.8	281.4	282.1	282.7	281.6	280.4
	7	268.8	269.6	270.2	270.9	271.5	272.1	271.1	270.1
	6	258.5	259.1	259.7	260.3	260.9	261.6	260.7	259.8
	5	248.2	248.6	249.2	249.8	250.4	251.0	250.2	249.4
	4	238.0	238.1	238.7	239.2	239.8	240.4	239.8	239.1
	3	227.7	227.6	228.1	228.7	229.2	229.8	229.3	228.8
SDs	8	37.7	37.6	37.4	37.1	37.1	37.1	37.3	37.5
	7	35.7	35.7	35.5	35.3	35.3	35.4	35.6	35.8
	6	33.8	33.7	33.6	33.4	33.6	33.7	33.9	34.0
	5	31.8	31.8	31.7	31.6	31.8	32.0	32.1	32.3
	4	29.9	29.8	29.8	29.7	30.0	30.3	30.4	30.5
	3	27.9	27.9	27.9	27.9	28.2	28.6	28.7	28.8

Note: Reported in 2009, 2011, 2013, and 2015 in grades 4 and 8, interpolated and extrapolated elsewhere. Lighter shaded cells are the basis for year-based scaling; darker shaded cells are the basis for cohort-based scaling. These are expanded population estimates and may differ slightly from those reported in public reports.

Because GSD test score moments are expressed on a state scale with mean 0 and unit variance, the estimated mapping of $\hat{\mu}_{dygb}^{\text{state}}$ to the NAEP scale is given by Equation (6.2) below, where $\hat{\rho}_{sygb}^{\text{state}}$ is the estimated reliability of the state test. This mapping yields an estimate of the of the GSD average performance on the NAEP scale; denoted $\hat{\mu}_{dygb}^{\widehat{\text{naep}}}$.

$$\hat{\mu}_{drygb}^{\widehat{\text{naep}}} = \hat{\mu}_{sygb}^{\text{naep}} + \frac{\hat{\mu}_{drygb}^{\text{state}}}{\sqrt{\hat{\rho}_{sygb}^{\text{state}}}} \cdot \hat{\sigma}_{sygb}^{\text{naep}} \tag{6.2}$$

Likewise, the estimated mapping of $\hat{\sigma}_{dygb}^{\text{state}}$ to the NAEP scale is given by Equation (6.3).

$$\hat{\sigma}_{drygb}^{\widehat{\text{naep}}} = \left[\frac{\left(\hat{\sigma}_{drygb}^{\text{state}}\right)^2 + \hat{\rho}_{sygb}^{\text{state}} - 1}{\hat{\rho}_{sygb}^{\text{state}}}\right]^{1/2} \cdot \hat{\sigma}_{sygb}^{\text{naep}} \tag{6.3}$$

The intuition behind Equations (6.2) and (6.3) is straightforward: GSDs that belong to states with relatively high NAEP averages should be placed higher on the NAEP scale. Within states, GSDs that are high or low relative to their state (positive and negative on the standardized state scale) should be relatively high or low on the NAEP scale in proportion to that state's NAEP standard deviation.

The reliability term, $\hat{\rho}_{sygb}^{\text{state}}$, in Equations (6.2) and (6.3) is necessary to account for measurement error in state accountability test scores. Note that GSD means and standard deviations on the state scale, $\hat{\mu}_{drygb}^{\text{state}}$ and $\hat{\sigma}_{drygb}^{\text{state}}$, are expressed in terms of standard deviation units of the state score distribution. The standardized means are biased toward zero due to measurement error. They must be disattenuated before being mapped to the NAEP scale, given that the NAEP scale accounts for measurement error due to item sampling. We disattenuate the means by dividing them by the square root of the state test score reliability estimate, $\hat{\rho}_{sygb}^{\text{state}}$. The GSD standard deviations on the state scale, $\hat{\sigma}_{drygb}^{state}$, are biased toward 1 due to measurement error; we adjust them before linking them to the NAEP scale, as shown in Equation (6.3).

The reliability data used to disattenuate the estimates come from Reardon and Ho (2015) and were supplemented with publicly available information from state technical reports. For cases where no information was available, test reliabilities were imputed using data from other grades and years in the same state. We compute the standard errors of the linked estimates $\hat{\mu}_{drygb}^{\text{naep}}$ and $\hat{\sigma}_{drygb}^{\text{naep}}$ using the formulas described in Reardon, Kalogrides, and Ho (2017).

7. Scaling the Estimates

In order to make these NAEP-linked estimates ($\hat{\mu}_{drygb}^{\text{naep}}$ and $\hat{\sigma}_{drygb}^{\text{naep}}$) usefully interpretable, they are standardized in ways. The standardizations rely on estimates of the year-, grade-, and subject-specific means and standard deviations of the national student-level NAEP score distributions. For year y, grade g, and subject b, we denote these $\hat{\mu}_{ygb}^{\text{naep}}$ and $\hat{\sigma}_{ygb}^{\text{naep}}$, respectively. For years and grades when NAEP was administered, we estimate these from NAEP micro-data; for other years and grades, we estimate these via interpolation/extrapolation, using Equation (6.1) above. The equations shown here are in terms of the

GSD-subgroups; however, the methodology is equivalent for whole GSDs, counties, and county-subgroups.

Cohort Standardized (cs) Scale. For this scale we standardize the GSD means and standard deviations by dividing by the national grade-subject-specific standard deviation for a given cohort. We use the cohort that was in 4th grade in 2009 (and in 8th grade in 2013), as this is a cohort for whom NAEP data are available and that is roughly in the middle of our data. To do this, we compute:

$$\hat{\mu}_{drygb}^{cs} = \frac{\widehat{\hat{\mu}_{drygb}^{\text{naep}}} - \hat{\mu}_{[(y,g)^*]b}^{\text{naep}}}{\hat{\sigma}_{[(y,g)^*]b}^{\text{naep}}}, \text{for } (y,g)^* \text{ s.t. } y - g = 2005$$
$$\hat{\sigma}_{drygb}^{cs} = \frac{\widehat{\hat{\sigma}_{drygb}^{\text{naep}}}}{\hat{\sigma}_{[(y,g)^*]b}^{\text{naep}}}, \text{for } (y,g)^* \text{ s.t. } y - g = 2005 \tag{7.1}$$

This metric is interpretable as an effect size, relative to the grade-specific standard deviation of scores in one cohort. This has the advantage of being able to describe aggregated changed over time in test scores.

In this metric, GSD test score distributions are standardized relative to the estimated grade-specific national student-level distribution of scores of the cohort of students who were in 4th grade in 2009 (and 8th grade in 2013, assuming regular progress through grades). The scale compares a GSD's average achievement in a given grade and year to the national average in that grade in the year when a specific cohort was in that grade. This scale retains information about absolute changes over time by relying on the stability of the NAEP scale over time and on the linear interpolation of NAEP distributions over time. This scale does not enable absolute comparisons across grades, however.

Grade (within Cohort) Standardized (gcs) Scale. For this scale we standardize the GSD means and standard deviations by dividing by the average difference in NAEP scores between students one grade level apart. A one-unit difference in this grade-equivalent unit scale is interpretable as equivalent to the national average difference in skills between students one grade level apart in school. This scale is a simple linear transformation of the NAEP scale. To do this, we first estimate the within-cohort change in subject b, for the cohort of students in 4th grade in 2009, using estimates of the national NAEP means and standard deviations in grade 8 in 2013 and grade 4 in 2009. This is denoted $\hat{\gamma}_{2009b}$, e.g.,:

$$\hat{\gamma}_{2009b} = \frac{\hat{\mu}_{2013,8b}^{\text{naep}} - \hat{\mu}_{2009,4b}^{\text{naep}}}{4} \tag{7.2}$$

We then identify the linear transformation that sets these grade 4 and 8 averages at the "grade level" values 4 and 8 respectively, and transform all other GSD scores accordingly:

$$\hat{\mu}_{drygb}^{gcs} = 4 + \frac{\widehat{\hat{\mu}_{drygb}^{\text{naep}}} - \hat{\mu}_{2009,4b}^{naep}}{\hat{\gamma}_{2009b}},$$
$$\hat{\sigma}_{drygb}^{gcs} = \frac{\widehat{\hat{\sigma}_{drygb}^{\text{naep}}}}{\hat{\gamma}_{c^*b}}. \tag{7.3}$$

On this basis, $\hat{\mu}^{gcs}_{drygb}$ can be interpreted as the estimated average national "grade-level performance" of students in GSD d, subgroup r, year y, grade g, and subject b. So if $\hat{\mu}^{gcs}_{dy4b} = 5$, 4th-grade students in GSD d and year y are one grade level ($\hat{\gamma}_{2009b}$) above the 4th grade 2009 national average ($\hat{\mu}^{naep}_{2009,4b}$) in performance on the tested subject b. This metric enables absolute comparisons across grades and over time, but it does so by relying not only on the fact that the NAEP scale is stable over time and is vertically linked across grades 4 and 8, but also on the linear interpolation of NAEP scores between grades and years. This metric is a simple linear transformation of the NAEP scale, intended to render the NAEP scale more interpretable. As such, this metric is useful for descriptive research to broad audiences not familiar with interpreting standard deviation units, but may not be appropriate in all statistical analyses. For statistical analyses that do not require a vertically-linked scale, the cohort-standardized scale is more appropriate. The standardization methods and interpretation of the different scales is described in more detail in Reardon, Kalogrides and Ho (2017).

In total, we produce the following estimates: grade-year-subject-specific estimated means and standard deviations for each GSD ($\hat{\mu}^{x}_{dygb}$ and $\hat{\sigma}^{x}_{dygb}$), GSD-subgroup ($\hat{\mu}^{x}_{drygb}$ and $\hat{\sigma}^{x}_{drygb}$), county ($\hat{\mu}^{x}_{cygb}$ and $\hat{\sigma}^{x}_{cygb}$), and county-subgroup ($\hat{\mu}^{x}_{crygb}$ and $\hat{\sigma}^{x}_{crygb}$), where x denotes a particular standardization: cohort standardization (cs) or grade standardization (gs).

8. Suppressing Data Post-Estimation

Post-estimation suppression is conducted in four stages: (A) removing GSD- and county-subject-grade-year cases where participation is less than 95%; (B) removing GSD- and county-subgroup-subject-grade-year cases where participation is less than 95%; (C) removing GSD- and county-subgroup-subject-grade-year cases where the number of test scores reported is less than 95% of the total reported test scores; and (D) removing GSD and county-subgroup-subject-grade-year estimates where the standard errors are greater than 2.

A. Removing GSD- and county-subject-grade-year cases where participation is less than 95%. We retain as much data as possible in the estimation because we need population data to recover the statewide distribution for linking to NAEP. However, we do not report estimates for cases with low participation because they may be biased (i.e., the population of tested students on which the mean and standard deviation estimates are based may not be representative of the population of students in that school). Therefore, we remove all GSD and county-subject-grade-year cases where participation was lower than 95%. By this rule, we remove the estimates for all students and all student subgroups (e.g., race) where the overall GSD or county participation in that subject, grade and year is less than 95%, where participation is defined as:

$$\widehat{part}_{dygb} = \frac{numscores_{dygb}}{numenrl_{dygb}}. \quad (8.1)$$

Note: We do not suppress any entire GSD or county-subject-grade-year cases with this rule prior to the 2012-13 school year as enrollment data is not available.

B. Removing GSD and county-subgroup-subject-grade-year cases where participation is less than 95%. For the same reasons outlined in 8A, we also remove individual GSD- and county-subgroup-subject-grade-year observations using their respective participation rates:

$$\widehat{part}_{drygb} = \frac{numscores_{drygb}}{numenrl_{drygb}}. \qquad (8.1)$$

Note: We also do not suppress any entire GSD or county-subject-grade-year cases with this rule prior to the 2012-13 school year as enrollment data is not available.

C. Removing GSD- and county-subgroup-subject-grade-year cases where the number of test scores reported is less than 95% of the total reported test scores. In addition to suppressing GSD- and county-subgroup-subject-grade-year estimates based on participation, we also suppress data based on whether the total number of test scores reported by race or gender is less than 95% of the total reported test scores for all students. For example, there may be 50 test scores reported for all students, but only 20 test scores for white students, 20 test scores for black students, and no test scores for other racial subgroups. In this case, we would not report the white or black test score means because insufficient test scores were reported by race. We calculate the reported percentage as:

$$\widehat{rep}_{drygb} = \frac{\sum_r numscores_{drygb}}{numscores_{dygb}}. \qquad (8.2)$$

We have this last measure in all years. In the early years (2008-09 through 2011-12), this is the only measure we use to suppress entire GSD- and county-subgroup-subject-grade-year cases. In the later years, we use this and the GSD- and county-subgroup-subject-grade-year specific participation measure described above.

D. Standard errors. For all years, we suppress any estimate with an estimated standard error greater than 2 in the state-standardized metric (the estimates produced in Step **4** above). Any individual estimate with such a large standard error is too imprecise to use in analysis.

Appendix Table A3 summarizes the number of cases removed by the four decision rules at the GSD and county-levels.

9. Calculating Achievement Gaps

In addition to the mean and standard deviation estimates, we provide achievement gap estimates SEDA 2.0, estimated as the difference in average achievement between subgroups. These estimates are derived from the GSD- or county-subgroup-subject-grade-year noisy, linked, and scaled means and their standard errors. We provide three types of achievement gaps in the current data: white-Black (wbg), white-Hispanic (whg), and white-Asian (wag). Future updates to the data will include achievement gaps between additional subgroups. Each gap is computed by calculating the difference in the order appearing in the label; for example, the White-Black gap is calculated as $mean(white) - mean(black)$.

More specifically, in each scale, the GSD-subject-grade-year gap is given by the difference in the means, e.g., the white-black gap is given by:

$$\widehat{wbg}^x_{dygb} = \hat{\mu}^x_{d(r=wht)ygb} - \hat{\mu}^x_{d(r=blk)ygb} \tag{9.1}$$

where x denotes a particular standardization (cohort standardization, grade standardization) described in Step 7 above. The standard error of the gap is given by:

$$se\left(\widehat{wbg}^x_{dygb}\right) = \sqrt{se\left(\hat{\mu}^x_{d(r=wht)ygb}\right)^2 + se\left(\hat{\mu}^x_{d(r=blk)ygb}\right)^2} \tag{9.2}$$

Note that gap estimates will only exist when both subgroups have an estimated mean, i.e. within the GSD there are at least 20 students in each subgroup. For example, to get a white-black gap for a given GSD-subject-grade-year, there must be at least 20 white students and 20 black students in that GSD-subject-grade-year. The gaps can be interpreted similarly to the means in the units defined by the scales described in Step 7. Note that the methodology for estimating achievement gaps in SEDA 2.0 differs from the methodology used in prior versions of SEDA, which used the V-statistic to estimate achievement gaps.

10. Pooling Mean, Standard Deviation, and Gap Estimates

SEDA 2.0 provides grade-year-subject-specific estimated means and standard deviations for each GSD ($\hat{\mu}^x_{dygb}$ and $\hat{\sigma}^x_{dygb}$), GSD-subgroup ($\hat{\mu}^x_{drygb}$ and $\hat{\sigma}^x_{drygb}$), county ($\hat{\mu}^x_{cygb}$ and $\hat{\sigma}^x_{cygb}$), and county-subgroup ($\hat{\mu}^x_{crygb}$ and $\hat{\sigma}^x_{crygb}$), where x denotes a particular standardization (cohort standardization, grade standardization) described in Section 7 above.

For each geographic unit (GSD or county) by subgroup (all students, white students, black students, Asian students), we have up to 42 grade-year estimates (7 years times 6 grades) per subject. For some analyses it is useful to pool these estimates in order to provide more precise estimates of average scores (or changes in scores across grades or years) within each unit. We pool the estimates within a GSD using precision-weighted random-coefficient models.

Pooling Model Specifications

Subject-specific estimates. The models allow each unit (GSD, GSD-subgroup, county, county-subgroup) to have a unit-subject-specific intercept (average score), a unit subject-specific linear grade slope (rate at which scores change across grades, within a cohort), and a unit subject-specific cohort trend (the rate at which scores change across student cohorts, within a grade). For each parameter y (μ or σ or a gap), and standardization x (cs and gcs), we fit the following model:

$$
\begin{aligned}
\hat{y}^{x}_{drygb} = {} & \left[\beta_{0md} + \beta_{1md}\left(cohort_{drygb} - 2006.5\right) \right. \\
& \left. + \beta_{2md}\left(grade_{drygb} - 5.5\right)\right] M_b \\
& + \left[\beta_{0ed} + \beta_{1ed}\left(cohort_{drygb} - 2006.5\right) \right. \\
& \left. + \beta_{2ed}\left(grade_{drygb} - 5.5\right)\right] E_b + u_{drygb} + e_{drygb} \\
\beta_{0md} &= \gamma_{0m0} + v_{0md} \\
\beta_{1md} &= \gamma_{1m0} + v_{1md} \\
\beta_{2md} &= \gamma_{2m0} + v_{2md} \\
\beta_{0ed} &= \gamma_{0e0} + v_{0ed} \\
\beta_{1ed} &= \gamma_{1e0} + v_{1ed} \\
\beta_{2ed} &= \gamma_{2e0} + v_{2ed} \\
e_{dygb} & \sim N\left(0, \omega^2_{dygb}\right);\ u_{dygb} \sim N(0, \sigma^2); \begin{bmatrix} v_{0md} \\ \vdots \\ v_{2ed} \end{bmatrix} \sim MVN(0, \boldsymbol{\tau}^2).
\end{aligned} \tag{10.1}
$$

In this model, M_b is an indicator variable equal to 1 if the subject is math and E_b is an indicator variable equal to 1 if the subject is ELA. β_{0bd} represents the mean test score in subject b, in unit d, in grade 5.5 for cohort 2006.5 ($cohort$ is defined as $year - grade$, so this pseudo-cohort and pseudo-grade represents the center of our data's grade and cohort ranges, since the middle year is 2012 and the middle grade is 5.5). The β_{1bd} parameter indicates the average within-grade (cohort-to-cohort) change per year in average test scores in unit d in subject b; and, the β_{2bd} indicates the average within-cohort change per grade in average test scores in unit d in subject b.

If the model is fit using one of the scales that standardizes scores within grades (the cs scale), the coefficients will be interpretable in NAEP student-level standard deviation units (relative to the specific standard deviation used to standardize the scale). Between-unit differences in β_{0bd}, β_{1bd}, and β_{2bd} will be interpretable relative to this same scale. If the model is fit using the grade-level scale (gcs), the coefficients will be interpretable as test score differences relative to the average between-grade difference among students.

Overall estimates. SEDA 2.0 also provides estimates pooled across grades, years, and subjects. This model is as follows:

$$
\begin{aligned}
\hat{y}^{x}_{dygb} = {} & \beta_{0d} + \beta_{1d}\left(cohort_{dygb} - 2006.5\right) + \beta_{2d}\left(grade_{dygb} - 5.5\right) \\
& + \beta_{3d}(M_b - .5) + u_{dygb} + e_{dygb} \\
\beta_{0d} &= \gamma_{00} + v_{0d} \\
\beta_{1d} &= \gamma_{10} + v_{1d} \\
\beta_{2d} &= \gamma_{20} + v_{2d} \\
\beta_{3d} &= \gamma_{30} + v_{3d} \\
e_{dygb} & \sim N\left(0, \omega^2_{dygb}\right);\ u_{dygb} \sim N(0, \sigma^2); \begin{bmatrix} v_{0d} \\ v_{1d} \\ v_{2d} \\ v_{3d} \end{bmatrix} \sim MVN(0, \boldsymbol{\tau}^2).
\end{aligned} \tag{10.2}
$$

This model allows each unit to have a unit-specific intercept (average score, pooled over subjects), a unit-specific linear grade slope (rate at which scores change across grades, within a cohort, pooled over subjects), and a unit-specific cohort trend (the rate at which scores change across student cohorts, within a grade, pooled over subjects), and a unit-specific math-ELA difference.

In Appendix C, we report the reliabilities and the variance and covariance terms from the estimated $\boldsymbol{\tau}^2$ matrices from these pooling models.

NOTES ON USING POOLED MEAN & SD ESTIMATES

SEDA 2.0 contains two sets of estimates derived from the pooling models described in Equations (10.1) and (10.2). First are what we refer to as the OLS estimates of $\beta_{0d}, \dots, \beta_{3d}$. Second are the Empirical Bayes (EB) shrunken estimates of $\beta_{0d}, \dots, \beta_{3d}$.

Note that SEDA 2.0 does not contain estimates of β_{1d}, the cohort slope from model (10.1) or (10.2). Those estimates will be included in a later SEDA release.

The OLS estimates are the estimates of $\beta_{0d}, \dots, \beta_{3d}$ that we would get if we took the fitted values from Model (10.1) or (10.2) and added in the residuals $v_{0d}, \dots, v_{3d}$. That is $\hat{\beta}_{0d}^{ols} = \hat{\gamma}_{00} + \hat{v}_{0d}$, for example. These estimates are unbiased estimates of $\beta_{0d}, \dots, \beta_{3d}$, but they may be noisy in small GSDs. We obtain standard errors of these as described in Appendix B-3.

In the interest of discouraging the over-interpretation of imprecisely estimated parameters, SEDA 2.0 does not report estimates of the β_d's with OLS reliability below 0.7. We compute the reliability of OLS estimate $\hat{\beta}_{kd}^{ols}$ as $\frac{\hat{\tau}_k^2}{\hat{\tau}_k^2 + \hat{V}_{kd}}$, where $\hat{\tau}_k^2$ is the k^{th} diagonal element of the estimated $\boldsymbol{\tau}^2$ matrix (the estimated true variance of β_{kd}) and $\hat{V}_{kd}$ is the square of the estimated standard error of $\hat{\beta}_{kd}^{ols}$. That is, we do not report $\hat{\beta}_{kd}^{ols}$ if $\hat{V}_{kd} > \frac{3}{7}\hat{\tau}_k^2$. Users who wish to obtain parameter estimates with lower reliability can obtain them by fitting model (10.1) or (10.2) themselves. For subgroups, we use the same procedure; however, we use the standard error threshold determined for all students to censor estimates (rather than calculate a subgroup-specific threshold).

The EB estimates are based on the fitted model as well, but they include the EB shrunken residual. That is, $\hat{\beta}_{0d}^{eb} = \hat{\gamma}_{00} + \hat{v}_{0d}^{eb}$, for example, where $\hat{v}_{0d}^{eb}$ is the EB residual from the fitted model. The EB estimates are biased toward $\hat{\gamma}_{00}$, but have statistical properties that make them suited for inclusion as predictor variables or when one is interested in identifying outlier GSDs. We report the square root of the posterior variance of the EB estimates as the standard error of the EB estimate.

NOTES ON WHEN TO USE OLS OR EB ESTIMATES

In general, the EB estimates should be used for descriptive purposes and as predictor variables on the right-hand side of a regression model. They should not be used as outcome variables in a regression model. Doing so may lead to biased parameter estimates in fitted regression models. The OLS estimates are appropriate for use as outcome variables in a regression model. When using the OLS estimates as outcome variables, we recommend fitting precision-weighted models that account for the known error variance of the OLS estimates.

Notes on Using Pooled Gap Estimates

For users interested in analyzing achievement gaps in the pooled data, it is important to use the pooled gap estimates rather than taking the difference between pooled estimates of group-specific means. For example, the pooled white-black gap estimate in GSD d is obtained by 1) computing the gap (the difference in mean white and black scores) in each GSD-grade-year-subject; 2) fitting model 10.1 or 10.2 above using these gap estimates on the lefthand side; and 3) constructing $\hat{\beta}_{0d}^{ols}$ and $\hat{\beta}_{0d}^{eb}$ from the estimates. This is the preferred method of computing the average gap in GSD d. The alternative approach (taking the difference of pooled white and black mean scores) will not yield the same estimates. That is, the approach above will not yield identical estimates of pooled gaps as: 1) fitting model 10.1 or 10.2 above using the white mean estimates on the left-hand side; 2) constructing $\hat{\beta}_{0dw}^{ols}$ and $\hat{\beta}_{0dw}^{eb}$ for white students from the estimates; 3) doing the same with black student mean scores to construct $\hat{\beta}_{0db}^{ols}$ and $\hat{\beta}_{0db}^{eb}$ for black students; and then 4) estimating gaps by subtracting $\hat{\beta}_{0dw}^{ols} - \hat{\beta}_{0db}^{ols}$ and $\hat{\beta}_{0dw}^{eb} - \hat{\beta}_{0db}^{eb}$. In particular, the EB shrunken mean of the gaps is not in general equal to the difference in the EB shrunken means. The former is preferred. Practically speaking, this means that users interested in the pooled gap in a GSD or county should use the gap estimates reported in the pooled data files, rather than taking the difference between the estimated pooled means in the files.

COVARIATE DATA CONSTRUCTION

SEDA 2.0 contains CCD and EDGE/ACS data that have been curated for use with the GSD-level achievement data. These data include raw measures as well derived measures (e.g., a composite socioeconomic status measure, segregation measures), and CCD data are imputed to reduce missingness in some years. The composite construction and imputation are described in detail in the following sections.

SES Composite Construction

We use the EDGE/ACS data to compute a composite measure of the SES of each GSD. This measure is computed as the first principal component score of the following measures (each standardized): median income, percent of adults ages 25 and older with a bachelor's degree or higher, poverty rate for households with children ages 5-17, SNAP receipt rate, single mother headed household rate, and employment rate for adults ages 25-64. We use the base 2 logarithm of median income in these computations. We calculate the component loadings by conducting the analysis at the GSD level and weighting by GSD enrollment. We then use the loadings from this principal component analysis to calculate SES composite values for subgroups within GSDs.

Table 5 shows the component loadings for the socioeconomic status composite as well as the mean and standard deviation of each measure it includes. The "standardized loadings" indicate the coefficients used to compute the overall GSD SES composite score from the 6 standardized indicator variables, resulting in an SES composite that has an enrollment-weighted mean of 0 and standard deviation of 1 across all GSDs. The "unstandardized loadings" are re-scaled versions of the coefficients that are used to construct an SES composite score from the raw (unstandardized) indicator variables, but which is on the same scale as the standardized SES composite scores. Also reported are the correlations between each of the 6 indicators and the SES composite measure and the enrollment-weighted mean and standard deviation of the 6 indicators across GSDs.

Table 5. Component Loadings and Summary Statistics for Socioeconomic Status Composite Construction.

	Standardized Loadings	Unstandardized Loadings	Correlation between Indicators and Composite	Mean	SD
Median Income	NA	NA	NA	$62,509	$26,565
Log2 of Median Income	0.22	0.40	0.96	15.82	0.56
Proportion of Adults, Aged 25+ with a Bachelor's Degree or Higher	0.16	1.08	0.67	0.29	0.14
Poverty Rate, Households with 5-17 Year Olds	-0.22	-2.21	-0.94	0.16	0.10
Unemployment Rate	-0.16	-8.97	-0.68	0.05	0.02
Proportion of Households Receiving Food Stamps or SNAP	-0.22	-2.01	-0.93	0.16	0.11
Proportion Single Mother Headed Households	-0.19	-1.75	-0.83	0.27	0.11

Based on 11,582 districts in ACS with non-missing data on all 6 measures and average per grade enrollment. All values are weighted by district enrollment. The log (in base 2) of median income is used in the construction of the SES composite, but the mean and standard deviation of median income are also shown above for ease of interpretation. Unstandardized component loadings are coefficients from a model that regresses the (standardized) SES composite on the (unstandardized) six variables used to construct the composite; the intercept from this model is -5.06. These coefficients and intercept are used to construct race-specific SES composites.

To provide context for interpreting values of the SES composite, Table 6 reports average values of the indicator variables at different values of the SES composite.

Table 6. Component Loadings and Summary Statistics for Socioeconomic Status Composite Construction.

	SES Composite					
	-3	-2	-1	0	1	2
Median Family Income	$24,038	$31,026	$39,634	$53,029	$78,644	$136,804
% With BA or Higher	13.5%	14.9%	14.6%	18.3%	32.3%	62.4%
Poverty Rate	48.0%	37.6%	25.9%	14.7%	6.0%	1.6%
SNAP Eligibility Rate	50.0%	39.9%	27.6%	15.5%	5.6%	0.2%
Unemployment Rate	10.5%	8.0%	6.0%	4.5%	3.4%	2.6%
Single Parent Family Rate	51.9%	41.9%	31.7%	22.2%	14.6%	10.0%

COMMON CORE OF DATA IMPUTATION

School-level data from the CCD are available from 1987 until 2015. There is some missing data on racial composition and free/reduced price lunch receipt for some schools in some years. We therefore impute missing data on race/ethnicity and free/reduced priced lunch counts at the school level prior to aggregating data to the GSD level. The imputation model includes school-level data from the 1991-92 through 2014-15 school years and measures of total enrollment, enrollments by race (black, Hispanic, white, Asian, and Native American), enrollments by free and reduced priced lunch receipt (note that reduced priced lunch is only available in 1998 and later), an indicator for whether the school is located in an urban area, and state fixed effects. To improve the imputation of free and reduced priced lunch in more recent years we also use the proportion of students at each school that are classified as economically disadvantaged in the ED*Facts* data for 2008-09 through 2014-15 in the imputation model. Different states use different definitions of economically disadvantaged but these measures are highly correlated with free lunch rates from the CCD (r=.90). The imputations are estimated using predictive mean matching in Stata's –**mi impute chained–** routine, which fills in missing values iteratively by using chained equations. The idea behind this method is to impute variables iteratively using a sequence of univariate imputation models, one for each imputation variable, with all variables except the one being included in the prediction equation on the right hand side. This method is flexible for imputing data of different types. For more information, see: https://www.stata.com/manuals13/mi.pdf.

Prior to the imputation, we make three changes to the reported raw CCD data. First, for states with especially high levels of missing free and reduced price lunch data in recent years, we searched state department of education websites for alternative sources of data. We were only able to locate the appropriate data for Oregon and Ohio. For these states we replace CCD counts of free and reduced price lunch receipt with the counts reported in state department of education data for 2008-09 through 2014-15. In Ohio, 8% of schools were missing CCD free lunch data in 4 or more of the 7 ED*Facts* years. In Oregon, 5% of schools were missing CCD free lunch data in 4 or more of the 7 ED*Facts* years. Other states with high rates of missing free lunch data in the CCD during the ED*Facts* years are Alaska, Arizona, Montana, Texas, and Idaho. Unfortunately, we were unable to locate alternative data sources for these states, and rely on the imputation model to fill in missing data.

Second, starting in the 2011-12 school year some states began using community eligibility for the delivery of school meals whereby all students attending schools in low-income areas would have access to free meals regardless of their individual household income. Free lunch counts in schools in the community eligibility program are not reported in the same way nation-wide in the CCD. In community eligible schools, some schools report that all of their students are eligible for free lunch while others report counts that are presumably based on the individual student-level eligibility. Because reported free lunch eligible rates of 100 percent in community eligible schools may not accurately reflect the number of children from poor families in the school, we impute free lunch eligible rates in these schools. We replace free and reduced priced lunch counts as equal to missing if the school is a community eligible program school in a given year and their reported CCD free lunch rate is 100 percent. We then impute their free lunch eligible rate as described above.

Third, and finally, prior to imputation we replaced free and reduced price lunch counts as missing if the count was equal to 0. Anomalies in the CCD data led some cases to be reported as zeros when they should have been missing so we preferred to delete these 0 values and impute them using other years of data from that school.

The structure of the data prior to imputation is wide – that is, there is one variable for each year for any given measure (i.e., total enrollment 1991, total enrollment 1992, total enrollment 1993,…, total enrollment 2014) for all the measures described above. The exception are time invariant measures –

urbanicity and state. We impute 6 datasets and use the average of the 6 imputed values for each school in each year.

VERSIONING AND PUBLICATION

New or revised data will be posted periodically to the SEDA website. If you indicate that you would like to be notified about new postings when filling out the data use agreement, you will receive an email notifying you of any updates.

SEDA updates that contain substantially new information are labeled as a new version (e.g. V1.0, V2.0). Updates that make corrections or minor revisions to previously posted data are labeled as a subsidiary of the current version (e.g. V1.1, V1.2, etc.). When citing any SEDA data set for presentation, publication or use in the field, please include the version number in the citation. All versions of the data will remain archived and available on the SEDA website to facilitate data verification and research replication.

SEDA 2.0 makes the following additions and modifications to SEDA 1.1:

- We include 2 additional years of data.
- We include county-level estimates in addition to GSD-level estimates.
- We include estimated standard deviations as well as means.
- We include OLS and EB estimates in pooled data files.
- We include estimates of means and standard deviations by race/ethnicity as well as for all students.
- We estimate achievement gaps as differences in means rather than as V statistics.
- We altered the PHOP constraints.
- We make minor corrections to the crosswalk file.
- We used imputation to provide better estimates of free lunch and variables derived from free lunch.
- We suppress some cases we did not in the prior release (e.g., CO in 2009-11, TX and VA 7th and 8th grade math in some years).
- We suppress estimates where participation <95% (in 2012/13-2014/15).
- We suppress very imprecise estimates (in long data, when SE>2; in pooled data when reliability <.7).
- Minor data cleaning, including suppressing a small number of data errors.

DATA USE AGREEMENT

You agree not to use the data sets for commercial advantage, or in the course of for-profit activities. Commercial entities wishing to use this Service should contact Stanford University's Office of Technology Licensing (info@otlmail.stanford.edu).

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 STANFORD MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED. STANFORD 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 State of California of the United States of America, applicable to agreements negotiated, executed, and performed within California.

You agree to acknowledge the Stanford Education Data Archive as the source of these data. In publications, please cite the data as:

Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & DiSalvo, R. (2017). Stanford Education Data Archive (Version 2.0). Retrieved from http://purl.stanford.edu/db586ns4974.

Subject to your compliance with the terms and conditions set forth in this Agreement, Stanford grants you a revocable, non-exclusive, non-transferable right to access and make use of the Data Sets.

ERRATA

December 8, 2017: There was an error in the pooled district and county racial achievement gap data in the GCS scale that was released as part of SEDA Version 2.0 on December 5, 2017. This error was contained in the Stata (.dta) and CSV (.csv) versions of following files:

- SEDA_geodist_poolsub_GCS_v20
- SEDA_geodist_pool_GCS_v20
- SEDA_county_poolsub_GCS_v20
- SEDA_county_pool_GCS_v20

In these files, all of the pooled achievement gap estimates were too large by a constant of 5.5. Specifically, the issue was with the mean estimates (variables starting with “mn_avg_”) for the subgroups: wbg (white-black gap), whb (white-Hispanic gap), and wag (white-Asian gap).

This error has been fixed and these eight files have been re-uploaded. When the files are downloaded, you will see the names have been changed (all eight updated files end in “b”) to ensure that users are downloading the corrected files:

- SEDA_geodist_poolsub_GCS_v20b
- SEDA_geodist_pool_GCS_v20b
- SEDA_county_poolsub_GCS_v20b
- SEDA_county_pool_GCS_v20b

Note that the documentation and tables on the website do not reflect the name change, and the files still are referred to by their original names.

REFERENCES

Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal "Proficiency" categories. *Journal of Educational and Behavioral Statistics*, *37*(4), 489–517. https://doi.org/10.3102/1076998611411918

Reardon, S. F., & Ho, A. D. (2015). Practical issues in estimating achievement gaps from coarsened data. *Journal of Educational and Behavioral Statistics*, *40*(2), 158–189. https://doi.org/10.3102/1076998615570944

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2017, June). *Linking U.S. school district test score distributions to a common scale*. Working Paper, Stanford Center for Education Policy Analysis. Retrieved from http://cepa.stanford.edu/content/linking-us-school-district-test-score-distributions-common-scale

Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*, 42(1), 3–45. https://doi.org/10.3102/1076998616666279

APPENDICES

APPENDIX A: MISSING DATA

TABLE A1. STATE-SUBJECT-YEAR-GRADE DATA NOT INCLUDED IN SEDA 2.0.

State Abbreviation	Reason for Missing	Cases missing (gyb)
CA	incomplete data due to pilot testing	2014: E 3-8; M 3-8
CA	math tests vary by course	2009: M 7-8; 2010: M 7-8; 2011: M 7-8; 2012: M 7-8; 2013: M 7-8; 2015: M 7-8
CO	insufficient data	2009: E 3-8; M 3-8; 2010: E 3-8; M 3-8; 2011: E 3-8; M 3-8
CO	participation below 0.95	2015: E 5-8; M 4-8
CT	incomplete data due to pilot testing	2014: E 3-8; M 3-8
DC	participation below 0.95	2015: E 8; M 8
FL	incomplete data due to pilot testing	2014: M 3-8
ID	incomplete data due to pilot testing	2014: E 3-8; M 3-8
KS	not in edfacts data	2014: E 3-8; M 3-8
MD	participation below 0.95	2014: E 3-4, 6-7; M 3-4, 6-7
ME	participation below 0.95	2015: E 7-8; M 6-8
MT	incomplete data due to pilot testing	2014: E 3-8; M 3-8; 2015: E 3-8; M 3-8
ND	other reasons	2015: E 3-5; M 3-5
ND	participation below 0.95	2015: E 6-8; M 7-8
NE	each district allowed to have their own test	2009: E 3-8; M 3-8; 2010: M 3-8
NH	participation below 0.95	2015: E 8; M 8
NJ	participation below 0.95	2015: E 3-8; M 3-8
NV	not in edfacts data	2015: E 3-8; M 3-8
NV	participation below 0.95	2014: E 3-8; M 3-8
NY	participation below 0.95	2014: E 6-8; M 6-8; 2015: E 3-8; M 3-8
OK	participation below 0.95	2013: M 8
OR	participation below 0.95	2014: E 3, 7-8; M 3-8
RI	participation below 0.95	2015: E 5-8; M 6-8
SD	incomplete data due to pilot testing	2014: E 3-8; M 3-8
TX	math tests vary by course	2012: M 7-8; 2013: M 7-8; 2014: M 7-8; 2015: M 7-8
VA	math tests vary by course	2009: M 7-8; 2010: M 7-8; 2011: M 7-8; 2012: M 7-8; 2013: M 7-8; 2014: M 7-8; 2015: M 7-8
WA	incomplete data due to pilot testing	2014: E 3-8; M 3-8
WA	participation below 0.95	2015: E 3-8; M 3-8
WV	other reasons	2014: M 3-7
WY	not in edfacts data	2010: E 3-8; M 3-8
WY	participation below 0.95	2013: M 3-8; 2014: E 3-8; M 3, 7-8

TABLE A2. INDIVIDUAL GSDS REMOVED PRIOR TO ESTIMATION.

District ID	District Name	State Abbreviation	Grade	Year	Subject
0200003	Lower Yukon School District	AK	3	2015	ela
0509750	Mena School District	AR	6	2009	ela
0509750	Mena School District	AR	6	2009	math
2201470	St. Helena Parish	LA	4	2010	ela
3910019	Marietta City	OH	7	2014	math

TABLE A3. REMOVED GSD SUMMARY STATISTICS.

Description of Dropped Cases	County Cases (cygbr)	Geo Dist Cases (dygbr)
Cases dropped because state participation < 95%	32,935 (4.8%)	153,700 (7.4%)
Cases dropped because participation of corresponding "all students" < 95% (or > 105%)	28,215 (4.1%)	131,728 (6.3%)
Cases dropped because participation of the case itself < 95% (or > 105%)	32,309 (4.7%)	142,107 (6.8%)
Cases dropped because subgroup category total is not within 5% of all students (we call this "representation" -- it applies for gender and race only)	686 (0.1%)	1,092 (0.1%)
Cases dropped because standard error > 2	13 (0.0%)	8 (0.0%)
Total cases dropped for any reason	*49,802 (7.2%)*	*217,702 (10.5%)*
Total cases not dropped	*643,015 (92.8%)*	*1,858,918 (89.5%)*
Total number of cases	*692,817 (100.0%)*	*2,076,620 (100.0%)*

Appendix B: Additional Detail on Statistical Methods

1. Fixed Cut Score Approach with HOMOP Model for Subgroups

This section briefly describes the approach used to estimate the subgroup means and standard deviations for state-grade-year-subjects in which there is only a single cut score.

First, we set the location of the estimates by fixing the single cut score to the value estimated in the GSD model for the appropriate state-grade-year-subject case. Next, using NAEP data, we calculate the ratio between the standard deviation of scores for subgroup r relative to the entire state as $\tilde{\sigma}_r = \hat{\sigma}_r^{NAEP} / \hat{\sigma}_T^{NAEP}$, where $\hat{\sigma}_r^{NAEP}$ and $\hat{\sigma}_T^{NAEP}$ are the NAEP estimates of the population standard deviation for subgroup r and for the total population in a given state-grade-year-subject. We use linear interpolation to get these values in non-NAEP years and grades. In cases where NAEP does not report a standard deviation by subgroup, we assume this ratio is 1.

Denote the single fixed cut score from the full GSD model as $\hat{c}_{gyb}^{state}$. We fit the HOMOP model with this fixed cutscore, constraining all groups to have standard deviation equal to 1, and obtain the estimated matrices $\widehat{\mathbf{M}}_{rdygb}^{raw}$, $\widehat{\mathbf{\Sigma}}_{rdygb}^{raw} = \mathbf{1}$, $\widehat{\mathbf{V}}_{rdygb}^{raw}$, and $\widehat{\mathbf{W}}_{rdygb}^{raw} = \mathbf{0}$. These are the estimated parameters in the metric in which c is set to $\hat{c}_{gyb}^{state}$ and $\hat{\sigma}_{rdygb}^{raw} = 1$ for all GSDs. We then transform these to a metric in which the population standard deviation with all GSD r subgroups pooled together is equal to $\tilde{\sigma}_r$ and the cut score is at $\hat{c}_{gyb}^{state}$ by computing:

$$\begin{aligned} \widehat{\mathbf{M}}_{rdygb}^{state} &= \hat{c}_{gyb}^{state}\mathbf{1} + \frac{\tilde{\sigma}_r}{\hat{\sigma}'_r}\left(\widehat{\mathbf{M}}_{rdygb}^{raw} - \hat{c}_{gyb}^{state}\mathbf{1}\right) \\ \widehat{\mathbf{\Sigma}}_{rdygb}^{state} &= \frac{\tilde{\sigma}_r}{\hat{\sigma}'_r}\mathbf{1} \end{aligned} \qquad \text{(B-1.1)}$$

where

$$\begin{aligned} \hat{\sigma}'_r &= \sqrt{\mathbf{P}\left(\mathbf{\Pi}\widehat{\mathbf{M}}_{rdygb}^{raw}\right)^{\circ 2} + \mathbf{Q1}} \\ \mathbf{Q} &= \mathbf{P} + \left[\mathbf{v} \circ \left(\mathbf{P^2} - \mathbf{P}\right)\right] \\ \mathbf{\Pi} &= \mathbf{I} - \mathbf{1}^t\mathbf{P} \end{aligned} \qquad \text{(B-1.2)}$$

and $\mathbf{1}$ is a vector of 1's and $\mathbf{P}$ is a vector of the proportion of all students in subgroup r that are in each GSD.

2. Estimating County-Level Means and Standard Deviations

This section briefly describes how means, standard deviations, and standard errors are estimated for counties. As described above, we first estimate GSD-level means and standard deviations. We then estimate the county means as weighted averages of the GSD means and the county standard deviations as estimates of total variance within a county based on the GSD means and standard deviations.

For each state-subject-grade-year case we start with $\widehat{\mathbf{M}}_{ygb}^{state}$, $\widehat{\mathbf{\Sigma}}_{ygb}^{state}$, $\widehat{\mathbf{V}}_{ygb}^{state}$, $\widehat{\mathbf{W}}_{ygb}^{state}$, and $\widehat{\mathbf{Z}}_{ygb}^{state}$ the vectors of GSD estimates and their sampling covariances for each state-year-grade-subject. In what follows we will refer to these vectors of GSD-level estimates as $\widehat{\mathbf{M}}_{D}^{state}$, $\widehat{\mathbf{\Sigma}}_{D}^{state}$, $\widehat{\mathbf{V}}_{D}^{state}$, $\widehat{\mathbf{W}}_{D}^{state}$, and $\widehat{\mathbf{Z}}_{D}^{state}$ and will refer to the vectors of county-level estimates as $\widehat{\mathbf{M}}_{C}^{state}$, $\widehat{\mathbf{\Sigma}}_{C}^{state}$, $\widehat{\mathbf{V}}_{C}^{state}$, and $\widehat{\mathbf{W}}_{C}^{state}$, hence omitting the ygb subscripts for clarity, but noting that these calculations are carried out separately for each state-year-grade-subject case.

Define $\mathbf{X}$ as a CxD design matrix, where C is the number of counties and D is the total number of GSDs. Based on the definition of counties and GSDs, each county will contain at least one GSD and each GSD falls within exactly one county. Each row of $\mathbf{X}$ corresponds to a county. Each element of $\mathbf{X}$, x_{cd}, is equal to 1 if GSD d is within county c and 0 otherwise.

Let $\mathbf{P}_C$ be a $1 \times D$ vector of the county-specific GSD proportions (i.e., the proportion of each county represented by each GSD, such that $\mathbf{P}_C\mathbf{X}^t = \mathbf{1}_C$, a 1 by C vector of 1's). This can be computed as:

$$\mathbf{P}_C = \mathbf{n} * inverse\big(diag(\mathbf{n}\mathbf{X}^t\mathbf{X})\big), \qquad \text{(B-2.1)}$$

where $\mathbf{n}$ is a 1 by D vector of GSD sample sizes and $diag(\mathbf{y})$ is a square, diagonal matrix with the elements of the vector $\mathbf{y}$ on the diagonal.

The county mean can be estimated as a weighted mean of the GSD means. Hence we can estimate the county means as

$$\widehat{\mathbf{M}}_{C}^{state} = \mathbf{E}\widehat{\mathbf{M}}_{D}^{state}, \qquad \text{(B-2.2)}$$

where

$$\mathbf{E} = \mathbf{X}[diag(\mathbf{P}_C)]. \qquad \text{(B-2.3)}$$

We can calculate the variance-covariance matrix of the county means as:

$$\widehat{\mathbf{V}}_{C}^{state} = \mathbf{E}\widehat{\mathbf{V}}_{D}^{state}\mathbf{E}^t. \qquad \text{(B-2.4)}$$

The estimate of the total county standard deviation is a combination of the GSD standard deviations and the variability across GSD means within a county. Hence, we can estimate county standard deviation for a single county as

$$\hat{\sigma}_{c}^{state} = \sqrt{\hat{\sigma}_{B_c}^2 + \hat{\sigma}_{W_c}^2}, \qquad \text{(B-2.5)}$$

where $\hat{\sigma}_{B_c}^2$ is the estimated variance between GSDs in county c and $\hat{\sigma}_{W_c}^2$ is the average estimated variance within GSDs in county c. We estimate these county standard deviations as:

$$\widehat{\mathbf{\Sigma}}_{C}^{state} = \Big(\mathbf{E}\big([\mathbf{I}_G - \mathbf{X}^t\mathbf{E}]\widehat{\mathbf{M}}_{D}^{state}\big)^{\circ 2} + \mathbf{Q}_C\big(\widehat{\mathbf{\Sigma}}_{D}^{state}\big)^{\circ 2}\Big)^{\circ\frac{1}{2}}, \qquad \text{(B-2.6)}$$

where $\mathbf{I}_G$ is a G by G identity matrix, $\mathbf{Q}_C$ is defined below, and the notation $\mathbf{Y}^{\circ a}$ indicates raising each element of the matrix $\mathbf{Y}$ to the power a. These calculations are based on the derivations used in the appendices of Reardon et al. (2017) and account for sampling error in the estimates of $\widehat{\mathbf{M}}_D^{state}$ and $\widehat{\boldsymbol{\Sigma}}_D^{state}$. To calculate the variance-covariance matrix of the county standard deviation estimates we use:

$$\widehat{\mathbf{W}}_C^{state} = inverse\left(diag\left(\widehat{\boldsymbol{\Sigma}}_C^{state}\right)\right) * \left[\mathbf{EA}\widehat{\mathbf{V}}_D^{state}\mathbf{AE}^t + \mathbf{Q}_C\mathbf{B}\widehat{\mathbf{W}}_D^{state}\mathbf{BQ}_C^t + 2 * \mathbf{EA}\widehat{\mathbf{Z}}_D^{state}\mathbf{BQ}_C^t\right], \tag{B-2.7}$$

where

$$\mathbf{A} = diag\left(\widehat{\mathbf{M}}_D^{state}\right)$$
$$\mathbf{B} = diag\left(\widehat{\boldsymbol{\Sigma}}_D^{state}\right), \tag{B-2.8}$$

and

$$\mathbf{Q}_C = inverse\left(diag(\mathbf{1}_C + 2\boldsymbol{\omega}^t)\right) * \left(\mathbf{X} * inverse\left(diag(\mathbf{n})\right) \circ [\mathbf{E} + \mathbf{X} * diag(\mathbf{n} - \mathbf{1}_G)] \circ \mathbf{E}\right), \tag{B-2.9}$$

where $\boldsymbol{\omega}$ is the C by 1 vector of county-specific omega-bar terms, computed as

$$\boldsymbol{\omega} = (\mathbf{1}_D\mathbf{X}^t)^{\circ-1}\mathbf{X}(2(\mathbf{n} - \mathbf{1}_D)^t)^{\circ-1}, \tag{B-2.10}$$

and where $\mathbf{1}_D$ is a 1 by D vector of 1's. This is equivalent to Equation (A8) in Reardon et al. (2017):

$$\overline{\omega_d^2} = \frac{1}{D}\sum_{d=1}^{D}\frac{1}{2(n_d - 1)}. \tag{B-2.11}$$

Following the derivations in Reardon et al. (2017), when the GSD estimates are based on a HOMOP model, we use:

$$\boldsymbol{\omega}_{HOMOP} = [2(\mathbf{1}_D\mathbf{n}^t - \mathbf{1}_D\mathbf{1}_D^t)]^{\circ-1}\mathbf{1}_C^t. \tag{B-2.12}$$

When the GSD estimates are based on a PHOP model, we use:

$$\boldsymbol{\omega}_{PHOP} = (\mathbf{1}_D\mathbf{X}^t)^{\circ-1}\mathbf{X}(2(\widehat{\mathbf{n}} - \mathbf{1}_D)^t)^{\circ-1}, \tag{B-2.13}$$

where

$$\widehat{\mathbf{n}} = \mathbf{YY}^t(\mathbf{n} - \mathbf{1}_D)^t + diag(\mathbf{1}_D^t - \mathbf{Y})\mathbf{n}^t, \tag{B-2.14}$$

with $\mathbf{Y}$ a D by 1 indicator vector, with $1's$ representing groups with a constrained standard deviation and 0 for groups with freely estimated standard deviations.

3. Constructing OLS Standard Errors from Pooled Models

In the SEDA 2.0 data, we release the OLS and EB estimates of the intercept and grade slope, as well as their standard errors, from the pooled models described in Section 10. The recovery of the OLS SEs is not straightforward from HLM. In order to recover these, we perform the estimation in two steps and calculate the OLS SEs post-estimation.

The remainder of this section describes the method and computational implementation. The equations are written to correspond to the pooling model shown in equation 10.2; however, this procedure is the same for the other variant of our pooling models.

Step 1. We estimate σ^2 using the three-level model described in equation 10.2 and define:

$$\hat{\phi}^2_{drygb} = \hat{\sigma}^2 + \omega^2_{drygb} \qquad \text{(B-3.1)}$$

Where ω^2_{drygb} is the variance of the $\hat{y}^x_{drygb}$ estimate (either μ or σ) . We assume that $\hat{\sigma}^2$ is a very precise estimate because of the large amount of data in the model.

Step 2. We then reweight the data and estimate a two-level HLM model:

Level-1:

$$\hat{\phi}^{-1}_{drygb}\hat{y}^x_{drygb} = [\beta_{0d} \quad \beta_{1d} \quad \beta_{2d} \quad \beta_{3d}]\begin{bmatrix} \hat{\phi}^{-1}_{drygb} \\ \hat{\phi}^{-1}_{drygb}(cohort_{drygb} - 2006.5) \\ \hat{\phi}^{-1}_{drygb}(grade_{drygb} - 5.5) \\ \hat{\phi}^{-1}_{drygb}(math_{drygb} - .5) \end{bmatrix} + \hat{\phi}^{-1}_{drygb}e_{drygb}$$

Level-2: (B-3.2)

$$\beta_{0d} = \gamma_{00} + \nu_{0d}$$
$$\beta_{0d} = \gamma_{10} + \nu_{1d}$$
$$\beta_{0d} = \gamma_{20} + \nu_{2d}$$
$$\beta_{0d} = \gamma_{30} + \nu_{3d}$$

After estimation, the HLM residual file contains the OLS and EB estimates, as well as the posterior variance matrices, $\boldsymbol{V}^{EB}_d$, for each GSD. From the model, we also recover an estimate of $\boldsymbol{\tau}^2$. Using $\boldsymbol{V}^{EB}_d$ and $\hat{\boldsymbol{\tau}}^2$, we can calculate the standard errors of the OLS estimates for each GSD as the inverse of:

$$\left(\boldsymbol{V}^{OLS}_d\right)^{-1} = \left(\boldsymbol{V}^{EB}_d\right)^{-1} - \hat{\boldsymbol{\tau}}^{-2}. \qquad \text{(B-3.3)}$$

Appendix C: Pooling Model Results

Table C1. Variance and Covariance Estimates from Pooling Models.

Identifiers				Pooled			Math			ELA		
Geo	Mt	Mn/SD	Sub	tau(int)	tau(grd)	cov (int,grd)	tau(int)	tau(grd)	cov (int,grd)	tau(int)	tau(grd)	cov (int,grd)
Geo Dist	cs	mean	all	0.11995	0.00203	0.00214	0.13147	0.00310	0.00325	0.11688	0.00170	0.00161
Geo Dist	cs	mean	asn	0.21410	0.00188	0.00880	0.23522	0.00328	0.01424	0.20360	0.00156	0.00582
Geo Dist	cs	mean	blk	0.06695	0.00185	0.00169	0.07311	0.00263	0.00234	0.06797	0.00154	0.00150
Geo Dist	cs	mean	hsp	0.07505	0.00206	-0.00057	0.07996	0.00302	0.00134	0.08102	0.00190	-0.00130
Geo Dist	cs	mean	wht	0.08090	0.00185	0.00239	0.09442	0.00285	0.00314	0.07526	0.00150	0.00202
Geo Dist	cs	mean	wag	0.09138	0.00088	0.00289	0.09955	0.00111	0.00344	0.08730	0.00080	0.00272
Geo Dist	cs	mean	wbg	0.04251	0.00048	0.00092	0.04347	0.00063	0.00143	0.04396	0.00041	0.00056
Geo Dist	cs	mean	whg	0.04311	0.00054	0.00001	0.04311	0.00052	0.00057	0.04634	0.00062	-0.00052
Geo Dist	cs	sd	all	0.00409	0.00017	-0.00012	0.00506	0.00026	-0.00005	0.00397	0.00019	-0.00018
County	cs	mean	all	0.06077	0.00126	0.00036	0.07100	0.00204	0.00099	0.05706	0.00105	0.00016
County	cs	mean	asn	0.14002	0.00143	0.00482	0.15184	0.00214	0.00802	0.13685	0.00133	0.00298
County	cs	mean	blk	0.03900	0.00134	0.00014	0.04581	0.00197	0.00082	0.03897	0.00116	0.00004
County	cs	mean	hsp	0.03851	0.00142	-0.00141	0.04369	0.00209	-0.00021	0.04188	0.00135	-0.00197
County	cs	mean	wht	0.04276	0.00121	0.00046	0.05460	0.00198	0.00113	0.03699	0.00099	0.00014
County	cs	mean	wag	0.09925	0.00097	0.00426	0.10489	0.00115	0.00504	0.09746	0.00094	0.00385
County	cs	mean	wbg	0.03735	0.00058	0.00114	0.04024	0.00079	0.00192	0.03670	0.00052	0.00058
County	cs	mean	whg	0.04010	0.00050	0.00078	0.04161	0.00054	0.00146	0.04199	0.00056	0.00010
County	cs	sd	all	0.00412	0.00014	0.00000	0.00516	0.00021	0.00008	0.00371	0.00016	-0.00006
Geo Dist	gcs	mean	all	1.06759	0.01920	0.03870	1.12417	0.03157	0.09264	1.11722	0.01619	-0.00419
Geo Dist	gcs	mean	asn	1.90253	0.02083	0.10881	2.02592	0.04573	0.23417	1.94286	0.01378	0.02219
Geo Dist	gcs	mean	blk	0.59350	0.01740	0.02595	0.62544	0.02534	0.05650	0.64920	0.01470	0.00335
Geo Dist	gcs	mean	hsp	0.65952	0.01847	0.00380	0.68323	0.02708	0.05207	0.77709	0.01904	-0.02665
Geo Dist	gcs	mean	wht	0.72083	0.01748	0.03613	0.80784	0.02859	0.07382	0.71890	0.01395	0.00659
Geo Dist	gcs	mean	wag	0.81045	0.00970	0.04005	0.85389	0.01484	0.07704	0.83378	0.00708	0.01218
Geo Dist	gcs	mean	wbg	0.37895	0.00457	0.01542	0.37228	0.00764	0.03324	0.42100	0.00392	-0.00171
Geo Dist	gcs	mean	whg	0.38245	0.00499	0.00688	0.36842	0.00593	0.02588	0.44444	0.00624	-0.01246
Geo Dist	gcs	sd	all	0.03508	0.00161	-0.00065	0.04311	0.00215	0.00223	0.03821	0.00189	-0.00242
County	gcs	mean	all	0.54235	0.01184	0.01484	0.60702	0.01943	0.04412	0.54703	0.01028	-0.00770
County	gcs	mean	asn	1.24759	0.01524	0.06537	1.30544	0.02878	0.14163	1.30656	0.01252	0.00674
County	gcs	mean	blk	0.34815	0.01241	0.00877	0.39148	0.01793	0.03010	0.37393	0.01138	-0.00597
County	gcs	mean	hsp	0.34102	0.01273	-0.00647	0.37147	0.01767	0.02075	0.40316	0.01388	-0.02624
County	gcs	mean	wht	0.38028	0.01132	0.01268	0.46716	0.01865	0.03720	0.35448	0.00955	-0.00481
County	gcs	mean	wag	0.88559	0.01073	0.05590	0.90195	0.01682	0.09340	0.93035	0.00815	0.02150
County	gcs	mean	wbg	0.33443	0.00558	0.01643	0.34530	0.00932	0.03599	0.35151	0.00488	-0.00042
County	gcs	mean	whg	0.35810	0.00487	0.01348	0.35713	0.00693	0.03285	0.40244	0.00542	-0.00588
County	gcs	sd	all	0.03550	0.00121	0.00062	0.04408	0.00196	0.00337	0.03565	0.00159	-0.00120

Abbreviations: Geo, Mt, Mn/SD, Sub = Geography, Metric, Mean/SD, Sugroup; geo dist = geographic district; cs = cohort scale; gcs = grade-cohort scale; wht, blk, hsp, asn = white, black, Hispanic, Asian; wag, wbg, whg = white-Asian gap, white-Black gap, white-Hispanic gap; tau = variance, rel = reliability

TABLE C2. ESTIMATED RELIABILITIES OF POOLED MODEL ESTIMATES BY GEOGRAPHIC UNIT, SCALE, ESTIMATE TYPE, AND SUBGROUP

Identifiers				Pooled		Math		ELA	
Geo	Mt	Mn/SD	Sub	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
Geo Dist	cs	mean	all	0.986	0.844	0.980	0.811	0.979	0.742
Geo Dist	cs	mean	asn	0.967	0.696	0.958	0.655	0.958	0.580
Geo Dist	cs	mean	blk	0.958	0.785	0.941	0.729	0.942	0.673
Geo Dist	cs	mean	hsp	0.956	0.769	0.940	0.721	0.943	0.680
Geo Dist	cs	mean	wht	0.978	0.820	0.970	0.787	0.966	0.701
Geo Dist	cs	mean	wag	0.943	0.564	0.925	0.444	0.925	0.446
Geo Dist	cs	mean	wbg	0.939	0.554	0.912	0.463	0.917	0.409
Geo Dist	cs	mean	whg	0.929	0.540	0.901	0.390	0.908	0.454
Geo Dist	cs	sd	all	0.973	0.782	0.959	0.742	0.954	0.709
County	cs	mean	all	0.995	0.897	0.991	0.877	0.990	0.814
County	cs	mean	asn	0.969	0.724	0.961	0.682	0.959	0.633
County	cs	mean	blk	0.967	0.820	0.954	0.773	0.952	0.717
County	cs	mean	hsp	0.950	0.784	0.936	0.747	0.935	0.694
County	cs	mean	wht	0.988	0.878	0.984	0.856	0.980	0.783
County	cs	mean	wag	0.961	0.674	0.951	0.580	0.949	0.578
County	cs	mean	wbg	0.963	0.696	0.947	0.620	0.947	0.565
County	cs	mean	whg	0.947	0.625	0.928	0.509	0.929	0.530
County	cs	sd	all	0.982	0.787	0.971	0.743	0.964	0.712
Geo Dist	gcs	mean	all	0.986	0.849	0.980	0.832	0.979	0.747
Geo Dist	gcs	mean	asn	0.967	0.720	0.959	0.740	0.958	0.570
Geo Dist	gcs	mean	blk	0.958	0.789	0.941	0.749	0.943	0.678
Geo Dist	gcs	mean	hsp	0.955	0.769	0.941	0.733	0.943	0.691
Geo Dist	gcs	mean	wht	0.978	0.825	0.970	0.808	0.967	0.703
Geo Dist	gcs	mean	wag	0.943	0.600	0.926	0.541	0.925	0.432
Geo Dist	gcs	mean	wbg	0.938	0.557	0.912	0.535	0.917	0.408
Geo Dist	gcs	mean	whg	0.928	0.538	0.902	0.453	0.908	0.463
Geo Dist	gcs	sd	all	0.950	0.680	0.959	0.744	0.955	0.719
County	gcs	mean	all	0.994	0.901	0.991	0.888	0.990	0.822
County	gcs	mean	asn	0.968	0.739	0.962	0.756	0.959	0.634
County	gcs	mean	blk	0.967	0.821	0.954	0.783	0.953	0.726
County	gcs	mean	hsp	0.950	0.784	0.937	0.751	0.936	0.709
County	gcs	mean	wht	0.988	0.880	0.984	0.868	0.980	0.790
County	gcs	mean	wag	0.961	0.704	0.952	0.679	0.949	0.561
County	gcs	mean	wbg	0.962	0.694	0.947	0.681	0.947	0.564
County	gcs	mean	whg	0.946	0.626	0.930	0.591	0.929	0.533
County	gcs	sd	all	0.970	0.682	0.971	0.758	0.965	0.720

Abbreviations: Geo, Mt, Mn/SD, Sub = Geography, Metric, Mean/SD, Sugroup; geo dist = geographic district; cs = cohort scale; gcs = grade-cohort scale; wht, blk, hsp, asn = white, black, Hispanic, Asian; wag, wbg, whg = white-Asian gap, white-Black gap, white-Hispanic gap; tau = variance, rel = reliability

Appendix D: Variables

Achievement Data

The tables below summarize the variables appearing in the "long," "poolsub," and "pool" SEDA 2.0 data files. Variable descriptions correspond to the CS scale reported for Geographic School Districts, but the variable names and file structures remain the same for other scales (GCS, State, or NAEP) and for the county-level estimates files.

Each long format file of estimates contains the following variables (variables shown here are from the CS scale file, but variables and structure remain the same for long files with alternative scales GCS, NAEP, and State):

Name	Label	Source
leaidC	NCES ID - Geographic School Districts	Identifier (ID) (common to multiple datasets)
leaname	District (LEA) Name	Common Core of Data (CCD)
fips	State FIPS Code	Identifier (ID) (common to multiple datasets)
stateabb	State Abbreviation	Identifier (ID) (common to multiple datasets)
grade	Tested Grade (g)	SEDA HETOP using EDFacts data
year	Spring of Tested Year (y)	SEDA HETOP using EDFacts data
subject	Tested Subject (b)	SEDA HETOP using EDFacts data
totgyb_all	Sample Size for All Estimates (number of tests in gyb)	SEDA HETOP using EDFacts data
mn_all	Geo Dist gyb Ach Mean, All Students, CS	SEDA HETOP using EDFacts data
mn_all_se	Geo Dist gyb SE of Ach Mean, All Students, CS	SEDA HETOP using EDFacts data
sd_all	Geo Dist gyb Ach SD, All Students, CS	SEDA HETOP using EDFacts data
sd_all_se	Geo Dist gyb SE of Ach SD, All Students, CS	SEDA HETOP using EDFacts data
totgyb_asn	Sample Size for Asian Estimates (number of tests in gyb)	SEDA HETOP using EDFacts data
mn_asn	Geo Dist gyb Ach Mean, Asian Students, CS	SEDA HETOP using EDFacts data
mn_asn_se	Geo Dist gyb SE of Ach Mean, Asian Students, CS	SEDA HETOP using EDFacts data
sd_asn	Geo Dist gyb Ach SD, Asian Students, CS	SEDA HETOP using EDFacts data

sd_asn_se	Geo Dist gyb SE of Ach SD, Asian Students, CS	SEDA HETOP using EDFacts data
totgyb_blk	Sample Size for Black Estimates (number of tests in gyb)	SEDA HETOP using EDFacts data
mn_blk	Geo Dist gyb Ach Mean, Black Students, CS	SEDA HETOP using EDFacts data
mn_blk_se	Geo Dist gyb SE of Ach Mean, Black Students, CS	SEDA HETOP using EDFacts data
sd_blk	Geo Dist gyb Ach SD, Black Students, CS	SEDA HETOP using EDFacts data
sd_blk_se	Geo Dist gyb SE of Ach SD, Black Students, CS	SEDA HETOP using EDFacts data
totgyb_hsp	Sample Size for Hispanic Estimates (number of tests in gyb)	SEDA HETOP using EDFacts data
mn_hsp	Geo Dist gyb Ach Mean, Hispanic Students, CS	SEDA HETOP using EDFacts data
mn_hsp_se	Geo Dist gyb SE of Ach Mean, Hispanic Students, CS	SEDA HETOP using EDFacts data
sd_hsp	Geo Dist gyb Ach SD, Hispanic Students, CS	SEDA HETOP using EDFacts data
sd_hsp_se	Geo Dist gyb SE of Ach SD, Hispanic Students, CS	SEDA HETOP using EDFacts data
totgyb_wht	Sample Size for White Estimates (number of tests in gyb)	SEDA HETOP using EDFacts data
mn_wht	Geo Dist gyb Ach Mean, White Students, CS	SEDA HETOP using EDFacts data
mn_wht_se	Geo Dist gyb SE of Ach Mean, White Students, CS	SEDA HETOP using EDFacts data
sd_wht	Geo Dist gyb Ach SD, White Students, CS	SEDA HETOP using EDFacts data
sd_wht_se	Geo Dist gyb SE of Ach SD, White Students, CS	SEDA HETOP using EDFacts data
mn_wag	Geo Dist gyb Estimated White-Asian Gap, CS	SEDA HETOP using EDFacts data
mn_wag_se	Geo Dist gyb SE of White-Asian Gap Estimate, CS	SEDA HETOP using EDFacts data
mn_wbg	Geo Dist gyb Estimated White-Black Gap, CS	SEDA HETOP using EDFacts data
mn_wbg_se	Geo Dist gyb SE of White-Black Gap Estimate, CS	SEDA HETOP using EDFacts data
mn_whg	Geo Dist gyb Estimated White-Hispanic Gap, CS	SEDA HETOP using EDFacts data
mn_whg_se	Geo Dist gyb SE of White-Hispanic Gap Estimate, CS	SEDA HETOP using EDFacts data

Each “poolsub” file of estimates, which pool across grades and years, contains the following variables (variables shown here are from the CS scale file, but variables and structure remain the same for the GCS “poolsub” file):

Name	Label	Source
leaidC	NCES ID - Geographic School Districts	Identifier (ID) (common to multiple datasets)
leaname	District (LEA) Name	Common Core of Data (CCD)
fips	FIPS State Code	Identifier (ID) (common to multiple datasets)
stateabb	State Abbreviation	Identifier (ID) (common to multiple datasets)
subgroup	Subgroup of estimates, or subgroups of gap estimates	HLM using SEDA HETOP est
gap_est	Row is a gap estimate (subgroup indicates which gap)	HLM using SEDA HETOP est
tot_asmts_ela	Total number of ELA tests for pooled estimates	HLM using SEDA HETOP est
tot_asmts_mth	Total number of math tests for pooled estimates	HLM using SEDA HETOP est
cellcount	Number of district-grade-year cases used in pooling	HLM using SEDA HETOP est
mn_avg_mth_ol	Geo Dist Mean Ach, Math, OLS est, CS	HLM using SEDA HETOP est
mn_avg_ela_ol	Geo Dist Mean Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
mn_grd_mth_ol	Geo Dist Grade Slope of Mean Ach, Math, OLS est, CS	HLM using SEDA HETOP est
mn_grd_ela_ol	Geo Dist Grade Slope of Mean Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
mn_avg_mth_eb	Geo Dist Mean Ach, Math, EB est, CS	HLM using SEDA HETOP est
mn_avg_ela_eb	Geo Dist Mean Ach, ELA, EB est, CS	HLM using SEDA HETOP est
mn_grd_mth_eb	Geo Dist Grade Slope of Mean Ach, Math, EB est, CS	HLM using SEDA HETOP est
mn_grd_ela_eb	Geo Dist Grade Slope of Mean Ach, ELA, EB est, CS	HLM using SEDA HETOP est
mn_avg_mth_eb_se	Geo Dist SE of Mean Ach, Math, EB est, CS	HLM using SEDA HETOP est
mn_avg_ela_eb_se	Geo Dist SE of Mean Ach, ELA, EB est, CS	HLM using SEDA HETOP est
mn_grd_mth_eb_se	Geo Dist SE of Grade Slope of Mean Ach, Math, EB est, CS	HLM using SEDA HETOP est
mn_grd_ela_eb_se	Geo Dist SE of Grade Slope of Mean Ach, ELA, EB est, CS	HLM using SEDA HETOP est
mn_avg_mth_ol_se	Geo Dist SE of Mean Ach, Math, OLS est, CS	HLM using SEDA HETOP est
mn_grd_mth_ol_se	Geo Dist SE of Grade Slope of Mean Ach, Math, OLS est, CS	HLM using SEDA HETOP est

mn_avg_ela_ol_se	Geo Dist SE of Mean Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
mn_grd_ela_ol_se	Geo Dist SE of Grade Slope of Mean Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
sd_avg_mth_ol	Geo Dist SD of Ach, Math, OLS est, CS	HLM using SEDA HETOP est
sd_avg_ela_ol	Geo Dist SD of Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
sd_grd_mth_ol	Geo Dist Grade Slope of SD of Ach, Math, OLS est, CS	HLM using SEDA HETOP est
sd_grd_ela_ol	Geo Dist Grade Slope of SD of Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
sd_avg_mth_eb	Geo Dist SD of Ach, Math, EB est, CS	HLM using SEDA HETOP est
sd_avg_ela_eb	Geo Dist SD of Ach, ELA, EB est, CS	HLM using SEDA HETOP est
sd_grd_mth_eb	Geo Dist Grade Slope of SD of Ach, Math, EB est, CS	HLM using SEDA HETOP est
sd_grd_ela_eb	Geo Dist Grade Slope of SD of Ach, ELA, EB est, CS	HLM using SEDA HETOP est
sd_avg_mth_eb_se	Geo Dist SE of SD of Ach, Math, EB est, CS	HLM using SEDA HETOP est
sd_avg_ela_eb_se	Geo Dist SE of SD of Ach, ELA, EB est, CS	HLM using SEDA HETOP est
sd_grd_mth_eb_se	Geo Dist SE of Grade Slope of SD of Ach, Math, EB est, CS	HLM using SEDA HETOP est
sd_grd_ela_eb_se	Geo Dist SE of Grade Slope of SD of Ach, ELA, EB est, CS	HLM using SEDA HETOP est
sd_avg_mth_ol_se	Geo Dist SE of SD of Ach, Math, OLS est, CS	HLM using SEDA HETOP est
sd_grd_mth_ol_se	Geo Dist SE of Grade Slope of SD of Ach, Math, OLS est, CS	HLM using SEDA HETOP est
sd_avg_ela_ol_se	Geo Dist SE of SD of Ach, ELA, OLS est, CS	HLM using SEDA HETOP est
sd_grd_ela_ol_se	Geo Dist SE of Grade Slope of SD of Ach, ELA, OLS est, CS	HLM using SEDA HETOP est

Each "pool" file of estimates, which pool across grades, years, and subjects, contains the following variables (variables shown here are from the CS scale file, but variables and structure remain the same for the GCS "pool" file):

Name	Label	Source
leaidC	NCES ID - Geographic School Districts	Identifier (ID) (common to multiple datasets)
leaname	District (LEA) Name	Common Core of Data (CCD)
fips	FIPS State Code	Identifier (ID) (common to multiple datasets)
stateabb	State Abbreviation	Identifier (ID) (common to multiple datasets)
subgroup	Subgroup of estimates, or subgroups of gap estimates	HLM using SEDA HETOP est
gap_est	Row is a gap estimate (subgroup indicates which gap)	HLM using SEDA HETOP est
tot_asmts	Total number of tests (math+ela) for pooled estimates	HLM using SEDA HETOP est
cellcount	Number of district-grade-year-subject cases used in pooling	HLM using SEDA HETOP est
mn_avg_ol	Geo Dist Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
mn_grd_ol	Geo Dist Grade Slope of Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
mn_mth_ol	Geo Dist Math-ELA Diff in Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
mn_avg_eb	Geo Dist Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_grd_eb	Geo Dist Grade Slope of Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_mth_eb	Geo Dist Math-ELA Diff in Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_avg_eb_se	Geo Dist SE of Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_grd_eb_se	Geo Dist SE of Grade Slope of Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_mth_eb_se	Geo Dist SE of Math-ELA Diff in Mean Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
mn_avg_ol_se	Geo Dist SE of Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
mn_grd_ol_se	Geo Dist SE of Grade Slope of Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
mn_mth_ol_se	Geo Dist SE of Math-ELA Diff in Mean Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_avg_ol	Geo Dist SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_grd_ol	Geo Dist Grade Slope of SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_mth_ol	Geo Dist Math-ELA Diff in SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_avg_eb	Geo Dist SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
sd_grd_eb	Geo Dist Grade Slope of SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est

sd_mth_eb	Geo Dist Math-ELA Diff in SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
sd_avg_eb_se	Geo Dist SE of SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
sd_grd_eb_se	Geo Dist SE of Grade Slope of SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
sd_mth_eb_se	Geo Dist SE of Math-ELA Diff in SD of Ach, Math&ELA, EB est, CS	HLM using SEDA HETOP est
sd_avg_ol_se	Geo Dist SE of SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_grd_ol_se	Geo Dist SE of Grade Slope of SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est
sd_mth_ol_se	Geo Dist SE of Math-ELA Diff in SD of Ach, Math&ELA, OLS est, CS	HLM using SEDA HETOP est

Covariate Data

Below is a list of variables included in the covariate files and information on their construction. The variables included in this list are from the file at the lowest level of aggregation—GSD by year by grade. Additional data files with the same measures are included at the GSD by year level and the GSD level. For the most part, the aggregated data files are derived by simply taking the means of the measures from the GSD-year-grade file and collapsing to the GSD-year level and to the GSD-level.

Name	Label	Source
leaidC	NCES Local Education Agency (District) Code	ID
leaname	LEA Name	CCD
year	spring of school year	CCD
grade	Grade Level	ID
fips	Fips State Code	ID
stateabb	State Abbreviation	ID
metroid03_orig	Metro ID: 2003 definition (original)	Census
metroname03	Metro Name, 2003 Definition	Census
micro03	micropolitan area, 2003 definition	Census
metro03	metropolitan area, 2003 definition	Census
metroid03	Metro ID: 2003, metro by fips	Census
metroid09_orig	Metro ID: 2009 definition (original)	Census
metroname09	Metro Name, 2009 Definition	Census
metro09	metropolitan area, 2009 definition	Census
micro09	micropolitan area, 2009 definition	Census
metroid09	Metro ID: 2009, metro by fips	Census
metroid13_orig	Metro ID: 2013 definition (original)	Census
metroname13	Metro Name, 2013 Definition	Census
metro13	metropolitan area, 2013 definition	Census
micro13	micropolitan area, 2013 definition	Census
metroid13	Metro ID: 2013, metro by fips	Census
czid	Commuting Zone ID, 2000 Definition	Census

countyid	county code	CCD
countyname	county name	CCD
cdcode	Congressional District Code	CCD
urban	city/urban locale	CCD
perind	percent native americans in the grade	CCD
perasn	percent asians in the grade	CCD
perhsp	percent hispanics in the grade	CCD
perblk	percent blacks in the grade	CCD
perwht	percent whites in the grade	CCD
perfrl	percent free lunch in the grade	CCD
pernonfrl	percent not free lunch in the grade	CCD
perrl	percent reduced lunch in the grade	CCD
pernonrl	percent not reduced lunch in the grade	CCD
perell	% of all Students in District that are ELL	CCD
perspeced	% of all Students in District that are Special Ed	CCD
ind	N native americans in the grade	CCD
asn	N asians in the grade	CCD
hsp	N hispanics in the grade	CCD
blk	N blacks in the grade	CCD
wht	N whites in the grade	CCD
frl	N free lunch in the grade	CCD
nonfrl	N not free lunch in the grade	CCD
rl	N reduced lunch in the grade	CCD
nonrl	N not reduced lunch in the grade	CCD
frlunch	N free or reduced lunch in the grade	CCD
nonfrlunch	N not free or reduced lunch in the grade	CCD
totenrl	Number of Students in Grade	CCD
nsch	Number of Schools in the District	CCD
ncharters	Number of Charter Schools in the District	CCD
gslo	Lowest Grade Offered in District	CCD

gshi	Highest Grade Offered in District	CCD
speced	Number of Special Ed (IEP) Students in District	CCD
ell	Number of Eng Language Learners in District	CCD
elmtch	Number of Elementary Teachers	CCD
tottch	Total Number of Teachers	CCD
aides	Number of Instructional Aides	CCD
corsup	Number of instructional coordinators and supervisors	CCD
elmgui	Number of Elementary Guidance Counselors	CCD
stutch_wht	pupil teacher ratio-- average white student's school	CCD
stutch_blk	pupil teacher ratio-- average black student's school	CCD
stutch_hsp	pupil teacher ratio-- average hispanic student's school	CCD
stutch_all	pupil-teacher ratio-- average all student's school	CCD
diffstutch_blkwht	stutch_blk-stutch_wht	CCD
diffstutch_hspwht	stutch_hsp-stutch_wht	CCD
ratstutch_whtblk	stutch_wht/stutch_blk	CCD
ratstutch_whthsp	stutch_wht/stutch_hsp	CCD
flunch_wht	percent free lunch in average white student's school	CCD
flunch_blk	percent free lunch in average black student's school	CCD
flunch_hsp	percent free lunch in average hisp student's school	CCD
diffexplch_blkwht	flunch_blk-flunch_wht	CCD
diffexplch_hspwht	flunch_hsp-flunch_wht	CCD
percharter_all	Percentage of Public School Students in Charters (all)	CCD
percharter_wht	Percentage of Public School Students in Charters (wht)	CCD
percharter_blk	Percentage of Public School Students in Charters (blk)	CCD
percharter_hsp	Percentage of Public School Students in Charters (hsp)	CCD
hswhtblk	Information index between schools: White/Black	CCD
hswhthsp	Information index between schools: White/Hispanic	CCD
hsflnfl	Information index between schools: FRPL/Non FRPL	CCD
ppexp_tot	Total PP Expenditures- Tot Exp/Enrl	CCD
ppexp_inst	Current PP Expenditures, Instruction- Inst Exp/Enrl	CCD

pprev_tot	Revenue Per Pupil- Total Revenue/Tot Enrl	CCD
totppe_fleslope	State Slope- Total PPE = % FLE	CCD
instppe_fleslope	State Slope- Instructional PPE = % FLE	CCD
baplus_wht	% of adults with ba+ (wht)	SDDS/ACS
poverty517_wht	% of hh with 5-17 yr olds in poverty (wht)	SDDS/ACS
snap_wht	% of hh receiving snap benefits (wht)	SDDS/ACS
singmom_wht	% hh with children, female head (wht)	SDDS/ACS
samehouse_wht	% living in same house as last year (wht)	SDDS/ACS
unemp_wht	% unemployed (wht)	SDDS/ACS
baplus_hsp	% of adults with ba+ (hsp)	SDDS/ACS
poverty517_hsp	% of hh with 5-17 yr olds in poverty (hsp)	SDDS/ACS
snap_hsp	% of hh receiving snap benefits (hsp)	SDDS/ACS
singmom_hsp	% hh with children, female head (hsp)	SDDS/ACS
samehouse_hsp	% living in same house as last year (hsp)	SDDS/ACS
unemp_hsp	% unemployed (hsp)	SDDS/ACS
baplus_blk	% of adults with ba+ (blk)	SDDS/ACS
poverty517_blk	% of hh with 5-17 yr olds in poverty (blk)	SDDS/ACS
snap_blk	% of hh receiving snap benefits (blk)	SDDS/ACS
singmom_blk	% hh with children, female head (blk)	SDDS/ACS
samehouse_blk	% living in same house as last year (blk)	SDDS/ACS
unemp_blk	% unemployed (blk)	SDDS/ACS
baplus_all	% of adults with ba+ (all)	SDDS/ACS
poverty517_all	% of hh with 5-17 yr olds in poverty (all)	SDDS/ACS
singmom_all	% hh with children, female head (all)	SDDS/ACS
snap_all	% of hh receiving snap benefits (all)	SDDS/ACS
samehouse_all	% living in same house as last year (all)	SDDS/ACS
unemp_all	% unemployed (all)	SDDS/ACS
pctenglish1	% hispanics- speak english only, very well or well	SDDS/ACS
pctenglish2	% hispanics - speak english only, very well	SDDS/ACS
pctenglish3	% hispanics - speak english only	SDDS/ACS

pctforeign	% hispanics- foreign born	SDDS/ACS
pctmexico	% hispanics- mexican	SDDS/ACS
pctpuerto	% hispanics- puerto rican	SDDS/ACS
pctcuba	% hispanics- cuban	SDDS/ACS
pctcentral	% hispanics- central american	SDDS/ACS
pctsouth	% hispanics- south american	SDDS/ACS
inc50all	income at 50th percentile (all)	SDDS/ACS
incrat9010all	90/10 income ratio (all)	SDDS/ACS
incrat9050all	90/50 income ratio (all)	SDDS/ACS
incrat5010all	50/10 income ratio (all)	SDDS/ACS
inc50blk	income at 50th percentile (blk)	SDDS/ACS
incrat9010blk	90/10 income ratio (blk)	SDDS/ACS
incrat9050blk	90/50 income ratio (blk)	SDDS/ACS
incrat5010blk	50/10 income ratio (blk)	SDDS/ACS
inc50hsp	income at 50th percentile (hsp)	SDDS/ACS
incrat9010hsp	90/10 income ratio (hsp)	SDDS/ACS
incrat9050hsp	90/50 income ratio (hsp)	SDDS/ACS
incrat5010hsp	50/10 income ratio (hsp)	SDDS/ACS
inc50wht	income at 50th percentile (wht)	SDDS/ACS
incrat9010wht	90/10 income ratio (wht)	SDDS/ACS
incrat9050wht	90/50 income ratio (wht)	SDDS/ACS
incrat5010wht	50/10 income ratio (wht)	SDDS/ACS
giniall	Gini Coefficient (all)	SDDS/ACS
giniwht	Gini Coefficient (wht)	SDDS/ACS
giniblk	Gini Coefficient (blk)	SDDS/ACS
ginihsp	Gini Coefficient (hsp)	SDDS/ACS
paredVblkwht	vgap for parent education, white-black	SDDS/ACS
paredVhspwht	vgap for parent education, white-hispanic	SDDS/ACS
incVblkwht	vgap for income, white-black	SDDS/ACS
incVhspwht	vgap for income, white-hispanic	SDDS/ACS

baplus_mal	Percent of Males with BA or Higher	SDDS/ACS
baplus_fem	Percent of Female with BA or Higher	SDDS/ACS
pov_mal	Percent of Males in Poverty	SDDS/ACS
pov_fem	Percent of Female in Poverty	SDDS/ACS
occbus_mal	Percent of Males in Management, Business and Financial Occs	SDDS/ACS
occbus_fem	Percent of Females in in Management, Business and Financial Occs	SDDS/ACS
occsci_mal	Percent of Males in Computer, Engineering and Science Occs	SDDS/ACS
occsci_fem	Percent of Females in Computer, Engineering and Science Occs	SDDS/ACS
occeduc_mal	Percent of Males in Education, Legal, Com Serv, Arts, Media Occs	SDDS/ACS
occeduc_fem	Percent of Females in Education, Legal, Com Serv, Arts, Media Occs	SDDS/ACS
occhealth_mal	Percent of Males in Health Practitioners and Technical Occs	SDDS/ACS
occhealth_fem	Percent of Females in Health Practitioners and Technical Occs	SDDS/ACS
occserv_mal	Percent of Males in Service Occs	SDDS/ACS
occserv_fem	Percent of Females in Service Occs	SDDS/ACS
occsales_mal	Percent of Males in Sales Occs	SDDS/ACS
occsales_fem	Percent of Females in Sales Occs	SDDS/ACS
occtrade_mal	Percent of Males in Natural Resources, Construction, Maintenance	SDDS/ACS
occtrade_fem	Percent of Females in Natural Resources, Construction, Maintenance	SDDS/ACS
inlf_mal	Percent of 25-64 Year Old Males in Labor Force	SDDS/ACS
inlf_fem	Percent of 25-64 Year Old Females in Labor Force	SDDS/ACS
unemp_mal	Percent of 25-64 Year Old Males in LF & Unemployed	SDDS/ACS
unemp_fem	Percent of 25-64 Year Old Females in LF & Unemployed	SDDS/ACS
incVmalfem	vgap for income, male-female	SDDS/ACS
educVmalfem	vgap for education, male-female	SDDS/ACS
teenbirth_all	percent of 15-19 year olds giving birth	SDDS/ACS
sesall	standardized ses composite (all races)	SDDS/ACS
seswht	standardized ses composite (whites)	SDDS/ACS
sesblk	standardized ses composite (black)	SDDS/ACS
seshsp	standardized ses composite (hispanic)	SDDS/ACS
sesallimp1	sesall imputed flag, 5 variables	flag

sesallimp2	sesall imputed flag, 3 variables	flag
seswhtimp1	seswht imputed flag, 5 variables	flag
seswhtimp2	seswht imputed flag, 3 variables	flag
sesblkimp1	sesblk imputed flag, 5 variables	flag
sesblkimp2	sesblk imputed flag, 3 variables	flag
seshspimp1	seshsp imputed flag, 5 variables	flag
seshspimp2	seshsp imputed flag, 3 variables	flag