Stanford Education Data Archive

Technical Documentation

Version 4.1

June 2021

Erin M. Fahle, St. John's University

Belen Chavez, Stanford University

Demetra Kalogrides, Stanford University

Benjamin R. Shear, University of Colorado Boulder

Sean F. Reardon, Stanford University

Andrew D. Ho, Harvard University

Suggested citation:

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

Contents

I. What is SEDA?

The Stanford Education Data Archive (SEDA) is part of the Educational Opportunity Project at Stanford University (https:\\edopportunity.org), an initiative aimed at harnessing data to help scholars, policymakers, educators, and parents learn how to improve educational opportunities for all children. SEDA includes a range of detailed data on educational conditions, contexts, and outcomes in schools, school districts, counties, commuting zones, and metropolitan statistical areas across the United States. Available measures differ by aggregation; see Sections I.A. and I.B. for a complete list of files and data.

By making the data files available to the public, we hope that anyone who is interested can obtain detailed information about U.S. 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, 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 and reported here are those of the authors alone; they do not represent the views of the U.S. Department of Education, NCES, or any of the aforementioned funding agencies.

I.A. Overview of Test Score Data Files

SEDA 4.1 contains test score data files for schools, geographically defined school districts, counties, commuting zones, metropolitan statistical areas, and states. Test score data files contain information about the average academic achievement as measured by standardized test scores administered in 3rd through 8th grade in mathematics and Reading Language Arts (RLA) over the 2008-09 through 2017-18 school years. The measures contained in the data files are detailed below.

School Files. There are two school-level test score data files, corresponding to the two different metrics in which the data are released: the cohort standardized (CS) scale and the grade cohort standardized (GCS) scale. In each file there are variables corresponding to the average test score in the middle grade of the data, the average “learning rate” across grades (grade slope), the average “trend” in test scores across cohorts (cohort slope), and the average difference between math and RLA test scores (math slope). Each measure is included along with its respective standard error. School estimates are only reported for all students; no estimates are provided by demographic subgroup.

Geographic School District, County, Commuting Zone, Metropolitan Statistical Area, and State Files. Thirty test score files are released corresponding to the five units (geographic school districts, counties, metropolitan areas, commuting zones, and states) by two scales (CS and GCS) by three pooling levels (long, pooled by subject, and pooled overall). “Long” files contain estimates for each grade and year separately; “pooled by subject” (or poolsub) files contain estimates that are averaged across grades and years within subjects; and “pooled overall” (or pool) files contain estimates that are averaged across grades, years, and subjects. In the long files there are variables corresponding to test score means by subgroup and their respective standard errors in each grade, year, and subject. In the poolsub files, there are variables corresponding to the average test score mean in math and in RLA (averaged across grades and years), the average “learning rate” across grades in math and in RLA, and the average “trend” in test scores across cohorts in math and in RLA, along with their standard errors. In the pooled overall file, there are variables corresponding to the average test score mean (averaged across grades, years, and subjects), the average “learning rate” across grades, the average “trend” in test scores across cohorts, and the average difference between math and RLA test scores, along with their standard errors. Estimates are reported for all students and by demographic subgroups.

Table 1 lists the files and file structures. Lists of variables can be found in the codebook that accompanies this documentation.

I.B. Covariate Data

SEDA 4.1 also provides estimates of socioeconomic, demographic, and segregation characteristics of schools, districts, counties, metropolitan areas, and states. The measures included in the district, county, metropolitan area, and state covariates files come from two primary sources. The first is the American Community Survey (ACS) detailed tables which we obtained from the National Historical Geographic Information System (NHGIS) web portal.[1] These data include demographic and socioeconomic characteristics of individuals and households residing in each unit. The second is the Common Core of Data (CCD) which is an annual survey of all public elementary and secondary schools and school districts in the United States. The data include basic descriptive information on schools and school districts, including demographic characteristics.[2] The measures included in the school covariates files come from the CCD as well as the Civil Rights Data Collection (CRDC). The CRDC includes data about school demographics, teacher experience, school expenditures, high school course enrollments as well as other information not used here.[3]

Twelve files (three per unit) in SEDA 4.1 contain CCD and ACS that data have been curated for use with the geographic school district-level, county-level, metropolitan area-level, and state-level achievement data. These data include raw measures as well derived measures (e.g., a composite socioeconomic status measure, segregation measures). Each of the three covariate files we construct for each unit contain the same variables but differ based on whether they report these variables separately for each grade and year, average across grades (providing a single value per unit per year), or average across grades and years (providing a single value per unit). Two data files are provided for schools; one includes an observation for each school in each year and the other reports a single record for each school that is the average across years. School level data from the CCD is used to aggregate various measures to the geographic school

[1] The ACS data is available for download from the IPUMS-NHGIS website at: https://www.nhgis.org/. Full citation: Steven Manson, Jonathan Schroeder, David Van Riper, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS. 2019. http://doi.org/10.18128/D050.V14.0
[2] The CCD is available for download from the NCES website: https://nces.ed.gov/ccd/.
[3] The CRDC data is available for download at: https://ocrdata.ed.gov/.

district, county, and metropolitan statistical area levels.[4] The measures from the ACS are downloaded separately at the school district, county, metropolitan area, and state levels of aggregation and are not available at the school level. The Covariate Data Construction section of the documentation describes more detail about the construction of these data files and the computation of derived variables. Table 2 lists the names and file structures of the covariate data files.

I.C. Data Use Agreement

Prior to downloading the data, users must sign the data use agreement, shown below.

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., Jang, H., & Chavez, B. (2021). Stanford Education Data Archive (Version 4.1). Retrieved from http://purl.stanford.edu/db586ns4974.

[4] The exception is the finance data (total instructional expenditures and per pupil expenditures) which are only available at the district level. These measures are aggregated from the district level to the county and metropolitan area levels and are not available at the school level.

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.

II. Achievement Data Construction

II.A. Source Data

The SEDA 4.1 achievement data is constructed using data from the ED*Facts* data system housed by the U.S. Department of Education. The ED*Facts* data system collects aggregated test score data from each state's standardized testing program as required by federal law. Specifically, each state is required to test every student in grades 3 through 8 in math and Reading Language Arts (RLA) each year.[5] States have the flexibility to select or design a test of their choice that measures student achievement relative to the state's standards. Additionally, states set their own benchmarks or thresholds for the levels of performance, e.g., "proficient," in each grade and subject. States select 2 to 5 performance levels, where one or more levels represent "proficient" grade-level achievement.

To ED*Facts*, states report the number of students in each school, subgroup, subject, grade, and year scoring at each of their defined performance levels; <u>*no individual student-level data is reported*</u>.[6] ED*Facts* currently contains these school assessment outcomes for ten consecutive school years from the 2008-09 school year to the 2017-18 school year in grades 3 to 8 in RLA and math. The student subgroups include race/ethnicity, gender, and socioeconomic disadvantage, among others not used in SEDA 4.1.

The raw ED*Facts* data used in SEDA include no suppressed cells, nor do they have a minimum cell size for reporting. Each row of data corresponds to a school-subgroup-subject-grade-year cell. Table 3 illustrates the structure of the raw data from ED*Facts* prior to use in constructing SEDA 4.1.

[5] Federal law also requires state to report data for one high school grade; however, that data is not currently used in SEDA.

[6] In recent years (2013-2018), the data is further broken out by the assessment type administered to students: regular assessments, regular assessments with accommodations, and alternate assessments with grade-level standards, modified standards and alternate standards. However, in 2009-2012, ED*Facts* does not distinguish students taking regular from alternate assessments. Therefore, for consistency in all years, we use all performance data reported in ED*Facts*, including results of students taking both regular and alternate assessments.

II.B. Definitions

Commuting Zone: Regions defined by the geographic boundaries of a local economy. We use the ERS 2010 boundary definitions (https://sites.psu.edu/psucz/data/) which are the most recent commuting zone definitions.

Geographic School District: The aggregate of all public schools in SEDA (except for special education[7] and virtual[8] schools) that are physically located within a geographically defined public Elementary or Unified school district. We use the 2019 Elementary and Unified School District Boundaries (https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries) to define these districts. Note that there are some districts in SEDA that are not geographically defined that are included in our analysis. In this document, the terms "district," "geographic school district," and "geographically-defined school district" are used interchangeably.

Group: The term "group" refers to a subgroup-unit. For schools, the only available subgroup is all students. For geographic school districts, counties, commuting zones, metropolitan areas, and states, data for other subgroups are available when estimates are sufficiently precise.

Metropolitan Statistical Area: A county or group of counties with a population exceeding 50,000 and encompassing an urban area, combined with any surrounding counties with strong commuting ties to the urban area (https://www.census.gov/programs-surveys/metro-micro/about/glossary.html). The U.S. Census Bureau revises the metropolitan statistical area definitions after each decennial census. We use the 2013 U.S. Census Bureau definitions, which are the definitions based on the 2010 census (https://www.census.gov/programs-surveys/metro-micro/geographies/geographic-reference-files.2013.html). We make one modification to the definitions: The Census defines very large metropolitan areas as Consolidated Metropolitan Statistical Areas (CMSAS); each CMSA is subdivided into Metropolitan Area Divisions. We treat each division as a separate metropolitan area for analysis purposes, as the CMSAs generally quite large.

[7] As defined by school type in the CCD Public Elementary/Secondary School Universe Survey Data.
[8] As defined by virtual text in the CCD Public Elementary/Secondary School Universe Survey Data.

State: States are identified by their FIPS state code. We include all 50 states plus Washington, DC. Data for Puerto Rico will be added in a future release of SEDA.

Subcategory: The term "subcategory" refers to the subcategory to which a subgroup belongs. In addition to data for all students, we have data for the following subcategories: gender, race, and economic status. The gender subcategory contains two subgroups, male and female. The race subcategory includes the Asian, Black, Hispanic, Multiracial, Native American, and White subgroups. The economic status subcategory includes the economically disadvantaged and not economically disadvantaged subgroups.

Subgroup: The term "subgroup" refers to the group of students to which an estimate pertains. Subgroups include: all, Asian, Black, Hispanic, Multiracial, Native American, White, female, male, economically disadvantaged, and not economically disadvantaged students. We do not currently report data for the Multiracial subgroup.

Unit: The term "unit" refers to the aggregation or the geography level of the data. This may be a school, geographic school district, county, commuting zone, metropolitan area, or state.

II.C. Construction Overview

The construction process produces mean test score estimates for schools, districts, counties, metropolitan areas, commuting zones, and states on two nationally comparable scales in a series of nine steps, outlined in Figure 1. We provide a brief conceptual description of each step here and the full technical details about each step in **Section II.D**.

Step 1: Creating the Crosswalk. This step links each public school to a unique stable school, geographic school district, county, commuting zone, metropolitan area, and state.

Step 2: Data Cleaning. This step removes data not used in SEDA 4.1.

Step 3: Estimating and Linking Cutscores. This step uses Heteroskedastic Ordered Probit (HETOP) models to estimate the state-grade-subject-year cutscores, link the estimated cutscores to the NAEP scale, and standardize the linked cutscores to the Cohort Standardized (CS) scale. The resulting cutscores are comparable across states and years.

Step 4: Selecting Data for Mean Estimation. This step selects data for *unit-subgroup-subject-grade-year* cases that will be used in estimation. We exclude cases with low participation in the assessment or high percentages of students taking alternate assessments. We also exclude cases for which we have insufficient data to produce an estimate.

Step 5: Estimating Means. This step uses the pooled HETOP model to estimate school, district, county, commuting zone, metropolitan area, and state subgroup-subject-grade-year means and standard deviations, along with their standard errors, based on the cutscores from Step 3 and the data prepared in Step 4.

Step 6: Creating Additional Reporting Scales. This step creates Grade Cohort Standardized (GCS) estimates for all units, such that each unit is interpreted as 1 grade level. From this point onward, we have two scales of the data for all units: CS and GCS. Subsequent steps are equivalent for both scales unless otherwise noted.

Step 7: Calculating Achievement Gaps. This step estimates White-Black, White-Hispanic, White-Asian, White-Native American, White-Multiracial, male-female, and nonpoor[9]-poor[10] achievement gaps for districts, counties, metropolitan areas, commuting zones, and states in each subject-grade-year where there is sufficient data.

Step 8: Pooling Mean and Gap Estimates. This step estimates the average achievement, learning rate, and trend in test scores by subject and overall for each unit and scale. From this point onward, we have three types data for districts, counties, metropolitan areas, commuting zones, and states: long (not pooled by grade, year, or subject), pooled by subject (poolsub; pooled over grades and years by subject), and pooled overall (pool; pooled over grades, years, and subjects). For schools, we only report the pooled overall (pool) estimates.

Step 9: Suppressing Data for Release. The step suppresses estimates that are too imprecise to be useful or do not reflect the performance of at least 20 unique students in

[9] "Non-poor" refers to the student subgroup that is not identified as "Economically Disadvantaged" in ED*Facts.*
[10] "Poor" refers to "Economically Disadvantaged" student subgroup in ED*Facts.*

both long and pooled files for all units and scales. For estimates reported in the long files, this step adds a small amount of random noise to meet the reporting requirements of the US Department of Education.

II.D. Notation

In the remainder of the documentation, we use the following mathematical notation:

- Mean estimates are denoted by $\hat{\mu}$ and standard deviation estimates by $\hat{\sigma}$.
- The cutscore estimates are denoted as $\hat{c}_1, \dots, \hat{c}_K$. There are K total cutscores in each state-subject-grade-year.
- A subscript indicates the aggregation of the estimate. We use the following subscripts:
    - u = unit (generic)
    - n = school
    - d = geographic school district
    - c = county
    - z = commuting zone
    - m = metropolitan area
    - f = state
  - r = subgroup
    - all = all students
    - asn = Asian
    - blk = Black
    - hsp = Hispanic
    - mtr = Multiracial
    - nam = Native American
    - wht = White
    - fem = female
    - mal = male
    - ecd = economically disadvantaged
    - nec = not economically disadvantaged
    - wag = White-Asian gap
    - wbg = White-Black gap
    - whg = White-Hispanic gap
    - wmg = White-Multiracial gap
    - wng = White-Native American gap
    - mfg = male-female gap
    - neg =not economically disadvantaged-economically disadvantaged gap
  - y = year
  - b = subject
  - g = grade

- A superscript indicates the scale of the estimate. The metric is generically designated as $\boldsymbol{x}$. There are four scales. The first two scales are only used in construction. The latter two scales are reported:

  $state$ = state-standardized metric
  $naep$ = NAEP test score scale metric
  cs = cohort scale metric
  gcs = grade within cohort scale metric

II.E. Detailed Construction Overview

Step 1. Creating the Crosswalk

The primary purpose of the crosswalk is to create stable school identifiers and assign schools to larger geographic units such as geographically-defined school districts, counties, metropolitan areas, commuting zones, and states. We use the CCD's *Public Elementary/Secondary School Universe Survey Data* (Directory and School Characteristics files) and the *Local Education Agency (School District) Universe Survey Data* (Directory files) as the basis for the crosswalk.

Stable School IDs. Since we want to be able to track schools as they change districts (district changes could be due to districts splitting, merging or some other administrative change), we create stable school IDs using the CCD's *Longitudinal ID Crosswalks*. According to the CCD documentation, "Schools are uniquely identified in CCD by the 12-digit variable **ncessch**. This variable is a combination of the state code (the first two digits or FIPST), the Local Education Agency (LEA) ID (the first seven digits or **leaid**) and the last five digits (**schid**). It was always intended that the **schid** should be unique within the state so that a school could be tracked from year-to-year even if a re-organization caused it to change LEAs. However, a system error created some duplicate **schid**s within some states."[11]

Because of some schools changed school IDs during the 2008-09 to 2017-18 time period, we use the CCD's longitudinal ID crosswalks[12] from the CCD's Reference Library (https://nces.ed.gov/ccd/reference_library.asp) to uniquely identify schools. These stable school IDs became the last 5 digits of the **sedasch** IDs. The final **sedasch** ID is comprised of 12 digits in the same format as the NCES school ID (ncessch). The **sedasch** ID's first 2 digits correspond to the state FIPS code, first 7 digits correspond to a stable district ID (**sedalea**), and the last 5 digits correspond to the school ID within the state. The next section describes how schools were assigned into geographic school districts. This assignment determines the 7-digit stable district ID that is used as the first part of the **sedasch** ID.

---

[11] See Page 1 in the NCES *School Crosswalk (SY 2014-15 to 2015-16)*. Retrieved from: https://nces.ed.gov/ccd/doc/3_Changes_to_NCES_School_ID_2015_16.docx. Bolding added for emphasis.

[12] *School Crosswalk (SY 2014-15 to 2015-16)*, *School Crosswalk (SY 2015-16 to 2016-17)*

Assignment of Schools to Geographically Defined Districts in SEDA. Most public school districts in the U.S. are geographically defined.[13] In SEDA we use the 2019 EDGE Unified and Elementary School District Boundaries to define districts used in SEDA which we call geographic school districts. Commonly, public school districts have administrative control over the traditional public schools that fall within their specific geographic boundaries. However, there may be some schools physically located within the geographic boundary of a school district that are not under its administrative control. For example, there may be charter schools or magnet schools located within the boundaries of a school district that are operated by a different school district or a charter school network (which may have no geographic boundary).

In SEDA we have several rules around what schools are placed or excluded from geographic school districts based on location (latitude and longitude coordinates), school type information, and school status information. The aim is for the district test score estimates in SEDA to reflect most of the public school students living within the geographic boundaries of the school district. The motivation for this assignment is to better align the average test score estimates with the demographic and socioeconomic data from ACS, which are reported for all families living within geographic school district boundaries.

We use a school's most recently observed CCD information on school status, charter status, magnet status, coordinates, and county ID to create time-invariant information for schools in SEDA. Below are the geographic district assignment rules in SEDA based on these time-invariant characteristics:

Charter schools: All (except for special education) charter schools are geolocated and reassigned to the Elementary or Unified District in which they physically reside.

Magnet schools: All (except for special education) magnet schools are geolocated and reassigned to the Elementary or Unified District in which they physically reside.

[13] According to NCES, "The US has more than 13,000 geographically defined school districts. These include districts that are administratively and fiscally independent of any other government, as well as public school systems that lack sufficient autonomy to be counted as separate governments and are classified as a dependent agency of some other government—a county, municipal, township, or state. Most public school systems are Unified districts that operate regular, special, and/or vocational programs for children in Prekindergarten through 12th grade." Retrieved from: https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries

Schools operated by secondary districts: All schools with LEAIDS corresponding to secondary school districts in the Secondary School District Boundary file are geolocated to the Elementary or Unified geographic district in which they physically reside. This is because the ED*Facts* data we use is for grades 3-8.

Virtual schools: By their nature, most virtual schools do not draw students from within district geographic boundaries. We identify schools as virtual using CCD data from 2013-14 through 2017-18 Public Elementary/Secondary School Universe Survey Data. The virtual school identifier did not exist in earlier years of data, so we flag schools as virtual in all years of our data if they were identified as virtual by any CCD indicators in the last year in which they were observed in the data.[14] Virtual schools are excluded from SEDA.

Special Education Schools: We classify schools as Special Education schools if they are ever classified as "Special Education" in the school-type variable in the CCD data between 2009 and 2018.[15] We exclude these schools from their geographic districts, counties, commuting zones, and metropolitan areas and instead assign them to a statewide "SEDA special education district." This ensures that their test scores are not used in estimating the test score distributions in any geographic unit. Because many special education schools enroll students who take alternative assessments, their school-average test scores are not comparable to those in other schools and we do not report test scores for such schools.

BIE Controlled Schools: Schools controlled by the Bureau of Indian Education (BIE) are placed in the Elementary or Unified District in which they are physically located.

Schools operated by supervisory unions: We place all (except for special education) schools that are part of supervisory unions in their supervisory union LEAs. This rule mostly affects schools in Vermont and New York. For example, New York City School District (LEA 3620580) is a supervisory union comprised of 33 subordinate school districts.

[14] In 2013-2015, we identified 12 non-virtual schools in Alabama identified as "virtual" by the CCD indicator. We treat these as regular schools in all subsequent steps.

[15] Special Education as defined by School Type in CCD Public Elementary/Secondary School Universe Survey Data

Closed Schools: We geolocate all closed schools (except for special education schools) to the Elementary or Unified Districts in which they physically reside.

District of Columbia Schools: All schools within Washington, DC are given DC's geographic district ID (1100030).

Hawaii Schools: All schools within Hawaii are given Hawaii's geographic district ID (1500030).

Puerto Rico Schools: All schools within Puerto Rico are given Puerto Rico's geographic district ID (720003).

All students in a school that is assigned to a particular geographically defined school district will be reflected in that district's estimate. School districts used in SEDA are identifiable by their **sedalea**. You can identify a given school's assigned district by looking at the first 7 digits of the **sedasch** ID, which will be the **sedalea** ID.

School Assignment to Higher Aggregations. For each school, we use the county code provided in CCD in the most recent year the school was observed. This county code (**sedacounty**) is stable over time. The county code is then used to merge on the 2013 metropolitan areas and 2010 commuting zones. Therefore, all schools in SEDA also have the same metropolitan area (**sedametro**), commuting zone (**sedacz**), and state (**fips**) over time.

Step 2. Data Cleaning

In this step, we first merge the ED*Facts* data (described under **II.A. Source Data**, above) by NCES school ID (**ncessch**) and year with the crosswalk developed in **Step 1**. With this merge, the ED*Facts* data now have stable unit IDs (**sedasch**, **sedalea**, **sedacounty**, **sedametro**, **sedacz**, and **fips**) which will be used throughout the SEDA process. We then create flags for schools (by state, grade, year, and subject) that we intend to drop before estimation. The flags we create are listed below:

State participation is less 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 2017-18 school years. 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. Participation is the ratio of the number of test scores reported to the number enrolled students in a given state-subject-grade-year:

$$participation_{fygb} = \frac{numscores_{fygb}}{numenrl_{fygb}} \tag{2.1}$$

for each state f, year y, grade g, and subject b. This state-level suppression is important because both the quality of the estimates and the linkage process depend on having the full 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 or pilot testing. Note that we do not suppress any entire state-subject-grade-year cases prior to the 2012-13 school year as enrollment data are not available in ED*Facts*. However, opt out and non-participation was low in 2012-13 (no state was excluded based on this threshold), which suggests states met 95% threshold in prior years when data are not available. A full list of the states, grades, years, and subjects this affects is in Table 4.

Duplicate BIE or ED*Facts* IDs: We remove a handful of places from the data that report data under both BIE school IDs and regular school IDs. These were identified by the NCES. According to the CCD documentation, "There is a possibility that some schools are reported in CCD by both the BIE and the state in which the schools are located, leading to a double counting of students and staff. (NCES allows for the possibility of co-located schools, so a double-counting of schools is not an issue.) This arises from situations where both the state and BIE share operational or financial responsibilities for a school."[16] In order for SEDA to also avoid double counting, we remove the schools from the list and retain their counterparts listed in Table 5.

Virtual schools: We flag all virtual schools in ED*Facts* based on the crosswalk and remove them from SEDA.

Not all students took the same content tests within the state-subject-grade-year: There are two common ways this appears within the data. First, there are cases where districts were permitted to administer locally selected assessments. This occurred in Nebraska

[16] See Page 1 in the NCES *Double Counting of BIE Reported Schools* documentatino. Retrieved from: https://nces.ed.gov/ccd/doc/5_Double_Counting_of_Bureau_of_Indian_Education_Schools_3.4.2020.docx

during SY 2008-2009 (RLA and Math) and SY 2009-2010 (Math). Second, in some cases students take end-of-course rather than end-of-grade assessments. This is the case in some or all years for 7th and 8th grade math for California for years SY 2008-09 to SY 2013-2014, Virginia, and Texas (among other states, reported in Table 4). When test scores measure different content and are reported on different scales using different cut scores, proficiency counts cannot be compared across districts or schools within these state-subject-grade-year cases. All of these flagged places are removed from SEDA.

Insufficient data was reported to ED*Facts*: Some states reported no data in certain years. Wyoming did not report any assessment outcomes in 2009-10. Others reported data from which we cannot recover reliable estimates. In the 2008-09, 2009-10, and 2010-11 school years, Colorado reported data 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/or standard deviations in most regions. In the 2014-15 and 2016-17 school years, New Mexico reported data in on two proficiency categories for RLA and did not report data for 2017-18, so we remove these cases because the last two years of data are consecutive and fall at the end of the time series. These places are all flagged and removed from SEDA. See full list reported in Table 4 (marked as manual removals).

NAEP data was not reported in any years or grades for a state-equivalent and subject. Puerto Rico does not take the NAEP assessment in Reading Language Arts, so linking the Puerto Rico RLA estimates to a common national RLA scale is not possible.

Alternate Assessments. In 2008-09 through 2011-12, ED*Facts* does not distinguish students taking regular from alternate assessments; these counts were combined in the reported data. Therefore, for consistency in all years, we combine the performance data for regular and alternate assessments as reported in ED*Facts*. In some states, alternate assessments have different performance categories relative to the regular assessment.[17] To ensure that all assessment's proficiency levels match the regular assessment's proficiency levels, we collapse the top categories for any places who have one higher proficiency level than the regular

[17] The ED*Facts* documentation notes proficiency levels by assessment type in years after 2011-12.

assessment. The affected state, subject, grade, and year cases include: Arkansas, math and RLA, grades 3-8, years 2012, 2013, and 2014; Colorado, math and RLA, grades 3-8, years 2012, 2013, and 2014; Iowa, math and RLA, grades 3 through 8, years 2015 and 2018.

Step 3. Cutscore Estimation and Linking

In this step, we use HETOP models and the all-student geographic school district proficiency count data to estimate state-subject-grade-year cutscores on a common scale linked to NAEP after dropping the flagged places in the previous step and also removing any BIE schools for the creation of cutscores. To address practical challenges that can arise in HETOP cutscore estimation for a specific state-subject-grade-year, we:

<u>Rearrange geographic school districts.</u> We reconfigure geographic school districts that meet certain criteria within a state-subject-grade-year in order to improve the HETOP estimation process. First, we combine vectors of counts that have fewer than 20 students into "overflow" groups because estimates based on small sample sizes can be inaccurate. Second, in some vectors with more than 20 students the pattern of counts does not provide enough information to estimate a mean or a standard deviation; we also place these count vectors into the "overflow" group. If the resulting overflow groups have parameters that cannot be estimated via maximum likelihood, they are removed from the data. This reconfiguration allows us to retain the maximum possible number of test scores in the estimation sample for the cutscores. This is important as the linking methods we use later in this step rely on having information about the full population in each state-grade-year-subject.

<u>Constrain geographic school districts.</u> For groups not in the "overflow" group, we always estimate a unique mean. But we can sometimes obtain more precise and identifiable estimates by placing additional constraints on group standard deviation parameters in the HETOP model. We constrain standard deviation parameter estimates for groups that meet the following conditions during estimation:

- There are fewer than 50 student assessment outcomes in a geographic school district.

- There are not sufficient data to estimate both a mean and standard deviation (all student assessment outcomes fall in only two adjacent performance level categories; all student assessment outcomes fall in the top and bottom performance categories; or all student assessment outcomes fall in a single performance level category).

After these data processing steps, we estimate a separate HETOP model for each state-subject-grade-year and save the cutscore estimates. For state-grade-year-subjects with only two proficiency categories, we cannot estimate unique geographic school district standard deviations and instead we use the model with a single, fixed standard deviation parameter (the HOMOP model). We denote the estimated cutscores as $\widehat{c_1}_{fygb}^{state}, \ldots, \widehat{c_{K-1}}_{fygb}^{state}$, for a state f, year y, grade g, and subject b, where the proficiency data are reported in K categories. These cutscores 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 cutscore estimates, denoted $se\left(\widehat{c_k}_{fygb}^{state}\right)\ for\ k = 1, .., K-1$, respectively (Reardon, Shear, Castellano, & Ho, 2017). Note that we do not use the group-specific means or standard deviations that are simultaneously estimated along with the cutscores. See Reardon et al. (2017) and the description in **Step 5** below for additional details about the HETOP model.

To place these cutscores on a common scale across states, grades, and years we use data from the National Assessment of Educational Progress (NAEP). NAEP data provide estimates of 4th and 8th grade test score means and standard deviations for each state on a common scale, denoted $\hat{\mu}(NAEP)_{fygb}$ and $\hat{\sigma}(NAEP)_{fygb}$, respectively, as well as their standard errors.[18] Because NAEP is administered only in 4th and 8th grades in odd-numbered years, we interpolate and extrapolate linearly to obtain estimates of these parameters for grades (3, 5, 6, and 7) and years (2010, 2012, 2014, 2016, and 2018) in which NAEP was not administered. First, within each NAEP-tested year (2009, 2011, 2013, 2015, 2017, and 2019) we linearly interpolate between grades 4 and 8 to grades 5, 6, and 7 and extrapolate to grade 3. Next, for all grades 3-8, we

[18] Note that the NAEP scales are not comparable across math and reading, but they are comparable across states, grades and years within each subject.

linearly interpolate between the odd NAEP-tested years to estimate parameters in 2010, 2012, 2014, 2016, and 2018 using the interpolation/extrapolation formulas here:

$$\begin{aligned}
\hat{\mu}(NAEP)_{fygb} &= \hat{\mu}(NAEP)_{fy4b} + \frac{g-4}{4}\left(\hat{\mu}(NAEP)_{fy8b} - \hat{\mu}(NAEP)_{fy4b}\right), \\
&\qquad \text{for g} \in \{3,5,6,7\} \\
\hat{\mu}(NAEP)_{fygb} &= \frac{1}{2}\left(\hat{\mu}(NAEP)_{f[y-1]gb} + \hat{\mu}(NAEP)_{f[y+1]gb}\right), \\
&\qquad \text{for y} \in \{2010, 2012, 2014, 2016, 2018\}
\end{aligned} \tag{3.1}$$

We do the same to interpolate/extrapolate the state NAEP standard deviations. We also do the same for the national NAEP means and standard deviations; these will be used in standardization. The reported national NAEP means and standard deviations, along with interpolated values, by year and grade, are shown in Table 6.

We then use these state-specific NAEP estimates to place each state's cutscores on the NAEP scale. The methods we use—as well as a set of empirical analyses demonstrating the validity of this approach—are described in more detail by Reardon, Kalogrides, and Ho (2019). We provide a brief summary here. Because geographic school district test score moments and the cutscores are expressed on a state scale with mean 0 and unit variance, the estimated mapping of $\widehat{c_k}_{fygb}^{state} for\ k = 1, \ldots, K-1$ to the NAEP scale is given by Equation (3.2) below, where $\hat{\rho}_{fygb}^{\text{state}}$ is the estimated reliability of the state test. This mapping yields an estimate of the k^{th} cutscore on the NAEP scale; denoted $\widehat{c_k}_{fygb}^{naep}$.

$$\widehat{c_k}_{fygb}^{naep} = \hat{\mu}_{fygb}^{naep} + \frac{\widehat{c_k}_{fygb}^{state}}{\sqrt{\hat{\rho}_{fygb}^{state}}} \cdot \hat{\sigma}_{fygb}^{naep} \tag{3.2}$$

The intuition behind Equation (3.2) is straightforward: cutscores in states with relatively high NAEP averages should be placed higher on the NAEP scale. The reliability term, $\hat{\rho}_{fygb}^{\text{state}}$, in Equation (3.2) is necessary to account for measurement error in state accountability test scores. Note that cutscores on the state scale are expressed in terms of standard deviation units of the state score distribution. The state scale cutscores are biased toward zero due to measurement error. They must be disattenuated during mapping 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, $\hat{\rho}_{fygb}^{\textbf{state}}$. 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.

Finally, we standardize the NAEP-linked cutscores relative to a reference cohort of students. This standardization is accomplished by subtracting the national grade-subject-specific mean and dividing by the national grade-subject-specific standard deviation for a reference cohort. We use the average of the four national cohorts that were in 4th grade in 2009, 2011, 2013, and 2015. We rescale at this step such that all means recovered in Step 5 will be interpretable as an effect size relative to the average of the four national cohorts that were in 4th grade in 2009, 2011, 2013, and 2015.

For each grade, year, and subject we calculate:

$$\begin{aligned} \hat{\mu}(NAEP)_{avg,gb} &= \sum_{Y\in\{2005,2007,2009,2011\}} \frac{1}{4}\hat{\mu}(NAEP)_{(y=Y+g)gb} \\ \hat{\sigma}(NAEP)_{avg,gb} &= \sum_{Y\in\{2005,2007,2009,2011\}} \frac{1}{4}\hat{\sigma}(NAEP)_{(y=Y+g)gb} \end{aligned} \tag{3.3}$$

In Equation (3.3), Y refers to the year in which the cohort was in the spring of kindergarten. For the 2009 4th grade cohort, this is equal to 2005 (or 2009 minus 4).

Then we standardize each cutscore:

$$\widehat{c_k}_{fygb}^{cs} = \frac{\widehat{c_k}_{fygb}^{naep} - \hat{\mu}(NAEP)_{avg,gb}}{\hat{\sigma}(NAEP)_{avg,gb}} \tag{3.4}$$

The resulting cutscores are on the CS scale, standardized to this nationally averaged reference cohort within subject, grade, and year.

PARCC & SBAC Cutscores for BIE Waiver Schools. Once we have scaled cutscores, we take all states, years, and subjects that took the PARCC and SBAC and average their cuts together to get the appropriate PARCC cuts and SBAC cuts to apply to BIE waiver schools. The table of states, years, subjects averaged for this cut creation is in Table 7.

Applying Cuts to BIE schools. Some BIE schools submitted data in different proficiency categories than the proficiency categories reported by the states in which the BIEs reside. In addition, the Navajo Nation had test waivers for the PARCC beginning in SY 2015-16 and Miccosukee Indian School which had a waiver for the SBAC starting in SY 2014-15. For these waiver schools, we use the averaged waiver cuts discussed above. For non-waiver BIE schools, we realign the BIE cuts to match the cuts for the states in which they were located. A few schools whose cuts we could not determine were omitted from SEDA. See Table 8.

Step 4. Selecting Data for Mean Estimation

In Step 5, we estimate a model separately for each unit-subgroup that draws only on the subject-grade-year data for that unit-subgroup. In some subjects, grades, and years, we are less confident in the quality of the unit-subgroup data and do not want to include these in the estimation as it may bias the parameter estimates.[19] We create flags for dropping these cases, which are described below:

The participation rate is less than 95%. In these cases, 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 flag and remove all unit-subgroup-subject-grade-year cases where participation was lower than 95%. Participation is defined as:

$$participation_{urygb} = \frac{numscores_{urygb}}{numenrl_{urygb}}. \tag{4.1}$$

[19] This logic of this data selection differs from the cleaning done in Step 2 to support cutscore estimation. For the cutscore estimation, we wanted to keep as much data as possible in the estimation process because the linking procedure at the end of the Step 3 requires population-based data. Moreover, the cutscores are not particularly sensitive to low-quality data for individual geographic school districts. In contrast, the individual school/geographic district mean/SD estimates will be strongly affected by low quality data (which is defined and described in this section).

This measure can be constructed in the 2012-13 through 2017-18 school years; we do not remove data based on this rule in earlier years. If the participation rate for "all students" is less than 95%, we do not report any estimates for demographic subgroups regardless of whether the subgroup-specific participation rate was greater than 95% because we are concerned about data quality in cases with low overall participation.

Incomplete data reported by student demographic subgroups. There are a small number of cases where 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 male students and 20 test scores for female students. In this case, we would not report the male or female test score means because insufficient test scores were reported by gender. We calculate the reported percentage as:

$$representation_{urygb} = \frac{\sum_r numscores_{urygb}}{numscores_{u,all,ygb}}. \tag{4.2}$$

This measure can be constructed in all years. We flag and remove unit-subgroup-subject-grade-year data from SEDA where there is not at least a 95% representation rate for the gender or race subcategories.

More than 40% of students take alternate assessments. Measurement error may affect unit-subgroup-subject-grade-year cases where over 40% of the students take alternate assessments. These assessments typically differ from the regular assessment and have different proficiency thresholds. This flag can be constructed only in the 2012-13 through 2017-18 school years, so we do not apply this rule in earlier years. We flag and remove places that meet this criterion from SEDA at this step.

Students scored only in the top or only in the bottom proficiency category. We cannot obtain maximum likelihood estimates of unique means for these cases and therefore remove them prior to estimation. This flag can be constructed in every year.

We next flag unit-subgroup-subject-grade-year cells that do not meet the minimum statistical estimation requirements, described below. First, we create a "type flag" for each cell.

It is considered “insufficient” if the case meets one of the following conditions: a) has all observations in a single middle category; b) has all observations in only 2 adjacent categories; c) has only 2 proficiency categories (one cut score); or, d) has all observations in only the top and bottom categories. Otherwise, cells are flagged as “sufficient.” Constraints on the parameter estimates for “insufficient” cells are needed during estimation because they do not provide sufficient data to freely estimate both a mean and a standard deviation. Second, we construct a “size flag.” We flag cells as “small” if they have fewer than 100 test scores; otherwise, cells are flagged as “large”. Each unit-subgroup-subject-grade-year cell, then, has two associated flags.

Prior to estimation, we use the type flag to drop data for entire unit-subgroup-subjects that have no “sufficient” cells. Our estimation methods, described in the next step, cannot produce a standard deviation under this condition. During estimation, the type and size flags are used again to place constraints on the standard deviation estimates for individual cells that are insufficient or small.

Table 9 shows how many cells these drops affect and what is excluded from mean estimation.

Step 5. Estimating Test Score Means

The goal of this step is to estimate the mean and standard deviation of test scores for each subgroup in each unit (schools, geographic school districts, counties, metropolitan areas, commuting zones, or states) across subjects, grades, and years.

Pooled HETOP estimation. In the prior steps, we prepared two pieces of information that we use in estimation: the observed proficiency counts for each unit-subgroup-grade-year-subject from **Step 4** and the estimated cutscores separating the proficiency categories in the associated state-grade-year-subject from **Step 3**. All schools are affiliated with a single state and, thus, a single test and a single set of cutscores. While larger units (districts, metropolitan areas, commuting zones, and states) are also typically affiliated with a single state, test, and set of cutscores, there are a few notable exceptions:

Units that contain BIE schools: As noted above, BIE schools often have different cut scores than the other schools assigned to the geographic school district, metropolitan area, commuting zone, or state. In these cases, the unit is split into two or more components where the schools assigned to each component take the same test and use the same cutscores. For example, we might split a unit into two pieces: unit-regular schools and unit-BIE waiver schools.

Metropolitan Areas or Commuting Zones that cross state lines: A subset of metropolitan areas and commuting zones cross state lines and therefore can be affiliated with several state's cutscores. We split these units into metropolitan area-by-state or commuting zone-by-state components, where the schools assigned to each component took the same test and used the same cutscores.

For both of these cases, we estimate pooled HETOP using data for each subcomponent and the appropriate cut scores, we then aggregate the components after estimation into overall unit estimates.

A pooled HETOP model (Shear & Reardon, 2021; Reardon et al., 2017) to estimate μ_{urygb}^{cs} and σ_{urygb}^{cs}, the mean and standard deviation of achievement on the CS scale for unit u (school, geographic school district, county, commuting zone-by-state, metropolitan area-by-state, or state), subgroup r, year y, grade g, and subject b. As described below, the pooled HETOP model is run separately for each unit-subgroup-subject, but combines data across grades and years when estimating these parameters. Combining data across grades and years allows us to get better estimates of σ_{urygb}^{cs} for years and grades in which sample sizes are small or where the proficiency count data are limited.

We use a pooled HETOP model in order to overcome three practical challenges. The challenges are: 1) in some states, years, and grades, the number of proficiency categories $K = 2$ and there is not sufficient information to estimate both a mean and a standard deviation for each unit-subgroup-grade-year-subject; 2) if $K \geq 3$ but there are sampling zeros because test scores were not observed in all K categories for a particular grade and year, there may not be sufficient information to estimate both a mean and a standard deviation; and 3) when the sample size n_{kurygb} is small, prior simulations (e.g., Reardon et al., 2017; Shear & Reardon, 2021) have shown that estimates of standard deviations can be biased and contain excessive sampling error.

We estimate a pooled HETOP model (Shear & Reardon, 2021) for each unit, separately for each subject and subgroup, by "pooling" data across all available grades and years, and maximizing the joint log likelihood function given by:

$$L = \ln\left[P\left(\mathbf{N}_{urb}\middle|\mathbf{M}_{urb}^{cs}, \mathbf{H}_{urb}^{cs}, \mathbf{C}_{fb}^{cs}\right)\right] = \sum_{y=1}^{Y}\sum_{g=1}^{G}\sum_{k=1}^{K} n_{kurygb} \ln\left(\pi_{kurygb}\right)$$

$$= \sum_{y=1}^{Y}\sum_{g=1}^{G}\sum_{k=1}^{K_{gy}} n_{kurygb} \ln\left(\Phi\left(\frac{\mu_{urygb}^{cs} - c_{k-1_{fygb}}^{cs}}{\exp\left(h_{urb}(g,y)\right)}\right) - \Phi\left(\frac{\mu_{urygb}^{cs} - c_{k_{fygb}}^{cs}}{\exp\left(h_{urb}(g,y)\right)}\right)\right),$$

where $\mathbf{N}_{urb}$ is a matrix of proficiency counts across all available grades (G) and years (Y) for unit u, subgroup r and subject b, $\mathbf{M}_{urb}^{cs}$ is a vector of estimated means across grades and years, $\mathbf{H}_{urb}^{cs}$ is a vector of estimated parameters for the function $h(\quad)$ described below, and $\mathbf{C}_{fb}^{cs}$ is a matrix of cutscores across grades and years. The cutscores are treated as fixed here, using the values estimated in Step 3. We have replaced σ_{urygb}^{cs} in the above equation with $\exp\left(h_{urb}(g,y)\right)$, where $h_{urb}(g,y)$ is a unit-specific function used to model the natural log of the standard deviations as a function of grade and year:

$$h_{urb}(g,y) = \ln\left(\sigma_{urygb}^{cs}\right) = \gamma_{urygb}^{cs}.$$

We do this for two reasons. First, estimating $\gamma_{urygb}^{cs} = \ln\left(\sigma_{urygb}^{cs}\right)$ rather than σ_{urygb}^{cs} directly ensures that the ML estimate will be positive. Second, defining γ_{urygb}^{cs} as a function of grade and year, rather than allowing this value to be unique in each grade and year, defines the pooled HETOP model. If we place no constraints on the model and allow $h_{urb}(g,y) = \gamma_{urbgy}$ to take on a unique value in each grade and year, maximization of this likelihood will result in identical estimates to those obtained by maximizing the likelihood separately for each grade and year.

To leverage the data available across multiple grades and years and overcome the limitations noted above, we define $h_{urb}(g,y)$ in the following way. First, we allow γ_{urygb} to be freely estimated in each grade-year cell that is both "sufficient" and "large", by the flags defined above. For all other grade-year cells, we constrain $h_{urb}(g,y)$ such that the estimate of γ_{urygb} is equal to the mean of the $\hat{\gamma}_{urygb}$ estimates across the freely estimated cells. That is, we estimate a common “pooled” standard deviation across the grades and years in which there are “insufficient” data and/or “small” cell sizes.

More formally, for all years and grades in which $n_{urygb} < 100$ and/or in which there are insufficient data to estimate both a mean and a standard deviation, we constrain $h_{urb}(g, y) = \gamma_{urb}^{cs}$ to be equal, while allowing $h_{urb}(g, y) = \gamma_{urygb}^{cs}$ to be freely estimated in grades and years with at least 100 test scores and sufficient data to estimate both a mean and standard deviation. We further constrain the model such that the “pooled” log standard deviation is equal to the (unweighted) mean of the unconstrained log standard deviations by defining the constraint:

$$\gamma_{urb}^{cs} = \frac{\sum_{g=1}^{G}\sum_{y=1}^{Y}\left(I_{urygb}^{100} \cdot I_{urygb}^{S} \cdot \gamma_{urygb}^{cs}\right)}{\sum_{g=1}^{G}\sum_{y=1}^{Y}\left(I_{urygb}^{100} \cdot I_{urygb}^{S}\right)},$$

where I_{urygb}^{100} is the size indicator flag (equal to 1 if size is “large”) and $I_{urygb}^{S}$is the sufficient data indicator flag (equal to 1 if there are sufficient data). If I_{urygb}^{100} and I_{urygb}^{S} are equal to 1 for all cells in a unit, then we estimate a unique mean and standard deviation for each cell. For all other units, there will be a mix of freely estimated and constrained standard deviation parameters. Recall in **Step 4** that we removed unit-subgroups where $I_{urygb}^{S} = 0$ for all cells because we are unable to estimate a standard deviation parameter.

In sum, the models described here are used to produce ML estimates of μ_{urygb}^{cs} and σ_{urygb}^{cs} (where $\hat{\sigma}_{urygb}^{cs}$ may be constrained to be equal in some grades and years), as well as estimated standard errors $se\left(\hat{\mu}_{urygb}^{cs}\right)$ and $se\left(\hat{\sigma}_{urygb}^{cs}\right)$ and the estimated sampling covariances $cov\left(\hat{\mu}_{urygb}^{cs}, \hat{\sigma}_{urygb}^{cs}\right)$, where unit can be a school, geographic school district, county, commuting zone-by-state, metropolitan area-by-state, or state. The estimates are produced on the CS scale described elsewhere, and can be transformed to other scales, as described in **Step 6**.

Aggregating unit components. For the subset of units where we split the unit into components for pooled HETOP estimation, we need to “re-aggregate” the components into complete unit estimates. The following summary is written for metropolitan areas that cross state lines; however, the same logic can be applied to units serving BIE schools or commuting zones that cross state lines. Suppose there are a set of M metropolitan areas that cross state lines (e.g., have two or more metropolitan area-by-state components). The metropolitan area mean is then estimated as the weighted average of metropolitan area-by-state means across all D_m metropolitan area components in metropolitan area m, computed as

$$\hat{\mu}^{cs}_{mrygb} = \sum_{d=1}^{D_m} p_{dm}\hat{\mu}^{cs}_{mrygb}, \tag{5.1}$$

where p_{dm} is the proportion of metropolitan area m represented by metropolitan area-by-state component d. The estimated metropolitan area standard deviation is estimated as the square root of the estimated total variance between and within metropolitan area-by-state components in the metropolitan area,

$$\hat{\sigma}^{cs}_{mrygb} = \sqrt{\hat{\sigma}^2_{B_m} + \hat{\sigma}^2_{W_m}} \tag{5.2}$$

where $\hat{\sigma}^2_{B_m}$ is the estimated variance between metropolitan area-by-state components in metropolitan area m and $\hat{\sigma}^2_{W_m}$ is the estimated variance within metropolitan area-by-state components in metropolitan area m. The formulas used to estimate $\hat{\sigma}^2_{B_m}$ and $\hat{\sigma}^2_{W_m}$ are based on equations in Reardon et al. (2017). These formulas and formulas for estimating the standard errors of the metropolitan area means and standard deviations, $\hat{\mu}^{cs}_{mrygb}$ and $\hat{\sigma}^{cs}_{mrygb}$, are included in Appendix A1.

Step 6. Creating Additional Reporting Scales

As described in **Step 3**, we standardize the cutscores prior to estimation such that all mean estimates are produced on the CS scale. In this step, we establish a second scale: The **Grade Cohort Standardized (GCS) scale.** We recommend CS-scaled estimates for research purposes and the GCS-scaled estimates for low-stakes reporting to non-research audiences.

Recall that the CS scale is standardized within subject and grade, relative to the average of the four cohorts in our data who were in 4th grade in 2009, 2011, 2013, and 2015. We use the average of four cohorts as our reference group because they provide a stable baseline for comparison. This metric is interpretable as an effect size, relative to the grade-specific standard deviation of student-level scores in this common, average cohort. For example, a district mean of 0.5 on the CS scale indicates that the average student scored approximately one half of a standard deviation higher than the average national reference cohort scored in that same grade. Means reported on the CS scale have an overall average near 0 as expected. Note that this scale retains information about absolute changes over time by relying on the stability of the NAEP scale over time. This scale does not enable absolute comparisons across grades, however.

The GCS scale standardizes the unit means relative to the average difference in NAEP scores between students one grade level apart. The average grade-level difference in national NAEP scores is estimated as the within-cohort grade-level change (separately by subject b), for the average of four cohorts of students in 4th grade in 2009, 2011, 2013, and 2015 (see detail on how $\hat{\mu}(NAEP)_{avg,gb}$ and $\hat{\sigma}(NAEP)_{avg,gb}$ are calculated in **Step 3**). It is denoted $\hat{\gamma}_{avg,b}$:

$$\hat{\gamma}_{avg,b} = \frac{\hat{\mu}(NAEP)_{avg,8b} - \hat{\mu}(NAEP)_{avg,4b}}{4} \tag{6.1}$$

We then identify the linear transformation that sets the grade 4 and 8 averages for this cohort at the “grade level” values of 4 and 8 respectively. Then transform unit means, standard deviations, and their variances accordingly:

$$\hat{\mu}^{gcs}_{urygb} = 4 + \frac{\hat{\mu}(NAEP)_{avg,gb} - \hat{\mu}(NAEP)_{avg,4b}}{\hat{\gamma}_{avg,b}} + \frac{\hat{\sigma}(NAEP)_{avg,gb}}{\hat{\gamma}_{avg,b}} \hat{\mu}^{cs}_{urygb} \tag{6.2}$$

$$var\left(\hat{\mu}^{gcs}_{urygb}\right) = \left(\frac{\hat{\sigma}(NAEP)_{avg,gb}}{\hat{\gamma}_{avg,b}}\right)^2 var\left(\hat{\mu}^{cs}_{urygb}\right) \tag{6.3}$$

$$\hat{\sigma}^{gcs}_{urygb} = \frac{\hat{\sigma}(NAEP)_{avg,gb}}{\hat{\gamma}_{avg,b}} \hat{\sigma}^{cs}_{urygb} \tag{6.4}$$

$$var\left(\hat{\sigma}^{gcs}_{urygb}\right) = \left(\frac{\hat{\sigma}(NAEP)_{avg,gb}}{\hat{\gamma}_{avg,b}}\right)^2 var\left(\hat{\sigma}^{cs}_{urygb}\right) \tag{6.5}$$

Then, $\hat{\mu}^{gcs}_{urygb}$ can be interpreted as the estimated average national “grade-level performance” of students in unit u, subgroup r, year y, grade g, and subject b. For example, if $\hat{\mu}^{gcs}_{ury4b} = 5$, 4th-grade students in unit u, subgroup r, and year y are one grade level ($\hat{\gamma}_b$) above the 4th grade 2009-2015 national average ($\hat{\mu}(NAEP)_{avg,4b}$) 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 assumption that the NAEP scale is stable over time but also that it is vertically linked across grades 4 and 8 and linear between grades. This metric is a simple linear transformation of the NAEP scale, intended to render the NAEP scale more interpretable. For reference, 1 CS scale standard deviation is approximately 3 grade levels. As such, this metric is useful for presenting descriptive research to broad audiences not familiar with interpreting standard deviation units. However, we do not advise it for analyses where the vertical linking across grades and the linear interpolation assumptions are not required nor defensible.

Step 7. Calculating Achievement Gaps

We calculate achievement gap estimates in SEDA 4.1 for all units except schools. Gaps are estimated as the difference in average achievement between subgroups, using the mean estimates from **Steps 5** and **6**. We calculate White-Black (wbg), White-Hispanic (whg), White-Asian (wag), White-Native American (wng), White-Multiracial (wmg),[20] male-female (mfg), and nonECD-ECD (neg) achievement.

In each scale, the unit-subject-grade-year gap is given by the difference in the means, e.g., the White-Black gap is given by:

$$\widehat{wbg}_{uygb}^{x} = \hat{\mu}_{u(r=wht)ygb}^{x} - \hat{\mu}_{u(r=blk)ygb}^{x} \tag{7.1}$$

where x denotes a particular scale (CS, GCS) described in Steps **3** and **7** above. The standard error of the gap is given by:

$$se\left(\widehat{wbg}_{uygb}^{x}\right) = \sqrt{se\left(\hat{\mu}_{u(r=wht)ygb}^{x}\right)^2 + se\left(\hat{\mu}_{u(r=blk)ygb}^{x}\right)^2} \tag{7.2}$$

The gaps can be interpreted similarly to the means in the units defined by the CS and GCS scales. If one or both of the subgroup means needed for the calculation is not available or excluded in a given unit-subject-grade-year, the gap estimate will be missing.

Step 8. Pooled Mean and Gap Estimates

[20] Estimates are calculated, but suppressed prior to public release.

Pooled Mean Estimates. For each unit-subgroup, we have up to 120 subject-grade-year mean estimates (10 years, 6 grades, 2 subjects). We pool the estimates within a unit using precision-weighted random-coefficient models. These models provide more precise estimates of average test scores in a unit (across grades and cohorts), as well as estimates of the grade slope (the average "learning rate" at which scores change across grades, within a cohort) and cohort slope (the average "trend" or rate at which scores change across student cohorts, within a grade). For geographic school districts, counties, metropolitan areas, commuting zones, and states, we provide both subject-specific and overall pooled estimates. For schools we provide only overall pooled estimates.

For each of the following model-types, we estimate a single model drawing on data for all 50 states plus DC to recover pooled estimates. Separate models are estimated for Puerto Rico because only math data is included in SEDA. These models are described at the end of the section.

<u>Subject-Specific Pooled Estimates</u>. This model allows each unit-subgroup to have a subject-specific intercept (average test score), a subject-specific linear grade slope (the "learning rate"), and a subject-specific cohort trend (the "trend"). We fit the following model for geographic districts, counties, metropolitan areas, commuting zones, and states:

$$
\begin{aligned}
\hat{\mu}^{x}_{urygb} = \big[\beta_{0md} &+ \beta_{1md}\big(cohort_{urygb} - mc\big) \\
&+ \beta_{2md}\big(grade_{urygb} - mg\big)\big]M_b \\
&+ \big[\beta_{0ed} + \beta_{1ed}\big(cohort_{urygb} - mc\big) \\
&+ \beta_{2ed}\big(grade_{urygb} - mg\big)\big]E_b + \epsilon_{urygb} + e_{urygb} \\
\beta_{0mu} &= \gamma_{0m0} + v_{0mu} \\
\beta_{1mu} &= \gamma_{1m0} + v_{1mu} \\
\beta_{2mu} &= \gamma_{2m0} + v_{2mu} \\
\beta_{0eu} &= \gamma_{0e0} + v_{0eu} \\
\beta_{1eu} &= \gamma_{1e0} + v_{1eu} \\
\beta_{2eu} &= \gamma_{2e0} + v_{2eu} \\
e_{uygb} \sim N\big(0, \widehat{\omega}^{2}_{uygb}\big);\ \epsilon_{uygb} &\sim N(0, \sigma^2); \begin{bmatrix} v_{0mu} \\ \vdots \\ v_{2eu} \end{bmatrix} \sim MVN(0, \boldsymbol{\tau}^2).
\end{aligned}
\tag{8.1}
$$

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 RLA. $grade$ is the grade-level. We center grade at mg, the middle grade of our sample ($\frac{3+8}{2} = 5.5$). $cohort$ is defined as $year - grade$. In the model we center cohort at mc, which is the middle cohort of our data ($mc = \left(\frac{2018-2009}{2} - \frac{8-3}{2}\right) = (2013.5 - 5.5) = 2008$). e_{uygb} is a normally distributed error term with mean zero and known variance equal to the sampling variance of the mean. The residual variance σ^2 and components of $\boldsymbol{\tau}^2$ are estimated.

In this model, β_{0bu} represents the mean test score in subject b, in unit u, in grade 5.5 for the 2008 cohort. The β_{1bu} parameter indicates the average within-grade (cohort-to-cohort) change per year in average test scores in unit u in subject b; and, the β_{2bu} indicates the average within-cohort change per grade in average test scores in unit u in subject b. If the model is fit using the cohort scale (cs), 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_{0bu}, β_{1bu}, and β_{2bu} will be interpretable relative to this same scale. If the model is fit using the grade cohort scale (gcs), the coefficients will be interpretable as test score differences relative to the average between-grade difference among students.

Overall Pooled Estimates. This model pools data across grades, years, and subjects to produce overall unit estimates. This model allows each unit to have a unit-specific intercept (average test score, pooled over subjects), linear grade slope (the average "learning rate" at which scores change across grades, within a cohort, pooled over subjects), cohort trend (the average "trend," or rate at which scores change across student cohorts, within a grade, pooled over subjects), and the math-RLA difference.

For districts, counties, metropolitan areas, commuting zones, and states, this model is as follows:

$$\hat{y}^x_{uvab} = \beta_{0u} + \beta_{1u}(cohort_{uvab} - mc) + \beta_{2u}(grade_{uvab} - mg) \tag{8.2}$$

$$\beta_{2u} = \gamma_{20} + v_{2u}$$
$$\beta_{3u} = \gamma_{30} + v_{3u}$$
$$e_{uygb} \sim N\left(0, \widehat{\omega}_{uygb}^2\right);\ \epsilon_{uygb} \sim N(0, \sigma^2); \begin{bmatrix} v_{0u} \\ v_{1u} \\ v_{2u} \\ v_{3u} \end{bmatrix} \sim MVN(0, \boldsymbol{\tau}^2).$$

$grade$ is the grade-level. We center grade at mg, the middle grade of our sample ($\frac{3+8}{2} = 5.5$). $cohort$ is defined as $year - grade$. In the model we center cohort at mc, which is the middle cohort of our data ($mc = \left(\frac{2018+2009}{2} - \frac{8+3}{2}\right) = (2013.5 - 5.5) = 2008.$ M_b is an indicator variable equal to 1 if the subject is math; we center M_b at 0.5 so that the intercept represents the average of math and RLA. e_{uygb} is a normally distributed error term with mean zero and known variance equal to the sampling variance of the mean. The residual variance σ^2 and components of $\boldsymbol{\tau}^2$ are estimated.

For schools, we estimate the same general model as shown in Equation (8.2). However, we use different grade and cohort centering. Specifically, we center relative to the middle grade of the school. We define the middle grade as the middle grade for which we have test score estimates from Step 5, regardless of whether or not the school serves additional grades or tested in other grades for which we could not produce estimates. For each school, the middle grade is: $mg_n = \frac{\max(grade)_n + \min(grade)_n}{2}$. Cohort is centered at: $mc_n = \left(\frac{2018+2009}{2} - mg_n\right)$. Note that we use this same middle year, $\frac{2018+2009}{2}$, for cohort centering regardless of whether or not the school was observed over that whole time period. For reference, the grade spans of schools are shown in Table 10.

In this model, β_{0bu} represents the mean test score in unit u in grade 5.5 for the 2008 cohort, averaging across math and RLA. The β_{1bu} parameter indicates the average within-grade (cohort-to-cohort) change per year in average test scores in unit u; and, the β_{2bu} indicates the average within-cohort change per grade in average test scores in unit u. If the model is fit using the cohort scale (cs), 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_{0bu}, β_{1bu}, and β_{2bu} will be interpretable relative to this same scale.

If the model is fit using the grade cohort scale (gcs), the coefficients will be interpretable as test score differences relative to the average between-grade difference among students.

Tables 11a-e and 12a-e report the variance and covariance terms from the $\hat{\boldsymbol{\tau}}^2$ matrices and the estimated reliabilities from estimated by the pooling models for all units.

OLS and EB Estimates from HLM Pooling Models. SEDA 4.1 contains two sets of estimates derived from the pooling models described in Equations (8.1) and (8.2): (1) the OLS estimates of $\beta_{0u}, \dots, \beta_{3u}$, and (2) the Empirical Bayes (EB) shrunken estimates of $\beta_{0u}, \dots, \beta_{3u}$. The OLS estimates are the estimates of $\beta_{0u}, \dots, \beta_{3u}$ that we would get if we took the fitted values from Equation (8.1) or (8.2) and added in the residuals $v_{0u}, \dots, v_{3u}$. That is $\hat{\beta}_{0u}^{ols} = \hat{\gamma}_{00} + \hat{v}_{0u}$, for example. These are unbiased estimates of $\beta_{0u}, \dots, \beta_{3u}$, but they may be noisy in small units. We obtain standard errors of these as described in Appendix A2.

The EB estimates are based on the fitted model as well, but they include the EB shrunken residual. That is, $\hat{\beta}_{0u}^{eb} = \hat{\gamma}_{00} + \hat{v}_{0u}^{eb}$, for example, where $\hat{v}_{0u}^{eb}$ 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 units. We report the square root of the posterior variance of the EB estimates as the standard error of the EB estimate.

In general, the EB estimates (marked as "eb" in the data files) should be used for descriptive purposes and as predictor variables on the right-hand side of a regression model; they are the estimates shown on the website (https://edopportunity.org). They should not be used as outcome variables in a regression model because they are shrunken estimates. Doing so may lead to biased parameter estimates in fitted regression models. The OLS estimates (marked as "ol" in the data files) 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.

Puerto Rico Pooled Estimates.[21] For schools, counties, and metropolitan areas in Puerto Rico, we pool the subgroup-subject-grade-year estimates using a model similar to that in Equation 8.2, but without the centered math term $(M_b - .5)$. We need to remove this term

[21] Note that Puerto Rico estimates are not included in SEDA 4.1 but will be included in a future version.

because we only produce estimates in Puerto Rico in math. The estimates produced in this model are reported as both the pooled overall and pooled subject estimates.

To pool Puerto Rico geographic district and state estimates, we use a simpler model:

$$\hat{y}^x_{uygb} = \beta_0 + \beta_1(cohort_{uygb} - mc) + \beta_2(grade_{uygb} - mg) + \epsilon_{uygb} + e_{uygb}$$
$$e_{dgyb} \sim N(0, \widehat{\omega}^2_{dsgy}), \epsilon_{dgyb} \sim N(0, \sigma^2),$$

All variables are defined as in the above equations. e_{uygb} is a normally distributed error term with mean zero and known variance equal to the sampling variance of the mean. The residual variance σ^2 is estimated. The model is estimated using the -metareg- command in Stata.

The estimates provided in the data files are the coefficients on the intercept, cohort, and grade terms, and their standard errors. Note that this model only produces OLS estimates (not EB shrunken estimates, as discussed in the prior section). In the data files, we report the Puerto Rico OLS estimates as both the "ol" and "eb" variables.

Pooled Gap Estimates. We use the models in Equations 8.1 and 8.2 to pool gaps in districts, counties, metropolitan areas, commuting zones, and states. For example, the pooled White-Black gap parameter estimates in unit u are obtained by 1) computing the gap (the difference in mean White and Black scores) in each unit-grade-year-subject; and, 2) fitting model 8.1 or 8.2 above using these gaps on the left-hand side. However, notably the interpretation of the estimated pooling model coefficients differs. These models recover the average test score gap across grades and years, the rate of the gap changes over grades within cohorts, and the trend in the gap across cohorts within grades.

For users interested in analyzing pooled achievement gaps, it is important to use the pooled gap estimates (described above) rather than taking the difference between pooled estimates of group-specific means.[22] For example, taking the difference of pooled White and Black mean scores will not yield the same White-Black pooled gap estimates as the above

[22] Taking the difference of the pooled means would entail: 1) fitting model 10.1 or 10.2 above using the White mean estimates on the left-hand side; 2) constructing $\hat{\beta}^{ols}_{0u(r=wht)}$ and $\hat{\beta}^{eb}_{0u(r=wht)}$ for White students from the estimates; 3) doing the same with Black student mean scores to construct $\hat{\beta}^{ols}_{0u(r=blk)}$ and $\hat{\beta}^{eb}_{0u(r=blk)}$ for Black students; and then 4) estimating gaps by subtracting $\hat{\beta}^{ols}_{0u(r=wht)} - \hat{\beta}^{ols}_{0u(r=blk)}$ and $\hat{\beta}^{eb}_{0u(r=wht)} - \hat{\beta}^{eb}_{0u(r=blk)}$.

approach because the difference in the EB shrunken means is not generally equal to the EB shrunken mean of the gaps. The latter (using the pooled gaps) is preferred.

Replicating the Pooled Estimates. Notably, we pooled non-noised long-form estimates prior to data suppression, described in **Step 9**. Users will not be able to identically replicate our pooled estimates given two differences between the public long files and the ones used to create the pooled estimates: added noise and fewer estimates (described in more detail below). However, the results should be similar.

Step 9. Suppressing Data for Release

Long Form Files. For the geographic school district, county, commuting zone, metropolitan area, and state long-form files, our agreement with the US Department of Education requires (1) that all reported cells reflect at least 20 students;[23] and (2) that a small amount of random noise is added to each estimate in proportion to the sampling variance of the respective estimate. The added noise is roughly equivalent to randomly removing one student's score from each unit-subgroup-subject-grade-year estimate. These measures are taken 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 unit-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. The SEs of the mean are adjusted to account for the additional error.

In addition, we remove any imprecise individual estimates where the CS scale standard error is greater than 2. Any individual estimate with such a large standard error is too imprecise to use in analysis. We also remove all estimates associated with units that are based on more than 20% alternate assessments across the grades and years in the ED*Facts* data. Table 13 summarizes the cases removed in the district, county, commuting zone, metropolitan area, and state long files.

[23] In the case of gap estimates, we require that each group has at least 20 unique students in each reported cell.

Pooled Files. For a small number of units, there is insufficient data to recover an OLS estimate or SE for a given parameter. While we are able to recover EB estimates for these parameters, we do not release them. Moreover, in the interest of discouraging the over-interpretation of imprecisely estimated parameters, SEDA 4.1 does not report EB or OLS parameter estimates (the average test score, learning rate, or trend in average test score) for a unit when the OLS reliabilities of the individual parameter are below 0.7. We compute the reliability of each OLS parameter estimate $\hat{\beta}_{ku}^{ols}$ as $\frac{\hat{\tau}_k^2}{\hat{\tau}_k^2+\hat{V}_{ku}}$, where $\hat{\tau}_k^2$ is the k^{th} diagonal element of the $\hat{\boldsymbol{\tau}}^{\mathbf{2}}$ matrix (the estimated true variance of β_{kd}) and $\hat{V}_{ku}$ is the square of the estimated standard error of $\hat{\beta}_{ku}^{ols}$. That is, we do not report $\hat{\beta}_{ku}^{ols}$ if $\hat{V}_{ku} > \frac{3}{7}\hat{\tau}_k^2$. For subgroups, we use the same procedure and for gaps we additionally suppress the estimate if either or both of the main subgroup is suppressed (e.g., suppress White-Black gap if either White and/or Black estimates are suppressed). We use the standard error threshold determined for all students to censor estimates rather than calculate a subgroup-specific threshold. In the case where the reliability of the intercept (average test score) for a unit is less than 0.7, we do not report any parameter estimates for that unit. We also remove all estimates associated with units that are based on more than 20% alternate assessments across the grades and years in the ED*Facts* data. Table 14 summarizes the cases removed in the school, district, county, commuting zone, metropolitan area, and state pooled files.

II.F. Additional Notes

Gender Mean and Gap Estimates. Recent research reported by Reardon, Kalogrides, et al. (2019) suggests that the magnitude of gender achievement gaps can be impacted by the proportion of test items that are multiple-choice versus constructed-response. As a result, differences in gender gaps across states (or across time when a state changes the format of its test) may confound true differences in achievement with differences in the format of the state test used to measure achievement. See Reardon, Fahle, et al. (2019) for a description of an analytic strategy that can be used to adjust for these potential effects.

Charter School Estimates. Research reported in Reardon, Papay, Kilbride, et al. (2019) shows that estimates of student learning rates (the coefficient on the “grade” term in the pooling models in Step 8) are generally unbiased and reliable, except when student mobility in and out of schools is high. In the three states' data they used, student mobility was higher, on average, in small schools and districts, schools with long grade spans, and in charter schools. In addition, in very small schools and charter schools, the estimated grade slope is biased upwards, as a result of differential mobility (more lower-achieving students leave schools than enter). As a result, we recommend that users interpret the school level grade slopes with some caution, particularly for small schools, schools that span 4 or more of the grades from 3 to 8, and charter schools. Moreover, users should be cautious in comparing grade slopes in charter schools to those of traditional public schools, given the evidence of systematic upward bias in the charter sector estimates.

II.G. How to Obtain Values Shown on edopportunity.org

In order to replicate the values shown on the EdOpportunity website from the downloadable data files (csv/Stata), please refer to the steps below:

1. Determine what geography you would like to focus on and download the corresponding GCS Pooled File: **seda_[geography]_pool_gcs_[seda_version]**, where **geography** is *school*, *geodist*, or *county* and where **seda_version** is a number, e.g., *4.1* for version 4.1.
2. To get average test score, do the following calculation: gcs_mn_avg_eb – gradecenter
3. To get the learning rate, do the following calculation: gcs_mn_grd_eb – 1
4. To get the trend, use gcs_mn_coh_eb
5. In order to get the corresponding Margins of Error (we use a 95% confidence interval) on our site, multiply the Standard Errors on the variable you would like to use by 1.959964 (the inverse cumulative standard normal distribution of .975). For example, to get the average test score margin of error, that would be gcs_mn_avg_eb_se * 1.959964. If you are working in Stata, you could also use gcs_mn_avg_eb_se * invnormal(.975).

III. Covariate Data Construction

SEDA 4.1 contains CCD and ACS data that have been curated for use with the school, geographic school district, county, metropolitan area, and state achievement data.

III.A. ACS Data and SES Composite Construction

For districts, counties, metropolitan areas and states we use data from the ACS to construct measures of median family income, proportion of adults with a bachelor's degree or higher, proportion of adults that are unemployed, the household poverty rate, the proportion of households receiving SNAP benefits, and the proportion of households with children that are headed by a single mother. We also combine these measures to construct a single socioeconomic status composite.

ACS data are available as 5-year pooled samples, from which we use samples from 2005-2009 through 2014-2018. The samples we use here reflect data for the total population of residents in each unit. In select years, district-level tabulations are also available for families who live in each school district in the U.S and who have children enrolled in public school. However, the most recent sample of this data that has all of the information we need is the 5-year 2007-2011 sample. We prefer to use the total population tabulation data from more recent years. We have compared measures constructed using the total population samples and the relevant children enrolled in public schools samples in years where both samples are available and the measures are highly correlated ($r > 0.99$) and not sensitive to which sample we use.

The construction of our derived measures from the ACS data occurs in a variety of steps, which we describe below. Our derivation of these measures is complicated by the fact that we use the ACS-reported margins of error to compute Empirical Bayes shrunken versions of our key ACS measures. The shrunken measures help account for attenuation bias that results from the fact that smaller units' measures include more measurement error due to smaller sample sizes. Appendix B2 describes the problems of measurement error and attenuation bias in detail. Below we describe the steps we take to create our derived measures from the raw ACS data. Note that we do not compute standard errors or Empirical Bayes shrunken versions of state-level measures. State samples are sufficiently large as to not be concerned about measurement error due to small samples. Therefore, many of the steps described below only refer to district, county, and metropolitan area data.

Step 1: We download and clean the raw ACS data for each year and unit, saving the measures of interest along with their margins of error. We use data from the 2005-2009, 2006-

2010, 2007-2011, 2008-2012, 2009-2013, 2010-2014, 2011-2015, 2012-2016, 2013-2017, and 2014-2018 samples. In Appendix B1 we provide a list of the raw ACS data tables we downloaded and use to compute each derived measure.

Step 2: Some of our derived measures require combining various fields from ACS in order to compute our desired metric. For example, in order to compute the proportion of adults with a bachelor's degree or higher we sum the number of men with a bachelor's degree, a master's degree or a professional degree with the number of women with a bachelor's degree, a master's degree or a professional degree and divide that sum by the total number of adults in the unit. Each of these component measures is reported with its own margin of error in the raw ACS data. We use the margins of error from each component measure to generate a single standard error for the combined bachelor's degree attainment rate variable (and do the same for all 6 socioeconomic measures that make up the SES composite). Appendix B3 describes our methodology for computing the sampling variance of sums of ACS variables in detail.

Step 3: After constructing the 6 SES measures and their standard errors we impute some missing data using Stata's –**mi impute chained–** routine, which fills in missing values iteratively by using chained equations. We reshape the data from long (one observation for each unit and race group [all, White, Black, Hispanic, Asian, and Native American] in each year) to wide (one observation for each unit and a separate variable for each of the 6 SES by race measures in each year). We use both the 6 SES measures and their standard errors in the imputation model as well as the total population count in each unit. The imputation model, therefore, includes median income, proportion of adults with a bachelor's degree or higher, child poverty rate, SNAP receipt rate, single mother headed household rate, and unemployment rate for each race group (all, White, Black, Hispanic, Asian, and Native American) in each of 10-year spans for both the estimates and their standard errors. We estimate the imputation model 5 times.

Step 4: Next, we use the imputed data to compute the SES composite. This is done 5 times for each imputed data set and then we take the average. 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, child poverty rate, SNAP receipt rate, single mother headed household rate, and employment rate for adults ages 16-64.

We use the logarithm of median income in these computations. We calculate the component loadings by conducting the analysis in 2008-2012 at the geographic school district level and weighting by geographic school district enrollment. We then use the loadings from this principal component analysis to calculate SES composite values for different subgroups, years and units. Note that only observations without any imputed ACS data are used in the computation of the factor weights.

Table 15 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 geographic school district SES composite score from the 6 standardized indicator variables in 2008-2012, resulting in an SES composite that has an enrollment-weighted mean of 0 and standard deviation of 1 across all geographic school districts in 2008-2012 without any imputed data. The "unstandardized loadings" are rescaled 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.

To provide context for interpreting values of the SES composite, Table 16 reports average values of the indicator variables at different values of the SES composite.

Step 5: The next step is to construct a standard error of the SES composite. We discuss our methodology in detail in Appendix B4.

Step 6: The final step is to do the Empirical Bayes shrinkage of the SES composites as well as for each of the 6 SES measures that go into making the composite. In addition to the time-varying versions of the SES composite, we also create an SES composite that is the average of SES in the 2005-2009 and 2014-2018 ACS (i.e., using the first and last years' ACS samples). The shrinkage is done using a random effects meta-analysis regression model weighted by the standard error of each measure. Note that we use the same suppression rules described on page 38 to suppress SES estimates with large standard errors. We do not report group specific SES if the reliability of the all group estimate for a given geography is below .70. Furthermore, if either the White or Black/Hispanic/Asian/Native American group specific SES is censored by this first rule then we censor the SES gap between the two groups as well. See page 38 for more details of the censoring rules.

III.B. Common Core of Data Imputation

School-level data from the CCD are available from Fall 1987 until Fall 2018. 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 geographic school district, county, metropolitan area, or state level. The imputation model includes school-level data from the 1991-92 through 2018-19 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 2017-18 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/manuals/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 2017-18. In Ohio, 8% of schools were missing CCD free lunch data in 4 or more of the ED*Facts* years. In Oregon, 5% of schools were missing CCD free lunch data in 4 or more of the 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–there is one variable for each year for any given measure (i.e., total enrollment 1991, total enrollment 1992, total enrollment 1993, …, total enrollment 2018) 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. We then aggregate this imputed school by year data file to different geographic levels, computing our desired measures.

IV. Versioning and Publication

New or revised data will be posted periodically to the SEDA website. SEDA updates that contain substantially new information are labeled as a new version (e.g., V1.0, V2.0, etc.). 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 4.1 makes the following additions to data contained in SEDA 4.1, we now release:

- BIE School Estimates
- Native-American (NAM) and White-Native American gap (WNG) estimates

SEDA 4.1 makes the following modifications to the procedures used in SEDA 4.0:

- In Step 4, "selecting the data for mean estimation" we no longer drop if a unit-subgroup has only one "insufficient" or "small" unit-subgroup-subject-grade-year case.
- Virtual Schools are now defined using the most recent information. That is for 4.1, a school is flagged as virtual if it most recently observed as a virtual school; whereas, for 4.0 we flagged schools as virtual if they were at any point identified as a virtual school.
- 4.1 uses updated imputed test reliabilities than 4.0
- 4.1 has different suppression rules for pooled output for gaps in public files than 4.0
- In SEDA 4.1 we create a crosswalk that links units to stable definitions over time using 2018 definitions of districts and counties and 2013 definitions of metros. This accounts for changing district, county, and metro definitions over time. We use this new geographic crosswalk and take weighted averages of measures from the American Community Survey to aggregate them units that have the same geographic boundaries over time.

References

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender Achievement Gaps in U.S. School Districts. *American Educational Research Journal*, 000283121984382. https://doi.org/10.3102/0002831219843824

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2019). Validation Methods for Aggregate-Level Test Scale Linking: A Case Study Mapping School District Test Score Distributions to a Common Scale. *Journal of Educational and Behavioral Statistics*, 107699861987408. http://doi.org/10.3102/1076998619874089

Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A., & Zárate, R. C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. Educational Researcher, 1–11. https://doi.org/10.3102/0013189X18762105

Reardon, S.F., Papay, J.P., Kilbride, T., Strunk, K.O., Cowen, J., An, L., & Donohue, K. (2019). Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive. (CEPA Working Paper No.19-08). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp19-08

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

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2019). Validation Methods for Aggregate-Level Test Scale Linking: A Case Study Mapping School District Test Score Distributions to a Common Scale. *Journal of Educational and Behavioral Statistics*, 107699861987408. http://doi.org/10.3102/1076998619874089

Shear, B. R., & Reardon, S. F. (2021). Using Pooled Heteroskedastic Ordered Probit Models to Improve Small-Sample Estimates of Latent Test Score Distributions. *Journal of Educational and Behavioral Statistics*, *46*(1), 3–33. https://doi.org/10.3102/1076998620922919

Tables

Table 1. Test Score Files

Test Score Estimations: Means and Achievement Gaps

File Name	Form	Metric	Unit					Disaggregated by				Subgroups							Download	
													Means			Gaps				
			School	District	County	Metro	Comm Zone	State	Year	Grade	Subject	All	Race	Gender	ECD	Race	Gender	ECD		
seda_school_pool_cs_4.1	Pooled	CS	X									X							Stata	CSV
seda_school_pool_gcs_4.1	Pooled	GCS	X									X							Stata	CSV
seda_geodist_long_cs_4.1	Long	CS		X					X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_geodist_long_gcs_4.1	Long	GCS		X					X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_geodist_poolsub_cs_4.1	Pooled	CS		X							X	X	X	X	X	X	X	X	Stata	CSV
seda_geodist_poolsub_gcs_4.1	Pooled	GCS		X							X	X	X	X	X	X	X	X	Stata	CSV
seda_geodist_pool_gcs_4.1	Pooled	CS		X								X	X	X	X	X	X	X	Stata	CSV
seda_geodist_pool_cs_4.1	Pooled	GCS		X								X	X	X	X	X	X	X	Stata	CSV
seda_county_long_cs_4.1	Long	CS			X				X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_county_long_gcs_4.1	Long	GCS			X				X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_county_poolsub_cs_4.1	Pooled	CS			X						X	X	X	X	X	X	X	X	Stata	CSV
seda_county_poolsub_gcs_4.1	Pooled	GCS			X						X	X	X	X	X	X	X	X	Stata	CSV
seda_county_pool_cs_4.1	Pooled	CS			X							X	X	X	X	X	X	X	Stata	CSV
seda_county_pool_gcs_4.1	Pooled	GCS			X							X	X	X	X	X	X	X	Stata	CSV
seda_metro_long_cs_4.1	Long	CS				X			X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_metro_long_gcs_4.1	Long	GCS				X			X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_metro_poolsub_cs_4.1	Pooled	CS				X					X	X	X	X	X	X	X	X	Stata	CSV
seda_metro_poolsub_gcs_4.1	Pooled	GCS				X					X	X	X	X	X	X	X	X	Stata	CSV
seda_metro_pool_cs_4.1	Pooled	CS				X						X	X	X	X	X	X	X	Stata	CSV
seda_metro_pool_gcs_4.1	Pooled	GCS				X						X	X	X	X	X	X	X	Stata	CSV
seda_commzone_long_cs_4.1	Long	CS					X		X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_commzone_long_gcs_4.1	Long	GCS					X		X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_commzone_poolsub_cs_4.1	Pooled	CS					X				X	X	X	X	X	X	X	X	Stata	CSV
seda_commzone_poolsub_gcs_4.1	Pooled	GCS					X				X	X	X	X	X	X	X	X	Stata	CSV
seda_commzone_pool_cs_4.1	Pooled	CS					X					X	X	X	X	X	X	X	Stata	CSV
seda_commzone_pool_gcs_4.1	Pooled	GCS					X					X	X	X	X	X	X	X	Stata	CSV
seda_state_long_cs_4.1	Long	CS						X	X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_state_long_gcs_4.1	Long	GCS						X	X	X	X	X	X	X	X	X	X	X	Stata	CSV
seda_state_poolsub_cs_4.1	Pooled	CS						X			X	X	X	X	X	X	X	X	Stata	CSV
seda_state_poolsub_gcs_4.1	Pooled	GCS						X			X	X	X	X	X	X	X	X	Stata	CSV
seda_state_pool_cs_4.1	Pooled	CS						X				X	X	X	X	X	X	X	Stata	CSV
seda_state_pool_gcs_4.1	Pooled	GCS						X				X	X	X	X	X	X	X	Stata	CSV

Notes:

Metric:	CS = Cohort Scale; GCS = Grade Scale
Unit	*Geodist = Geographically Defined School District; Metro = Metropolitan Statistical Area; Commzone = Commuting Zone*
Academic Years:	2008/09 – 2017/18
Grades:	3 – 8
Subjects:	Math, RLA
Race:	Asian, Black, Hispanic, Multiracial, Native American, White
Race Gaps:	White-Asian, White-Black, White-Hispanic, White-Multiracial, White-Native American
Gender:	male, female
Gender Gaps:	male-female
ECD:	economically disadvantaged, not disadvantaged (as defined by states)
ECD Gaps:	not disadvantaged-economically disadvantaged

Table 2. Covariate Data Files

File Name	Form	Disaggregated by			Download	
		Unit	Year	Grade		
seda_cov_school_pool_4.1	Pooled	X			Stata	CSV
seda_cov_school_poolyr_4.1	Pooled	X	X		Stata	CSV
seda_cov_geodist_long_4.1	Long	X	X	X	Stata	CSV
seda_cov_geodist_poolyr_4.1	Pooled	X	X		Stata	CSV
seda_cov_geodist_pool_4.1	Pooled	X			Stata	CSV
seda_cov_county_long_4.1	Long	X	X	X	Stata	CSV
seda_cov_county_poolyr_4.1	Pooled	X	X		Stata	CSV
seda_cov_county_pool_4.1	Pooled	X			Stata	CSV
seda_cov_metro_long_4.1	Long	X	X	X	Stata	CSV
seda_cov_metro_poolyr_4.1	Pooled	X	X		Stata	CSV
seda_cov_metro_pool_4.1	Pooled	X			Stata	CSV
seda_cov_state_long_4.1	Long	X	X	X	Stata	CSV
seda_cov_state_poolyr_4.1	Pooled	X	X		Stata	CSV
seda_cov_state_pool_4.1	Pooled	X			Stata	CSV

Table 3. Example ED*Facts* Data Structure

FIPS	NCESSCH	Subgroup	Subject	Grade	Year	Number of students scoring at…			
						Level 1	Level 2	Level 3	Level 4
99	997777755555	All Students	Math	3	2009	26	87	185	32
99	997777755555	All Students	RLA	3	2009	13	102	195	20
99	997777755556	All Students	Math	3	2009	35	238	192	7
99	997777755556	All Students	RLA	3	2009	7	278	187	0

Note. The data shown in this table are not real.

Table 4. State-Subject-Year-Grade Data Not Included in SEDA 4.1

State Abbreviation	Reason for Missing	Years/Subjects/Grades Excluded
AK	No EdFacts Data	2016
AK	Participation Drop	2017: Math 3-8, RLA 3-8; 2018: Math 3-8, RLA 3-8
AZ	Participation Drop	2018: Math 3-8, RLA 3-8
AR	Manual Drop	2009: Math 8; 2010: Math 8; 2015: Math 8
CA	Manual Drop	2009: Math 7-8; 2010: Math 7-8; 2011: Math 7-8; 2012: Math 7-8; 2013: Math 7-8
CA	Participation Drop	2014: Math 3-8, RLA 3-8
CO	Manual Drop	2009: Math 3-8, RLA 3-8; 2010: Math 3-8, RLA 3-8; 2011: Math 3-8, RLA 3-8
CO	Participation Drop	2015: Math 4-8, RLA 4-8; 2016: Math 5-8, RLA 5-8; 2017: Math 5-8, RLA 5-8; 2018: Math 6-8, RLA 6-8
CT	Participation Drop	2014: Math 3-8, RLA 3-8
DC	Manual Drop	2011: RLA 3
DC	Participation Drop	2015: Math 8, RLA 8; 2017: Math 3,5-8, RLA 3-8
FL	Participation Drop	2014: Math 3-8
ID	Participation Drop	2014: Math 3-8, RLA 3-8
IL	Participation Drop	2015: Math 8, RLA 8
KS	No EdFacts Data	2014
LA	Participation Drop	2018: Math 3-4, RLA 3-4
ME	Participation Drop	2015: Math 6-8, RLA 7-8
MD	Participation Drop	2014: Math 3-4,6-7, RLA 3-4,6-7
MD	No EdFacts Data	2018
MA	Manual Drop	2015: Math 3-8, RLA 3-8
MO	Manual Drop	2013: Math 8; 2014: Math 8; 2015: Math 8; 2016: Math 8; 2018: Math 8
MO	Participation Drop	2017: Math 8
MT	Participation Drop	2014: Math 3-8, RLA 3-8; 2015: Math 3-8, RLA 3-8; 2017: RLA 3-5
NE	Manual Drop	2009: Math 3-8, RLA 3-8; 2010: Math 3-8
NV	Participation Drop	2014: Math 3-8, RLA 3-8
NV	No EdFacts Data	2015
NH	Participation Drop	2015: Math 8, RLA 8; 2016: RLA 8; 2017: Math 8, RLA 8
NJ	Participation Drop	2015: Math 3-8, RLA 3-8; 2016: Math 7-8, RLA 7-8
NM	Manual Drop	2015: RLA 3-8; 2016: RLA 3-8; 2017: RLA 3-8
NM	No EdFacts Data	2018: RLA
NY	Participation Drop	2014: Math 6-8, RLA 6-8; 2015: Math 3-8, RLA 3-8; 2016: Math 3-8, RLA 3-8; 2017: Math 3-8, RLA 3-8; 2018: Math 3-8, RLA 3-8
ND	Manual Drop	2015: Math 3-4, RLA 3-4
ND	Participation Drop	2015: Math 6-8, RLA 5-8
OH	Manual Drop	2014: Math 3-8, RLA 3-8; 2015: Math 8
OK	Manual Drop	2012: Math 8
OK	Participation Drop	2013: Math 8
OR	Participation Drop	2014: Math 3-8, RLA 3-8; 2017: Math 7-8, RLA 8; 2018: Math 7-8, RLA 7-8
RI	Participation Drop	2015: Math 6-8, RLA 5-8
SD	Participation Drop	2014: Math 3-8, RLA 3-8
TN	Manual Drop	2014: Math 8; 2016: Math 3-7, RLA 3-7
TN	Participation Drop	2016: Math 8, RLA 8
TX	Manual Drop	2012: Math 7-8; 2013: Math 7-8; 2014: Math 7-8; 2015: Math 7-8; 2016: Math 7-8; 2017: Math 7-8; 2018: Math 7-8
UT	Manual Drop	2009: Math 8; 2010: Math 8; 2011: Math 8; 2012: Math 8; 2013: Math 8
UT	Participation Drop	2016: Math 7-8, RLA 8; 2017: Math 5-8, RLA 5-8; 2018: Math 5-8, RLA 5-8
VT	Participation Drop	2014: Math 3-8, RLA 3-8
VT	No EdFacts Data	2018
VA	Manual Drop	2009: Math 5-8; 2010: Math 5-8; 2011: Math 5-8; 2012: Math 5-8; 2013: Math 5-8; 2014: Math 5-8; 2015: Math 5-8; 2016: Math 5-8; 2017: Math 7-8
VA	No EdFacts Data	2018: Math
WA	Participation Drop	2014: Math 3-8, RLA 3-8; 2015: Math 3-8, RLA 3-8; 2016: Math 3-8, RLA 3-8; 2017: Math 3-8, RLA 3-8
WV	Participation Drop	2014: Math 3-7; 2016: Math 3-5
WV	Manual Drop	2014: Math 8
WY	No EdFacts Data	2010
WY	Manual Drop	2012: Math 3-8, RLA 3-8
WY	Participation Drop	2013: Math 3-8; 2014: Math 3,7-8, RLA 3-8
BI	No EdFacts Data	2013; 2014; 2015; 2018
PR	No NAEP Data	2009: Math 3-8, RLA 3-8; 2010: Math 3-8, RLA 3-8; 2011: RLA 3-8; 2012: RLA 3-8; 2013: RLA 3-8; 2014: RLA 3-8; 2015: RLA 3-8; 2016: RLA 3-8; 2017: RLA 3-8; 2018: RLA 3-8

Note. Year is spring of year, so 2018 is the 2017-18 school year.

Table 5. Double Counting of Bureau of Indian Education Schools

NCESSCH Dropped from ED*Facts*	School Name	NCESSCH Kept in SEDA	Note
590002500172	TURTLE MOUNTAIN COMMUNITY MIDDLE SCHOOL	380253000751	CCD Task Force
590010600170	TURTLE MOUNTAIN COMMUNITY ELEMENTARY SCHOOL	380253000750	CCD Task Force
590007700173	TWIN BUTTES ELEMENTARY SCHOOL	381860000757	CCD Task Force
590011700174	WHITE SHIELD ELEMENTARY SCHOOL	381968000807	CCD Task Force
590018700086	NAH TAH WAHSH PUBLIC SCHOOL ACADEMY	260010300646	CCD Task Force
590005400080	JOSEPH K LUMSDEN BAHWETING ANISHNABE ACADEMY	260007100492	CCD Task Force
590018900167	MANDAREE ELEMENTARY SCHOOL	381185000747	SEDA Team
590018900167	MANDAREE HIGH SCHOOL	381185000006	SEDA Team
230006400664	INDIAN TOWNSHIP SCHOOL	590004200052	SEDA Team
370015302953	CHEROKEE ELEMENTARY	590006600044	SEDA Team
230006900630	BEATRICE RAFFERTY SCHOOL	590013700042	SEDA Team
230006600671	INDIAN ISLAND SCHOOL	590016000051	SEDA Team

Note. CCD Task Force indicates the school was listed in the CCD's Documentation (Double Counting of Bureau of Indian Education Schools). SEDA Team indicates that it was determined by looking at school coordinates, assessments received by grade, and school contact information since the CCD task force only began with SY 2009-10.

Table 6. NAEP Means and Standard Deviations by Year and Grade.

	Grade	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
		Reading/English Language Arts											
Means	8	259.1	260.1	260.9	261.7	263.3	264.8	263.9	263.0	263.5	264.0	262.3	260.6
	7	248.5	249.3	250.0	250.7	252.1	253.4	252.8	252.3	252.6	252.9	251.4	250.0
	6	237.9	238.6	239.2	239.8	240.9	241.9	241.7	241.5	241.6	241.8	240.6	239.4
	5	227.3	227.8	228.3	228.8	229.7	230.5	230.6	230.8	230.7	230.6	229.7	228.7
	4	216.7	217.0	217.4	217.8	218.5	219.1	219.6	220.0	219.8	219.5	218.8	218.1
	3	206.1	206.2	206.5	206.8	207.3	207.7	208.5	209.3	208.8	208.4	207.9	207.5
SDs	8	36.8	36.3	36.0	35.8	35.5	35.3	35.5	35.8	36.4	36.9	38.1	39.3
	7	37.1	36.6	36.5	36.3	36.2	36.1	36.2	36.3	36.9	37.4	38.4	39.4
	6	37.5	37.0	36.9	36.9	36.9	36.9	36.9	36.9	37.4	38.0	38.8	39.5
	5	37.9	37.4	37.4	37.4	37.5	37.6	37.5	37.4	38.0	38.5	39.1	39.7
	4	38.2	37.7	37.8	37.9	38.2	38.4	38.2	38.0	38.5	39.0	39.4	39.8
	3	38.6	38.1	38.2	38.4	38.8	39.2	38.9	38.6	39.0	39.5	39.7	39.9
		Math											
	Grade	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Means	8	279.1	280.1	280.8	281.4	282.1	282.7	281.6	280.4	280.6	280.9	280.4	279.9
	7	268.8	269.6	270.2	270.8	271.5	272.1	271.1	270.1	270.2	270.2	269.9	269.6
	6	258.5	259.1	259.7	260.3	260.9	261.6	260.7	259.8	259.7	259.6	259.5	259.4
	5	248.2	248.6	249.2	249.7	250.4	251.0	250.2	249.4	249.2	249.0	249.0	249.1
	4	238.0	238.1	238.7	239.2	239.8	240.4	239.8	239.1	238.7	238.3	238.6	238.9
	3	227.7	227.6	228.1	228.6	229.2	229.8	229.3	228.8	228.2	227.7	228.1	228.6
SDs	8	37.7	37.6	37.3	37.1	37.1	37.1	37.3	37.5	38.5	39.6	40.1	40.6
	7	35.7	35.6	35.4	35.2	35.3	35.4	35.6	35.8	36.8	37.8	38.2	38.7
	6	33.8	33.7	33.5	33.4	33.5	33.7	33.8	34.0	35.0	35.9	36.3	36.8
	5	31.8	31.7	31.6	31.6	31.8	32.0	32.1	32.3	33.2	34.1	34.5	34.8
	4	29.8	29.8	29.8	29.7	30.0	30.3	30.4	30.5	31.4	32.3	32.6	32.9
	3	27.9	27.8	27.9	27.9	28.2	28.6	28.7	28.8	29.6	30.5	30.7	31.0

Note. Table shows the interpolated national NAEP estimates. We use the expanded population estimates, which may differ slightly from those reported publicly on the website.

Table 7. PARCC and SBAC States and Years Used to Create BIE Waiver Cut Scores

Test	Year	States
	2014	HI
	2015	CA,CT,DE,HI,ID,ME,MI,MO,MT,NV,NH,ND,OR,SD,VT,WA,WV
	2016	CA,CT,DE,HI,ID,MI,MT,NV,NH,ND,OR,SD,VT,WA,WV
	2017	CA,CT,DE,HI,ID,MI,MT,NV,NH,ND,OR,SD,VT,WA,WV
SBAC	2018	CA,CT,DE,HI,ID,MT,NV,NH,OR,SD,WA,WV
	2015	AR,CO,DC,IL,MD,MS,NJ,RI
	2016	CO,DC,IL,MD,NJ,RI
	2017	CO,DC,IL,MD,NJ,RI
PARCC	2018	CO,DC,IL,MD,NJ,NM

Table 8. BIE Schools Dropped from SEDA

State Abbreviation	Year	NCES School ID	School Name
MS	2016	590005300056	Standing Pine Elementary School
MS	2017	590005300056	Standing Pine Elementary School
MS	2016	590007800048	Choctaw Central Middle School
MS	2017	590007800048	Choctaw Central Middle School
MS	2016	590010000057	Tucker Elementary School
MS	2017	590010000057	Tucker Elementary School
MS	2016	590011100050	Conehatta Elementary School
MS	2017	590011100050	Conehatta Elementary School
MS	2016	590012300054	Pearl River Elementary School
MS	2017	590012300054	Pearl River Elementary School
MS	2016	590017200055	Red Water Elementary School
MS	2017	590017200055	Red Water Elementary School
MS	2016	590019400043	Bogue Chitto Elementary School
MS	2017	590019400043	Bogue Chitto Elementary School
WI	2012	590010400087	Lac Courte Oreilles Ojibwa School
WI	2012	590011400090	Oneida Nation School
WI	2012	590014400088	Menominee Tribal School

Note. Year indicates the spring of the school year.

Table 9. Subject-Grade-Year Cases Removed Pre-Estimation

Cases Dropped Pre-Estimation	sedasch	sedalea	sedacounty	sedametro	sedacz	sedafips
Virtual Schools	34,274 (0.71%)					
Manual Drops	176,945 (3.68%)	Data is aggregated from schools. Drops already incorporated.				
State Participation < 95%	276,662 (5.75%)					
All Participation < 95%	147,761 (3.07%)	278,065 (2.39%)	36,990 (1.12%)	10,625 (0.92%)	7,226 (0.85%)	10 (0.02%)
Subgroup participation <95%	147,761 (3.07%)	285,214 (2.45%)	53,869 (1.63%)	19,564 (1.70%)	12,130 (1.42%)	456 (0.72%)
Representation < 95%	0 (0.00%)	94,070 (0.81%)	34,060 (1.03%)	11,865 (1.03%)	7,342 (0.86%)	583 (0.92%)
Alternative Assessments > 40%	42,688 (0.89%)	20,721 (0.18%)	523 (0.02%)	0 (0.00%)	7 (0.00%)	0 (0.00%)
Pathological Cases	90,278 (1.88%)	776,480 (6.66%)	148,575 (4.50%)	22,557 (1.96%)	17,863 (2.09%)	413 (0.65%)
Total Cases Dropped for Any Reasor	731,481 (15.19%)	1,269,202 (10.89%)	248,866 (7.54%)	57,752 (5.02%)	39,765 (4.65%)	1,445 (2.29%)
Total Cases Not Dropped	4,082,479 (84.81%)	10,382,669 (89.11%)	3,051,522 (92.46%)	1,093,031 (94.98%)	814,804 (95.35%)	61,668 (97.71%)
Total Number of Cases	4,813,960 (100.00%)	11,651,871 (100.00%	3,300,388 (100.00%)	1,150,783 (100.00%)	854,569 (100.00%)	63,113 (100.00%)

Table 10. Grade Spans by School

Grade Span	Schools (N)
3 - 3	1,325
3 - 4	3,283
3 - 5	26,275
3 - 6	14,121
3 - 7	1,407
3 - 8	11,537
4 - 4	30
4 - 5	268
4 - 6	423
4 - 7	99
4 - 8	868
5 - 5	58
5 - 6	450
5 - 7	139
5 - 8	3,002
6 - 6	147
6 - 7	142
6 - 8	12,965
7 - 7	66
7 - 8	5,126
8 - 8	764

Note. Grade span is “start grade”-“end grade.”

Table 11a. State Variances and Covariances

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedafips	cs	all	0.02527	0.00029	-0.00057	0.03002	0.00030	-0.00051	0.02270	0.00037	-0.00079
sedafips	cs	asn	0.06883	0.00042	0.00168	0.08004	0.00060	0.00356	0.06300	0.00046	0.00036
sedafips	cs	blk	0.01963	0.00039	-0.00014	0.02238	0.00043	0.00013	0.01907	0.00046	-0.00041
sedafips	cs	ecd	0.01164	0.00049	-0.00086	0.01466	0.00047	-0.00079	0.01125	0.00062	-0.00105
sedafips	cs	f	0.02292	0.00030	-0.00048	0.02588	0.00032	-0.00048	0.02242	0.00037	-0.00053
sedafips	cs	hsp	0.01481	0.00055	-0.00069	0.01671	0.00049	-0.00037	0.01548	0.00073	-0.00109
sedafips	cs	m	0.02896	0.00030	-0.00069	0.03506	0.00032	-0.00057	0.02526	0.00043	-0.00098
sedafips	cs	mfg	0.00228	0.00005	-0.00002	0.00198	0.00003	-0.00010	0.00383	0.00012	0.00015
sedafips	cs	mtr	0.02786	0.00039	-0.00059	0.03261	0.00049	-0.00042	0.02577	0.00037	-0.00099
sedafips	cs	nam	0.06070	0.00052	-0.00266	0.05583	0.00055	-0.00167	0.06880	0.00064	-0.00383
sedafips	cs	nec	0.01052	0.00031	-0.00027	0.01429	0.00036	-0.00026	0.00953	0.00036	-0.00042
sedafips	cs	neg	0.00847	0.00022	0.00003	0.00891	0.00019	0.00002	0.00883	0.00030	0.00003
sedafips	cs	wag	0.06914	0.00037	0.00195	0.08012	0.00042	0.00257	0.06228	0.00039	0.00160
sedafips	cs	wbg	0.03326	0.00017	0.00082	0.03335	0.00023	0.00079	0.03482	0.00019	0.00090
sedafips	cs	whg	0.03623	0.00024	0.00104	0.03637	0.00022	0.00105	0.03788	0.00032	0.00087
sedafips	cs	wht	0.02642	0.00026	0.00025	0.02960	0.00032	0.00049	0.02580	0.00031	-0.00014
sedafips	cs	wmg	0.00815	0.00009	0.00026	0.00892	0.00009	0.00039	0.00782	0.00010	0.00014
sedafips	cs	wng	0.06848	0.00020	0.00045	0.06889	0.00025	0.00140	0.06997	0.00021	-0.00031
sedafips	gcs	all	0.26852	0.00315	0.00231	0.30731	0.00367	0.01487	0.25639	0.00440	-0.01124
sedafips	gcs	asn	0.72902	0.00598	0.03631	0.81518	0.01361	0.08687	0.71300	0.00520	-0.00281
sedafips	gcs	blk	0.20905	0.00406	0.00456	0.22718	0.00538	0.01657	0.21583	0.00539	-0.00672
sedafips	gcs	ecd	0.12349	0.00477	-0.00500	0.14900	0.00406	0.00218	0.12798	0.00725	-0.01322
sedafips	gcs	f	0.24432	0.00329	0.00217	0.26520	0.00368	0.01249	0.25316	0.00433	-0.00831
sedafips	gcs	hsp	0.15741	0.00564	-0.00286	0.16941	0.00504	0.00773	0.17519	0.00848	-0.01411
sedafips	gcs	m	0.30778	0.00330	0.00233	0.35899	0.00397	0.01771	0.28539	0.00513	-0.01378
sedafips	gcs	mfg	0.02471	0.00052	0.00033	0.01995	0.00027	0.00039	0.04311	0.00136	0.00127
sedafips	gcs	mtr	0.29287	0.00419	0.00159	0.33222	0.00596	0.01689	0.29063	0.00442	-0.01384
sedafips	gcs	nam	0.64175	0.00451	-0.01314	0.55886	0.00552	0.01850	0.77984	0.00819	-0.05106
sedafips	gcs	nec	0.11215	0.00336	0.00056	0.14738	0.00379	0.00711	0.10802	0.00422	-0.00586
sedafips	gcs	neg	0.08839	0.00212	0.00297	0.08880	0.00216	0.00598	0.10002	0.00343	-0.00061
sedafips	gcs	wag	0.73355	0.00580	0.04146	0.81570	0.01074	0.07701	0.70301	0.00411	0.01149
sedafips	gcs	wbg	0.35799	0.00248	0.01923	0.33956	0.00468	0.02975	0.39236	0.00193	0.00617
sedafips	gcs	whg	0.39265	0.00363	0.02379	0.37232	0.00500	0.03431	0.42651	0.00338	0.00551
sedafips	gcs	wht	0.28181	0.00337	0.01118	0.30278	0.00516	0.02455	0.29165	0.00353	-0.00437
sedafips	gcs	wmg	0.08712	0.00134	0.00549	0.09173	0.00189	0.00988	0.08813	0.00112	0.00079
sedafips	gcs	wng	0.72737	0.00279	0.02342	0.69803	0.00691	0.05734	0.79172	0.00256	-0.01092

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Table 11b. Commuting Zone Variances and Covariances

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedacz	cs	all	0.03778	0.00059	-0.00043	0.04185	0.00082	-0.00002	0.03799	0.00057	-0.00077
sedacz	cs	asn	0.08667	0.00134	-0.00004	0.09995	0.00161	0.00159	0.08113	0.00131	-0.00100
sedacz	cs	blk	0.03272	0.00075	0.00064	0.03544	0.00098	0.00096	0.03357	0.00068	0.00042
sedacz	cs	ecd	0.02700	0.00067	-0.00088	0.03049	0.00083	-0.00054	0.02798	0.00068	-0.00119
sedacz	cs	f	0.03466	0.00059	-0.00027	0.03670	0.00080	0.00005	0.03777	0.00056	-0.00042
sedacz	cs	hsp	0.02196	0.00077	-0.00131	0.02410	0.00092	-0.00085	0.02473	0.00081	-0.00166
sedacz	cs	m	0.04196	0.00059	-0.00057	0.04797	0.00080	-0.00013	0.03982	0.00060	-0.00095
sedacz	cs	mfg	0.00259	0.00006	0.00001	0.00238	0.00004	-0.00007	0.00410	0.00011	0.00017
sedacz	cs	mtr	0.03094	0.00073	-0.00007	0.03794	0.00108	-0.00048	0.02735	0.00054	0.00007
sedacz	cs	nam	0.05591	0.00111	-0.00213	0.05417	0.00136	-0.00154	0.06182	0.00104	-0.00279
sedacz	cs	nec	0.02124	0.00069	-0.00022	0.02620	0.00095	-0.00010	0.02052	0.00063	-0.00019
sedacz	cs	neg	0.01900	0.00027	0.00025	0.02043	0.00032	0.00058	0.01850	0.00029	-0.00006
sedacz	cs	wag	0.07946	0.00084	0.00030	0.09063	0.00090	0.00133	0.07263	0.00090	-0.00040
sedacz	cs	wbg	0.04127	0.00038	0.00115	0.04336	0.00055	0.00200	0.04094	0.00032	0.00038
sedacz	cs	whg	0.03345	0.00027	0.00040	0.03431	0.00036	0.00118	0.03513	0.00029	-0.00045
sedacz	cs	wht	0.02390	0.00052	-0.00005	0.03129	0.00080	0.00021	0.02037	0.00046	-0.00020
sedacz	cs	wmg	0.01327	0.00020	0.00032	0.01439	0.00024	0.00052	0.01263	0.00018	0.00013
sedacz	cs	wng	0.06200	0.00070	0.00066	0.06227	0.00095	0.00179	0.06371	0.00054	-0.00025
sedacz	gcs	all	0.39996	0.00654	0.00643	0.42600	0.00982	0.02781	0.42999	0.00668	-0.01292
sedacz	gcs	asn	0.91045	0.01497	0.02408	1.01719	0.02173	0.08160	0.91723	0.01517	-0.02002
sedacz	gcs	blk	0.34672	0.00847	0.01602	0.36113	0.01213	0.03377	0.37947	0.00769	0.00095
sedacz	gcs	ecd	0.28492	0.00692	-0.00158	0.30952	0.00870	0.01541	0.31703	0.00803	-0.01669
sedacz	gcs	f	0.36753	0.00656	0.00704	0.37429	0.00954	0.02529	0.42735	0.00649	-0.00899
sedacz	gcs	hsp	0.23078	0.00781	-0.00757	0.24218	0.00897	0.00823	0.28019	0.00961	-0.02171
sedacz	gcs	m	0.44298	0.00655	0.00644	0.48742	0.00981	0.03061	0.45071	0.00704	-0.01523
sedacz	gcs	mfg	0.02809	0.00062	0.00074	0.02396	0.00044	0.00098	0.04620	0.00118	0.00153
sedacz	gcs	mtr	0.32466	0.00794	0.00826	0.38713	0.01159	0.02049	0.30924	0.00606	-0.00237
sedacz	gcs	nam	0.58624	0.01077	-0.00839	0.54153	0.01338	0.02013	0.70111	0.01266	-0.03867
sedacz	gcs	nec	0.22480	0.00760	0.00443	0.26577	0.01038	0.01713	0.23259	0.00728	-0.00447
sedacz	gcs	neg	0.20115	0.00315	0.00807	0.20732	0.00468	0.01925	0.20903	0.00334	-0.00263
sedacz	gcs	wag	0.83738	0.00984	0.02763	0.92662	0.01426	0.07220	0.81993	0.01026	-0.01220
sedacz	gcs	wbg	0.43917	0.00504	0.02337	0.44048	0.00942	0.04804	0.46292	0.00363	-0.00021
sedacz	gcs	whg	0.35548	0.00337	0.01362	0.35059	0.00646	0.03471	0.39696	0.00343	-0.00897
sedacz	gcs	wht	0.25192	0.00591	0.00708	0.31932	0.00957	0.02342	0.23060	0.00528	-0.00452
sedacz	gcs	wmg	0.14074	0.00253	0.00745	0.14821	0.00364	0.01427	0.14270	0.00200	0.00013
sedacz	gcs	wng	0.65379	0.00782	0.02447	0.63125	0.01365	0.05785	0.72136	0.00631	-0.00999

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Table 11c. Metropolitan Area Variances and Covariances

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedametro	cs	all	0.04263	0.00071	0.00004	0.04884	0.00099	0.00039	0.04109	0.00068	-0.00032
sedametro	cs	asn	0.09326	0.00149	0.00081	0.10675	0.00177	0.00231	0.08771	0.00141	-0.00025
sedametro	cs	blk	0.03815	0.00099	0.00074	0.04213	0.00133	0.00112	0.03849	0.00086	0.00037
sedametro	cs	ecd	0.02568	0.00083	-0.00091	0.03114	0.00104	-0.00068	0.02495	0.00085	-0.00118
sedametro	cs	f	0.04009	0.00073	0.00019	0.04387	0.00100	0.00046	0.04173	0.00069	0.00000
sedametro	cs	hsp	0.03091	0.00102	-0.00139	0.03362	0.00122	-0.00091	0.03403	0.00106	-0.00186
sedametro	cs	m	0.04640	0.00074	-0.00013	0.05474	0.00100	0.00034	0.04256	0.00073	-0.00062
sedametro	cs	mfg	0.00236	0.00007	0.00001	0.00230	0.00005	-0.00005	0.00389	0.00011	0.00018
sedametro	cs	mtr	0.04210	0.00115	0.00008	0.04927	0.00161	-0.00021	0.03861	0.00085	0.00013
sedametro	cs	nam	0.05453	0.00138	-0.00139	0.05504	0.00178	-0.00121	0.05762	0.00123	-0.00175
sedametro	cs	nec	0.02937	0.00077	0.00046	0.03696	0.00108	0.00058	0.02648	0.00070	0.00040
sedametro	cs	neg	0.01928	0.00031	0.00023	0.02080	0.00036	0.00061	0.01878	0.00033	-0.00015
sedametro	cs	wag	0.07079	0.00093	0.00129	0.07953	0.00104	0.00215	0.06624	0.00093	0.00061
sedametro	cs	wbg	0.04533	0.00045	0.00130	0.04762	0.00062	0.00207	0.04483	0.00039	0.00061
sedametro	cs	whg	0.04156	0.00038	0.00059	0.04149	0.00046	0.00152	0.04433	0.00041	-0.00052
sedametro	cs	wht	0.03412	0.00070	0.00021	0.04250	0.00103	0.00044	0.03011	0.00063	-0.00001
sedametro	cs	wmg	0.01563	0.00030	0.00023	0.01719	0.00033	0.00042	0.01447	0.00028	0.00006
sedametro	cs	wng	0.05154	0.00097	0.00150	0.05182	0.00125	0.00239	0.05247	0.00075	0.00068
sedametro	gcs	all	0.45226	0.00815	0.01326	0.49775	0.01227	0.03675	0.46476	0.00789	-0.00824
sedametro	gcs	asn	0.98029	0.01722	0.03638	1.08640	0.02451	0.09351	0.99131	0.01618	-0.01239
sedametro	gcs	blk	0.40292	0.01102	0.01867	0.42749	0.01593	0.03959	0.43528	0.00975	-0.00021
sedametro	gcs	ecd	0.26994	0.00861	-0.00191	0.31439	0.01061	0.01450	0.28290	0.01001	-0.01622
sedametro	gcs	f	0.42577	0.00829	0.01379	0.44781	0.01224	0.03427	0.47185	0.00794	-0.00468
sedametro	gcs	hsp	0.32484	0.01042	-0.00599	0.33778	0.01214	0.01391	0.38559	0.01249	-0.02502
sedametro	gcs	m	0.49045	0.00838	0.01279	0.55723	0.01254	0.03999	0.48148	0.00846	-0.01171
sedametro	gcs	mfg	0.02552	0.00069	0.00071	0.02320	0.00053	0.00110	0.04386	0.00123	0.00158
sedametro	gcs	mtr	0.44242	0.01243	0.01391	0.50331	0.01757	0.03152	0.43667	0.00947	-0.00289
sedametro	gcs	nam	0.57055	0.01400	-0.00033	0.55177	0.01766	0.02491	0.65289	0.01451	-0.02667
sedametro	gcs	nec	0.31140	0.00890	0.01440	0.37653	0.01295	0.03094	0.29968	0.00793	0.00158
sedametro	gcs	neg	0.20404	0.00357	0.00802	0.21093	0.00506	0.01983	0.21235	0.00380	-0.00374
sedametro	gcs	wag	0.74588	0.01143	0.03516	0.81248	0.01617	0.07290	0.74746	0.01045	-0.00035
sedametro	gcs	wbg	0.48145	0.00568	0.02594	0.48436	0.01038	0.05162	0.50686	0.00434	0.00187
sedametro	gcs	whg	0.44153	0.00471	0.01747	0.42432	0.00816	0.04287	0.50099	0.00483	-0.01081
sedametro	gcs	wht	0.36074	0.00804	0.01305	0.43290	0.01246	0.03317	0.34067	0.00719	-0.00346
sedametro	gcs	wmg	0.16572	0.00346	0.00744	0.17750	0.00454	0.01541	0.16355	0.00304	-0.00084
sedametro	gcs	wng	0.54316	0.01098	0.02976	0.52671	0.01671	0.05710	0.59343	0.00843	0.00189

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Table 11d. County Variances and Covariances

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedacounty	cs	all	0.05924	0.00101	0.00024	0.06810	0.00150	0.00101	0.05577	0.00089	-0.00028
sedacounty	cs	asn	0.10373	0.00182	0.00130	0.11802	0.00217	0.00251	0.09860	0.00173	0.00055
sedacounty	cs	blk	0.04519	0.00130	0.00060	0.04941	0.00174	0.00114	0.04597	0.00110	0.00023
sedacounty	cs	ecd	0.03655	0.00113	-0.00060	0.04383	0.00154	-0.00012	0.03458	0.00105	-0.00095
sedacounty	cs	f	0.05447	0.00101	0.00046	0.06041	0.00147	0.00113	0.05474	0.00088	0.00011
sedacounty	cs	hsp	0.03689	0.00146	-0.00120	0.03939	0.00187	-0.00048	0.04097	0.00135	-0.00184
sedacounty	cs	m	0.06539	0.00105	0.00010	0.07705	0.00147	0.00096	0.05898	0.00094	-0.00055
sedacounty	cs	mfg	0.00367	0.00009	0.00003	0.00365	0.00006	-0.00007	0.00526	0.00013	0.00022
sedacounty	cs	mtr	0.05365	0.00150	-0.00001	0.06307	0.00203	-0.00014	0.04843	0.00112	-0.00001
sedacounty	cs	nam	0.07417	0.00171	-0.00145	0.07544	0.00216	-0.00064	0.07820	0.00149	-0.00234
sedacounty	cs	nec	0.04290	0.00106	0.00068	0.05319	0.00157	0.00126	0.03829	0.00091	0.00041
sedacounty	cs	neg	0.02410	0.00035	0.00033	0.02552	0.00040	0.00073	0.02367	0.00036	-0.00005
sedacounty	cs	wag	0.06904	0.00120	0.00079	0.07592	0.00133	0.00119	0.06680	0.00118	0.00059
sedacounty	cs	wbg	0.05191	0.00061	0.00179	0.05410	0.00078	0.00266	0.05185	0.00053	0.00106
sedacounty	cs	whg	0.04801	0.00052	0.00097	0.04839	0.00060	0.00178	0.05067	0.00053	0.00001
sedacounty	cs	wht	0.04521	0.00098	0.00025	0.05611	0.00148	0.00083	0.03943	0.00083	-0.00008
sedacounty	cs	wmg	0.02299	0.00047	0.00050	0.02461	0.00052	0.00075	0.02176	0.00042	0.00027
sedacounty	cs	wng	0.06397	0.00138	0.00159	0.06557	0.00169	0.00279	0.06447	0.00115	0.00052
sedacounty	gcs	all	0.62613	0.01165	0.02021	0.69455	0.01874	0.05586	0.63091	0.01033	-0.00943
sedacounty	gcs	asn	1.09283	0.02104	0.04430	1.20362	0.02901	0.10333	1.11407	0.01973	-0.00433
sedacounty	gcs	blk	0.47744	0.01432	0.01969	0.50345	0.02015	0.04519	0.51996	0.01252	-0.00275
sedacounty	gcs	ecd	0.38460	0.01213	0.00478	0.44548	0.01668	0.02908	0.39173	0.01221	-0.01471
sedacounty	gcs	f	0.57608	0.01163	0.02100	0.61756	0.01830	0.05220	0.61922	0.01006	-0.00490
sedacounty	gcs	hsp	0.38791	0.01522	-0.00207	0.39801	0.01915	0.02269	0.46427	0.01586	-0.02570
sedacounty	gcs	m	0.68895	0.01196	0.02097	0.78506	0.01878	0.06111	0.66727	0.01096	-0.01280
sedacounty	gcs	mfg	0.03950	0.00092	0.00135	0.03684	0.00071	0.00186	0.05924	0.00144	0.00193
sedacounty	gcs	mtr	0.56265	0.01611	0.01615	0.64562	0.02207	0.04125	0.54765	0.01254	-0.00549
sedacounty	gcs	nam	0.77762	0.01753	0.00496	0.76033	0.02285	0.04400	0.88659	0.01771	-0.03537
sedacounty	gcs	nec	0.45223	0.01240	0.02065	0.54260	0.01924	0.04882	0.43341	0.01027	0.00038
sedacounty	gcs	neg	0.25440	0.00410	0.00986	0.25937	0.00578	0.02398	0.26748	0.00411	-0.00313
sedacounty	gcs	wag	0.72939	0.01390	0.02937	0.77428	0.01764	0.06134	0.75394	0.01326	-0.00045
sedacounty	gcs	wbg	0.54712	0.00780	0.03236	0.55120	0.01280	0.06138	0.58611	0.00585	0.00611
sedacounty	gcs	whg	0.50842	0.00632	0.02315	0.49536	0.01000	0.04958	0.57262	0.00602	-0.00554
sedacounty	gcs	wht	0.47560	0.01121	0.01671	0.57253	0.01797	0.04645	0.44619	0.00949	-0.00537
sedacounty	gcs	wmg	0.24351	0.00541	0.01226	0.25490	0.00709	0.02352	0.24615	0.00457	0.00073
sedacounty	gcs	wng	0.67252	0.01539	0.03481	0.66632	0.02200	0.07048	0.72964	0.01315	-0.00122

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Table 11e.District and School Variances and Covariances

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedalea	cs	all	0.12158	0.00186	0.00174	0.13193	0.00274	0.00301	0.11830	0.00159	0.00075
sedalea	cs	asn	0.16630	0.00224	0.00288	0.18601	0.00286	0.00507	0.15725	0.00202	0.00122
sedalea	cs	blk	0.07795	0.00213	0.00154	0.08048	0.00275	0.00213	0.08143	0.00178	0.00100
sedalea	cs	ecd	0.05923	0.00195	-0.00028	0.06761	0.00265	0.00030	0.05743	0.00172	-0.00067
sedalea	cs	f	0.11174	0.00175	0.00163	0.11572	0.00258	0.00267	0.11563	0.00147	0.00086
sedalea	cs	hsp	0.07718	0.00226	-0.00039	0.07659	0.00294	0.00070	0.08550	0.00203	-0.00148
sedalea	cs	m	0.12335	0.00193	0.00159	0.13889	0.00270	0.00308	0.11443	0.00167	0.00042
sedalea	cs	mfg	0.00501	0.00014	0.00009	0.00499	0.00012	0.00001	0.00669	0.00018	0.00028
sedalea	cs	mtr	0.10685	0.00271	0.00190	0.11758	0.00352	0.00251	0.10117	0.00207	0.00135
sedalea	cs	nam	0.09140	0.00311	-0.00103	0.09434	0.00398	-0.00019	0.09412	0.00245	-0.00182
sedalea	cs	nec	0.08372	0.00183	0.00159	0.09791	0.00268	0.00227	0.07698	0.00154	0.00128
sedalea	cs	neg	0.02908	0.00045	0.00008	0.03097	0.00047	0.00043	0.02824	0.00044	-0.00026
sedalea	cs	wag	0.06942	0.00113	0.00049	0.07878	0.00128	0.00074	0.06500	0.00106	0.00041
sedalea	cs	wbg	0.05399	0.00079	0.00164	0.05626	0.00093	0.00234	0.05399	0.00070	0.00101
sedalea	cs	whg	0.04552	0.00068	0.00065	0.04564	0.00072	0.00128	0.04813	0.00070	-0.00011
sedalea	cs	wht	0.08973	0.00177	0.00166	0.10157	0.00265	0.00236	0.08474	0.00145	0.00120
sedalea	cs	wmg	0.03096	0.00095	0.00065	0.03260	0.00103	0.00090	0.02966	0.00086	0.00041
sedalea	cs	wng	0.06090	0.00190	0.00119	0.06099	0.00213	0.00205	0.06246	0.00171	0.00034
sedalea	gcs	all	1.28600	0.02153	0.05259	1.34511	0.03559	0.11740	1.33845	0.01808	-0.00446
sedalea	gcs	asn	1.75477	0.02613	0.07736	1.89945	0.04162	0.17334	1.77635	0.02295	-0.00334
sedalea	gcs	blk	0.82177	0.02349	0.03795	0.81998	0.03227	0.07595	0.92047	0.02027	0.00209
sedalea	gcs	ecd	0.62208	0.02091	0.01439	0.68719	0.02879	0.04965	0.65037	0.01984	-0.01413
sedalea	gcs	f	1.18378	0.02024	0.04866	1.18054	0.03293	0.10383	1.30800	0.01672	-0.00298
sedalea	gcs	hsp	0.81370	0.02417	0.01722	0.77900	0.03238	0.05971	0.96798	0.02361	-0.02639
sedalea	gcs	m	1.30060	0.02216	0.05223	1.41543	0.03547	0.12251	1.29453	0.01914	-0.00785
sedalea	gcs	mfg	0.05350	0.00152	0.00234	0.05036	0.00133	0.00335	0.07547	0.00196	0.00255
sedalea	gcs	mtr	1.12804	0.03041	0.05217	1.20695	0.04235	0.10465	1.14325	0.02320	0.00385
sedalea	gcs	nam	0.95911	0.03238	0.01583	0.95360	0.04170	0.06457	1.06672	0.02867	-0.03112
sedalea	gcs	nec	0.88707	0.02099	0.04265	0.99823	0.03293	0.08884	0.87084	0.01733	0.00607
sedalea	gcs	neg	0.30630	0.00499	0.00889	0.31524	0.00628	0.02454	0.31927	0.00507	-0.00600
sedalea	gcs	wag	0.73167	0.01291	0.02572	0.80211	0.01653	0.05844	0.73411	0.01205	-0.00242
sedalea	gcs	wbg	0.56765	0.00952	0.03209	0.57285	0.01392	0.05943	0.61033	0.00783	0.00538
sedalea	gcs	whg	0.48085	0.00781	0.01964	0.46617	0.01037	0.04272	0.54426	0.00801	-0.00670
sedalea	gcs	wht	0.95076	0.02033	0.04449	1.03643	0.03273	0.09176	0.95866	0.01644	0.00427
sedalea	gcs	wmg	0.32974	0.01065	0.01609	0.33887	0.01275	0.03019	0.33500	0.00946	0.00169
sedalea	gcs	wng	0.64088	0.02049	0.03027	0.61961	0.02506	0.06095	0.70649	0.01955	-0.00300

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Identifiers			Pooled			Math			RLA		
Geo	Scale	Subgroup	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)	tau(int)	tau(grd)	cov(int,grd)
sedasch	cs	all	0.20778	0.00385	0.00248	0.22153	0.00564	0.00536	0.20246	0.00293	0.00013
sedasch	gcs	all	3.34049	0.04209	0.09167	3.34381	0.06529	0.19831	3.45927	0.03311	-0.01355

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau = variance; rel = reliability

Table 12a. State Reliabilities

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedafips	cs	all	0.999	0.969	0.999	0.949	0.999	0.964
sedafips	cs	asn	0.999	0.918	0.998	0.898	0.998	0.879
sedafips	cs	blk	0.997	0.936	0.995	0.899	0.995	0.911
sedafips	cs	ecd	0.996	0.962	0.994	0.931	0.994	0.952
sedafips	cs	f	0.999	0.962	0.998	0.939	0.998	0.954
sedafips	cs	hsp	0.997	0.957	0.994	0.921	0.995	0.949
sedafips	cs	m	0.999	0.963	0.998	0.936	0.998	0.958
sedafips	cs	mfg	0.989	0.813	0.973	0.578	0.989	0.854
sedafips	cs	mtr	0.984	0.912	0.979	0.879	0.977	0.864
sedafips	cs	nam	0.997	0.900	0.994	0.839	0.995	0.863
sedafips	cs	nec	0.997	0.953	0.995	0.925	0.993	0.932
sedafips	cs	neg	0.997	0.950	0.994	0.895	0.995	0.939
sedafips	cs	wag	0.999	0.913	0.998	0.860	0.998	0.858
sedafips	cs	wbg	0.999	0.886	0.997	0.843	0.997	0.824
sedafips	cs	whg	0.999	0.930	0.998	0.872	0.998	0.911
sedafips	cs	wht	0.999	0.965	0.999	0.949	0.999	0.952
sedafips	cs	wmg	0.972	0.790	0.960	0.649	0.961	0.704
sedafips	cs	wng	0.998	0.817	0.995	0.742	0.996	0.727
sedafips	gcs	all	0.999	0.969	0.999	0.958	0.999	0.967
sedafips	gcs	asn	0.998	0.918	0.998	0.950	0.998	0.883
sedafips	gcs	blk	0.997	0.927	0.995	0.918	0.995	0.917
sedafips	gcs	ecd	0.996	0.955	0.994	0.921	0.994	0.956
sedafips	gcs	f	0.999	0.962	0.998	0.947	0.998	0.957
sedafips	gcs	hsp	0.996	0.951	0.994	0.922	0.995	0.952
sedafips	gcs	m	0.999	0.962	0.999	0.947	0.999	0.962
sedafips	gcs	mfg	0.990	0.805	0.973	0.550	0.989	0.854
sedafips	gcs	mtr	0.984	0.910	0.979	0.895	0.977	0.871
sedafips	gcs	nam	0.997	0.876	0.994	0.839	0.995	0.877
sedafips	gcs	nec	0.996	0.941	0.994	0.924	0.994	0.935
sedafips	gcs	neg	0.994	0.903	0.994	0.909	0.995	0.941
sedafips	gcs	wag	0.999	0.933	0.998	0.937	0.998	0.855
sedafips	gcs	wbg	0.998	0.877	0.997	0.913	0.997	0.816
sedafips	gcs	whg	0.999	0.931	0.998	0.938	0.998	0.908
sedafips	gcs	wht	0.999	0.964	0.999	0.967	0.999	0.954
sedafips	gcs	wmg	0.972	0.830	0.961	0.781	0.960	0.697
sedafips	gcs	wng	0.997	0.821	0.995	0.870	0.996	0.739

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Table 12b. Commuting Zone Reliabilities

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedacz	cs	all	0.998	0.944	0.995	0.920	0.996	0.906
sedacz	cs	asn	0.975	0.734	0.959	0.648	0.955	0.629
sedacz	cs	blk	0.968	0.769	0.945	0.706	0.946	0.672
sedacz	cs	ecd	0.994	0.927	0.988	0.890	0.989	0.885
sedacz	cs	f	0.997	0.928	0.992	0.899	0.994	0.880
sedacz	cs	hsp	0.974	0.829	0.955	0.766	0.958	0.757
sedacz	cs	m	0.997	0.926	0.994	0.894	0.994	0.883
sedacz	cs	mfg	0.953	0.647	0.909	0.430	0.944	0.639
sedacz	cs	mtr	0.935	0.695	0.902	0.635	0.892	0.543
sedacz	cs	nam	0.959	0.674	0.919	0.584	0.938	0.552
sedacz	cs	nec	0.991	0.924	0.985	0.892	0.984	0.869
sedacz	cs	neg	0.988	0.840	0.978	0.773	0.978	0.776
sedacz	cs	wag	0.978	0.665	0.959	0.550	0.955	0.565
sedacz	cs	wbg	0.975	0.687	0.955	0.628	0.956	0.567
sedacz	cs	whg	0.981	0.699	0.965	0.629	0.967	0.606
sedacz	cs	wht	0.994	0.925	0.991	0.904	0.988	0.869
sedacz	cs	wmg	0.889	0.496	0.827	0.392	0.832	0.372
sedacz	cs	wng	0.968	0.600	0.933	0.521	0.945	0.433
sedacz	gcs	all	0.998	0.946	0.995	0.931	0.996	0.912
sedacz	gcs	asn	0.974	0.736	0.959	0.696	0.955	0.633
sedacz	gcs	blk	0.968	0.774	0.944	0.735	0.946	0.675
sedacz	gcs	ecd	0.994	0.924	0.988	0.893	0.990	0.893
sedacz	gcs	f	0.997	0.930	0.993	0.912	0.994	0.885
sedacz	gcs	hsp	0.973	0.823	0.954	0.762	0.959	0.766
sedacz	gcs	m	0.997	0.929	0.994	0.909	0.994	0.890
sedacz	gcs	mfg	0.954	0.644	0.908	0.442	0.944	0.637
sedacz	gcs	mtr	0.935	0.698	0.903	0.649	0.893	0.544
sedacz	gcs	nam	0.959	0.657	0.918	0.582	0.939	0.566
sedacz	gcs	nec	0.991	0.922	0.985	0.898	0.984	0.872
sedacz	gcs	neg	0.987	0.831	0.978	0.820	0.978	0.779
sedacz	gcs	wag	0.977	0.680	0.959	0.630	0.955	0.567
sedacz	gcs	wbg	0.975	0.704	0.955	0.706	0.956	0.566
sedacz	gcs	whg	0.981	0.709	0.965	0.720	0.967	0.613
sedacz	gcs	wht	0.994	0.927	0.991	0.916	0.988	0.874
sedacz	gcs	wmg	0.889	0.520	0.830	0.466	0.832	0.365
sedacz	gcs	wng	0.968	0.604	0.933	0.587	0.945	0.440

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Table 12c. Metropolitan Area Reliabilities

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedametro	cs	all	0.998	0.950	0.996	0.929	0.996	0.911
sedametro	cs	asn	0.976	0.696	0.958	0.602	0.951	0.568
sedametro	cs	blk	0.971	0.777	0.951	0.720	0.948	0.664
sedametro	cs	ecd	0.995	0.939	0.992	0.909	0.991	0.900
sedametro	cs	f	0.997	0.941	0.995	0.918	0.995	0.895
sedametro	cs	hsp	0.983	0.853	0.971	0.795	0.971	0.776
sedametro	cs	m	0.998	0.941	0.996	0.914	0.995	0.897
sedametro	cs	mfg	0.961	0.648	0.924	0.427	0.954	0.617
sedametro	cs	mtr	0.939	0.720	0.906	0.655	0.900	0.560
sedametro	cs	nam	0.934	0.622	0.888	0.541	0.894	0.482
sedametro	cs	nec	0.995	0.932	0.992	0.905	0.990	0.874
sedametro	cs	neg	0.993	0.857	0.987	0.782	0.985	0.783
sedametro	cs	wag	0.971	0.619	0.948	0.509	0.941	0.495
sedametro	cs	wbg	0.976	0.669	0.958	0.602	0.956	0.536
sedametro	cs	whg	0.988	0.724	0.976	0.638	0.977	0.623
sedametro	cs	wht	0.997	0.938	0.994	0.917	0.993	0.885
sedametro	cs	wmg	0.884	0.496	0.824	0.380	0.820	0.368
sedametro	cs	wng	0.949	0.579	0.903	0.490	0.908	0.405
sedametro	gcs	all	0.998	0.952	0.996	0.941	0.996	0.917
sedametro	gcs	asn	0.975	0.706	0.957	0.660	0.951	0.571
sedametro	gcs	blk	0.970	0.781	0.951	0.746	0.949	0.666
sedametro	gcs	ecd	0.995	0.937	0.992	0.909	0.991	0.907
sedametro	gcs	f	0.997	0.944	0.995	0.930	0.996	0.900
sedametro	gcs	hsp	0.983	0.848	0.970	0.793	0.971	0.784
sedametro	gcs	m	0.998	0.943	0.996	0.929	0.996	0.903
sedametro	gcs	mfg	0.962	0.640	0.923	0.440	0.954	0.614
sedametro	gcs	mtr	0.939	0.724	0.908	0.673	0.900	0.559
sedametro	gcs	nam	0.933	0.613	0.886	0.541	0.895	0.491
sedametro	gcs	nec	0.995	0.932	0.992	0.917	0.990	0.877
sedametro	gcs	neg	0.992	0.849	0.987	0.832	0.986	0.786
sedametro	gcs	wag	0.970	0.644	0.948	0.590	0.941	0.494
sedametro	gcs	wbg	0.976	0.684	0.957	0.685	0.956	0.534
sedametro	gcs	whg	0.987	0.736	0.976	0.736	0.977	0.630
sedametro	gcs	wht	0.996	0.940	0.994	0.929	0.993	0.890
sedametro	gcs	wmg	0.884	0.512	0.827	0.441	0.820	0.361
sedametro	gcs	wng	0.948	0.587	0.903	0.546	0.908	0.404

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Table 12d. County Reliabilities

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedacounty	cs	all	0.997	0.924	0.992	0.889	0.993	0.862
sedacounty	cs	asn	0.944	0.621	0.904	0.516	0.908	0.498
sedacounty	cs	blk	0.945	0.720	0.907	0.643	0.914	0.599
sedacounty	cs	ecd	0.990	0.901	0.979	0.853	0.981	0.833
sedacounty	cs	f	0.995	0.902	0.988	0.862	0.990	0.825
sedacounty	cs	hsp	0.951	0.757	0.913	0.682	0.927	0.645
sedacounty	cs	m	0.996	0.902	0.990	0.856	0.991	0.831
sedacounty	cs	mfg	0.920	0.509	0.849	0.311	0.898	0.462
sedacounty	cs	mtr	0.892	0.610	0.834	0.526	0.833	0.446
sedacounty	cs	nam	0.920	0.586	0.866	0.494	0.880	0.439
sedacounty	cs	nec	0.989	0.884	0.979	0.841	0.978	0.798
sedacounty	cs	neg	0.976	0.728	0.952	0.617	0.958	0.621
sedacounty	cs	wag	0.942	0.561	0.893	0.440	0.900	0.437
sedacounty	cs	wbg	0.958	0.604	0.920	0.510	0.928	0.473
sedacounty	cs	whg	0.965	0.588	0.929	0.482	0.942	0.472
sedacounty	cs	wht	0.992	0.899	0.985	0.862	0.984	0.822
sedacounty	cs	wmg	0.835	0.412	0.743	0.303	0.756	0.289
sedacounty	cs	wng	0.935	0.564	0.880	0.459	0.891	0.397
sedacounty	gcs	all	0.997	0.928	0.992	0.906	0.994	0.868
sedacounty	gcs	asn	0.943	0.631	0.905	0.569	0.908	0.498
sedacounty	gcs	blk	0.945	0.723	0.907	0.665	0.914	0.602
sedacounty	gcs	ecd	0.989	0.901	0.979	0.860	0.981	0.841
sedacounty	gcs	f	0.995	0.907	0.988	0.882	0.990	0.829
sedacounty	gcs	hsp	0.951	0.754	0.913	0.687	0.928	0.652
sedacounty	gcs	m	0.995	0.907	0.990	0.879	0.991	0.837
sedacounty	gcs	mfg	0.921	0.509	0.848	0.329	0.898	0.458
sedacounty	gcs	mtr	0.892	0.613	0.838	0.544	0.833	0.445
sedacounty	gcs	nam	0.919	0.577	0.865	0.506	0.880	0.449
sedacounty	gcs	nec	0.988	0.889	0.979	0.862	0.979	0.801
sedacounty	gcs	neg	0.975	0.730	0.953	0.681	0.958	0.623
sedacounty	gcs	wag	0.943	0.576	0.894	0.492	0.900	0.436
sedacounty	gcs	wbg	0.958	0.626	0.920	0.592	0.928	0.469
sedacounty	gcs	whg	0.965	0.605	0.929	0.573	0.942	0.474
sedacounty	gcs	wht	0.992	0.903	0.985	0.879	0.985	0.827
sedacounty	gcs	wmg	0.836	0.427	0.748	0.355	0.756	0.282
sedacounty	gcs	wng	0.935	0.571	0.880	0.512	0.891	0.399

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Table 12e. District and School Reliabilities

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedalea	cs	all	0.990	0.888	0.986	0.857	0.985	0.802
sedalea	cs	asn	0.946	0.618	0.919	0.531	0.915	0.483
sedalea	cs	blk	0.924	0.677	0.884	0.599	0.890	0.544
sedalea	cs	ecd	0.976	0.845	0.965	0.796	0.961	0.742
sedalea	cs	f	0.989	0.854	0.982	0.817	0.982	0.744
sedalea	cs	hsp	0.944	0.715	0.909	0.642	0.917	0.587
sedalea	cs	m	0.988	0.859	0.983	0.814	0.981	0.762
sedalea	cs	mfg	0.841	0.397	0.754	0.242	0.793	0.319
sedalea	cs	mtr	0.884	0.581	0.831	0.494	0.828	0.414
sedalea	cs	nam	0.877	0.582	0.820	0.493	0.826	0.415
sedalea	cs	nec	0.980	0.839	0.971	0.799	0.966	0.729
sedalea	cs	neg	0.941	0.565	0.908	0.441	0.901	0.435
sedalea	cs	wag	0.923	0.499	0.877	0.382	0.868	0.365
sedalea	cs	wbg	0.923	0.509	0.875	0.405	0.876	0.372
sedalea	cs	whg	0.929	0.492	0.880	0.370	0.884	0.374
sedalea	cs	wht	0.985	0.856	0.977	0.821	0.975	0.750
sedalea	cs	wmg	0.788	0.395	0.693	0.283	0.694	0.268
sedalea	cs	wng	0.881	0.505	0.807	0.383	0.815	0.349
sedalea	gcs	all	0.990	0.894	0.985	0.881	0.985	0.807
sedalea	gcs	asn	0.945	0.629	0.920	0.599	0.915	0.484
sedalea	gcs	blk	0.924	0.682	0.884	0.626	0.891	0.546
sedalea	gcs	ecd	0.976	0.846	0.965	0.807	0.961	0.749
sedalea	gcs	f	0.989	0.862	0.982	0.845	0.982	0.748
sedalea	gcs	hsp	0.944	0.716	0.909	0.659	0.917	0.593
sedalea	gcs	m	0.988	0.866	0.983	0.846	0.981	0.766
sedalea	gcs	mfg	0.842	0.400	0.752	0.264	0.793	0.315
sedalea	gcs	mtr	0.885	0.592	0.834	0.531	0.828	0.413
sedalea	gcs	nam	0.876	0.577	0.819	0.501	0.828	0.423
sedalea	gcs	nec	0.980	0.845	0.971	0.823	0.966	0.731
sedalea	gcs	neg	0.941	0.570	0.908	0.496	0.901	0.439
sedalea	gcs	wag	0.923	0.511	0.877	0.430	0.868	0.365
sedalea	gcs	wbg	0.922	0.527	0.875	0.477	0.876	0.369
sedalea	gcs	whg	0.928	0.503	0.879	0.438	0.884	0.376
sedalea	gcs	wht	0.985	0.862	0.977	0.844	0.975	0.753
sedalea	gcs	wmg	0.790	0.407	0.698	0.320	0.694	0.262
sedalea	gcs	wng	0.880	0.506	0.806	0.416	0.816	0.351

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Identifiers			Pooled		Math		RLA	
Geo	Scale	Subgroup	rel(int)	rel(grd)	rel(int)	rel(grd)	rel(int)	rel(grd)
sedasch	cs	all	0.969	0.706	0.955	0.652	0.956	0.536
sedasch	gcs	all	0.976	0.714	0.965	0.680	0.966	0.545

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School; cs = cohort scale; gcs = grade-cohort scale; asn = Asian; blk = Black; hsp = Hispanic; nam = Native-American; mtr = Multiracial; wht = White; m = male; f = female; wag = White-Asian gap; wbg = White-Black gap; whg = White-Hispanic gap; wng = White-Native American gap; wmg = White-Multiracial gap; mfg = male-female gap; tau

Table 13. Suppressed Estimates by Unit Post-Estimation, Long Form Data for Districts, Counties, Metropolitan Areas, Commuting Zones, and States

Cases Dropped Post-Estimation	sedasch	sedalea	sedacounty	sedametro	sedacz	sedafips
SE > 2	185 (0.00%)	1,261 (0.01%)	151 (0.00%)	6 (0.00%)	11 (0.00%)	0 (0.00%)
Alternative Assessment > 40%	70,622 (1.73%)	9,990 (0.10%)	1,626 (0.05%)	141 (0.01%)	56 (0.01%)	0 (0.00%)
Students < 20	411,351 (10.08%)	3,733,274 (35.96%)	718,621 (23.55%)	182,905 (16.73%)	107,935 (13.25%)	226 (0.37%)
Any Suppression/Drop	459,068 (11.24%)	3,733,822 (35.96%)	718,682 (23.55%)	182,905 (16.73%)	107,937 (13.25%)	226 (0.37%)
Total Cases in Public Long Files	3,622,270 (88.73%)	6,604,060 (63.61%)	2,314,315 (75.84%)	831,990 (76.12%)	565,943 (69.46%)	57,255 (92.84%)
Total Cases used in Estimation	4,082,479 (100.00%)	10,382,669 (100.00%	3,051,522 (100.00%)	1,093,031 (100.00%)	814,804 (100.00%)	61,668 (100.00%)

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School

Table 14. Suppressed Estimates by Unit Post-Estimation, Pooled Data for Schools, Districts, Counties, Metropolitan areas, commuting zones, and States

Cases Dropped Post-Pooling	sedasch	sedalea	sedacounty	sedametro	sedacz	sedafips
Suppressed Due to Reliability	30,811 (37.35%)	75,146 (40.86%)	20,729 (40.06%)	6,223 (37.28%)	3,950 (36.15%)	95 (10.30%)
Alternative Assessment > 40%	180 (0.22%)	209 (0.11%)	34 (0.07%)	3 (0.02%)	1 (0.01%)	0 (0.00%)
Unique Students < 20	1,507 (1.83%)	24,341 (13.24%)	4,194 (8.11%)	385 (2.31%)	286 (2.62%)	0 (0.00%)
Any Suppression/Drop	32,463 (39.35%)	99,520 (54.11%)	24,930 (48.18%)	6,609 (39.59%)	4,237 (38.78%)	95 (10.30%)
Unsuppressed Cases in Public Pooled Files	50,025 (60.65%)	84,389 (45.89%)	26,810 (51.82%)	10,085 (60.41%)	6,689 (61.22%)	827 (89.70%)
Total Observations in Public Pooled Files	82,488 (100.00%)	183,909 (100.00%	51,740 (100.00%)	16,694 (100.00%)	10,926 (100.00%)	922 (100.00%)

Note: sedafips = State; sedacz = Commuting zone; sedametro = Metro; sedacounty = County; sedalea = Geographic district; sedasch = School

Table 15. Component Loadings and Summary Statistics for Socioeconomic Status Composite Construction.

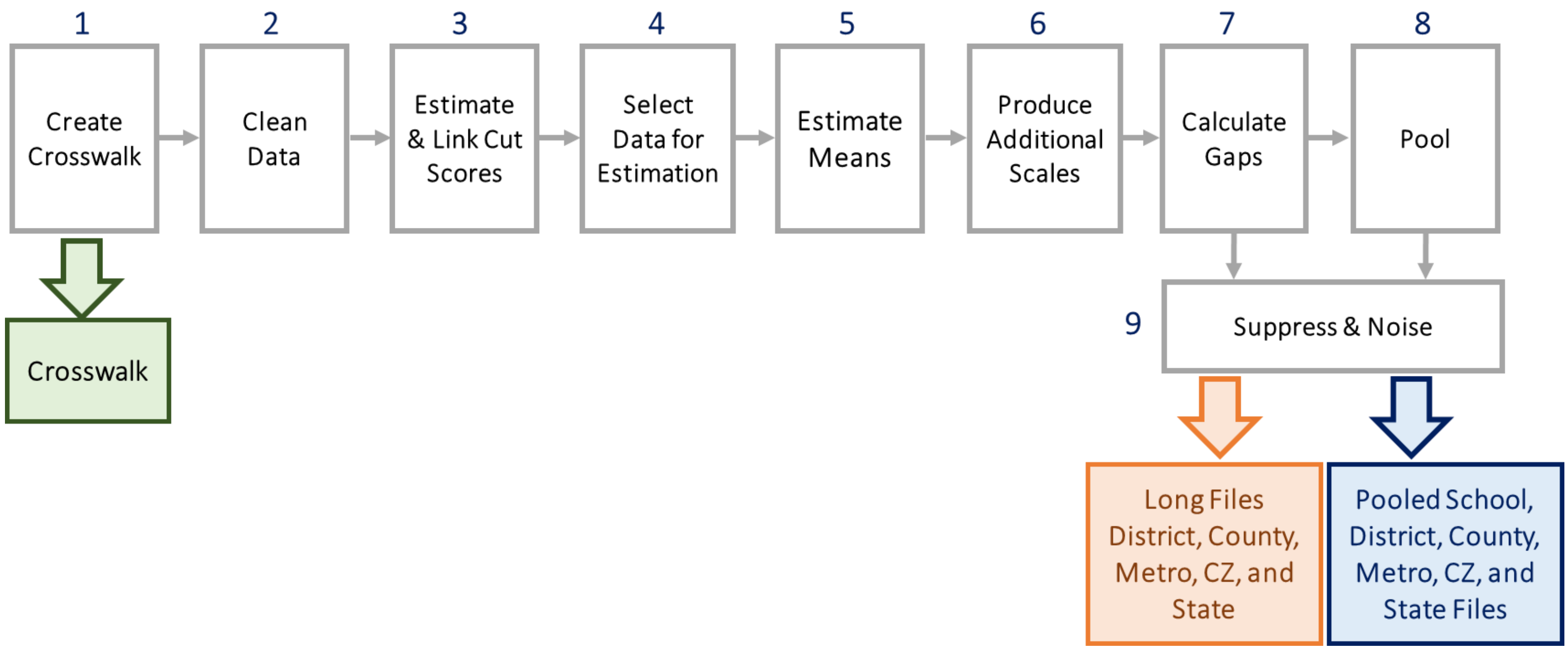
	Standardized Loadings	Unstandardized Loadings	Mean	SD
log(Median Family Income)	0.914	0.638	10.886	0.335
% with BA or Higher	0.674	1.170	0.285	0.135
Poverty Rate	-0.925	-2.664	0.151	0.081
SNAP Eligibility Rate	-0.931	-2.958	0.115	0.074
Unemployment Rate	-0.791	-5.221	0.097	0.035
Single Mother Headed Household Rate	-0.799	-2.315	0.199	0.081

Table 16. Summary Statistics at Different Values of the Socioeconomic Status Composite.

	SES Composite						
	below -2.5	-2.5 to -1.5	-1.5 to -.5	-.5 to .5	.5 to 1.5	1.5 to 2.5	above 2.5
log(Median Family Income)	10.17	10.31	10.51	10.74	11.03	11.55	12.11
% with BA or Higher	0.10	0.12	0.14	0.18	0.27	0.54	0.79
Poverty Rate	0.35	0.29	0.21	0.14	0.08	0.04	0.03
SNAP Eligibility Rate	0.32	0.26	0.18	0.11	0.06	0.02	0.01
Unemployment Rate	0.14	0.10	0.08	0.06	0.05	0.04	0.04
Single Mother Headed Household Rate	0.39	0.30	0.22	0.16	0.12	0.10	0.07

Figures

Figure 1. SEDA 4.1 Construction Process



Appendices

Appendix A: Additional Detail on Statistical Methods

1. Estimating Means and Standard Deviations for Units that Cross State Lines

This section briefly describes how means, standard deviations, and standard errors are estimated for units serving BIE schools or units that cross state lines. As described above, we first estimate unit "component" means and standard deviations. We then estimate the overall unit mean as weighted averages of the component means and the unit standard deviations as estimates of total variance within a unit based on the component means and standard deviations.

Let $\hat{\mu}_d$ and $\hat{\sigma}_d$ be the estimated means and standard deviations for the D components $d = 1, \ldots,$ that will be aggregated for a given unit. We also have estimates of the standard errors for each mean and standard deviation, $se(\hat{\mu}_d)$ and $se(\hat{\sigma}_d)$. We do not include grade, subject, year, or state subscripts here for clarity.

To estimate the aggregate parameters, we make the simplifying assumption that $cov(\hat{\mu}_i, \hat{\mu}_j) = cov(\hat{\sigma}_i, \hat{\sigma}_j) = cov(\hat{\mu}_i, \hat{\sigma}_i) = 0$ for $i \neq j$. The derivations for these expressions are based on the formulas in the appendix of Reardon et al. (2017) used to estimate to overall mean and variance of a set of groups in the HETOP model. Let

$$p_d = \frac{n_d}{\sum_{d=1}^{D} n_d} = \frac{n_d}{N_c}$$

be the proportion of all students in the aggregate unit c that are in component d. We estimate the aggregate mean for aggregate unit c as the weighted average of the component estimated means,

$$\hat{\mu}_c = \sum_{d=1}^{D} p_d \hat{\mu}_d,$$

with an estimated standard error of

$$se(\hat{\mu}_c) = \sqrt{\sum_{d=1}^{D} [p_d^2 \cdot se(\hat{\mu}_d)^2]}.$$

We estimate the standard deviation for aggregate unit c as the square root of the sum of the estimated between and within-unit variance,

$$\hat{\sigma}_c = \sqrt{\sum_{d=1}^{D} [p_d(\hat{\mu}_d - \hat{\mu}_c)^2 + q_d\hat{\sigma}_d^2]},$$

with the associated estimated standard error

$$se(\hat{\sigma}_c) = \sqrt{z_c * \left(\frac{1}{\hat{\sigma}_c}\right)}.$$

In these expressions we define

$$q_d = \left(\frac{p_d + (n_d - 1)}{n_d}\right)\left(\frac{p_d}{1 + 2\left(\frac{1}{2\tilde{n}_c}\right)}\right),$$

$$\tilde{n}_c = \left[\left(\frac{1}{D}\right)\sum_{d=1}^{D}\left(\frac{1}{n_d - 1}\right)\right]^{-1},$$

and

$$z_c = \sum_{d=1}^{D} [(p_d^2(\hat{\mu}_d - \hat{\mu}_c)^2 se(\hat{\mu}_d)^2) + (q_d^2 \cdot \hat{\sigma}_d^2 \cdot se(\hat{\sigma}_d)^2)].$$

2. Constructing OLS Standard Errors from Pooled Models

In the SEDA 4.1 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 9. 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 8.2; 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 8.2 and define:

$$\hat{\phi}^2_{urygb} = \hat{\sigma}^2 + \omega^2_{urygb} \tag{A-2.1}$$

Where ω^2_{urygb} is the variance of the $\hat{y}^x_{urygb}$ 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:

$$\hat{\phi}^{-1}_{urygb}\hat{y}^x_{urygb} = [\beta_{0u} \quad \beta_{1u} \quad \beta_{2u} \quad \beta_{3u}] \begin{bmatrix} \hat{\phi}^{-1}_{drygb} \\ \hat{\phi}^{-1}_{urygb}(cohort_{urygb} - 2008) \\ \hat{\phi}^{-1}_{urygb}(grade_{urygb} - 5.5) \\ \hat{\phi}^{-1}_{urygb}(math_{urygb} - .5) \end{bmatrix} + \hat{\phi}^{-1}_{urygb} e_{urygb} \tag{A-2.2}$$

Level-2:

$$\begin{aligned} \beta_{0u} &= \gamma_{00} + \nu_{0u} \\ \beta_{0u} &= \gamma_{10} + \nu_{1u} \\ \beta_{0u} &= \gamma_{20} + \nu_{2u} \\ \beta_{0u} &= \gamma_{30} + \nu_{3u} \end{aligned}$$

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

$$(\boldsymbol{V}^{OLS}_u)^{-1} = (\boldsymbol{V}^{EB}_u)^{-1} - \hat{\boldsymbol{\tau}}^{-2}. \tag{A-2.3}$$

Appendix B: Covariates

1. List of Raw ACS Tables Used for SES Composite

Table Description	Table ID	Universe	Description	Usage	Derived Construct
Median household income	B19013	Households	median family income in the past 12 months	we adjust the reported median income for inflation (2012 constant dollars)	Median Income
Median household income	B19013B	Families with a householder who is Black or African American alone	median family income in the past 12 months	we adjust the reported median income for inflation (2012 constant dollars)	White Median Income
Median household income	B19013H	Families with a householder who is white alone (not Hispanic or Latino)	median family income in the past 12 months	we adjust the reported median income for inflation (2012 constant dollars)	Hispanic Median Income
Median household income	B19013I	Families with a householder who is Hispanic or Latino	median family income in the past 12 months	we adjust the reported median income for inflation (2012 constant dollars)	Black Median Income
Sex by Educational Attainment for the Population 25 and Older	B15002	Population 25 years and over	counts of number of individuals that fall into each of 16 educational attainment categories, by sex	we use the counts of men and women with a bachelor's degree or higher along with the total count to generate the BA+ rate	Bachelor's Degree Rate
Sex by Educational Attainment for the Population 25 and Older	C15002B	Black or African American alone population 25 years and over	counts of number of individuals that fall into each of 4 educational attainment categories, by sex	we use the counts of men and women with a bachelor's degree or higher along with the total count to generate the BA+ rate	Black Bachelor's Degree Rate
Sex by Educational Attainment for the Population 25 and Older	C15002H	White alone, not Hispanic or Latino population 25 years and over	counts of number of individuals that fall into each of 4 educational attainment categories, by sex	we use the counts of men and women with a bachelor's degree or higher along with the total count to generate the BA+ rate	White Bachelor's Degree Rate
Sex by Educational Attainment for the Population 25 and Older	C15002I	Hispanic or Latino population 25 years and over	counts of number of individuals that fall into each of 4 educational attainment categories, by sex	we use the counts of men and women with a bachelor's degree or higher along with the total count to generate the BA+ rate	Hispanic Bachelor's Degree Rate

Poverty Status in the Last 12 Months by Age	B17020	Population for whom poverty status is determined	counts of number of individuals living in households above and below the poverty line in various age bins	we use the counts of those living in poverty that are school aged (6-17 years old)	Poverty Rate, 6-17 Year Olds
Poverty Status in the Last 12 Months by Age	B17020B	Black or African American alone population for whom poverty status is determined	counts of number of individuals living in households above and below the poverty line in various age bins	we use the counts of those living in poverty that are school aged (6-17 years old)	Black Poverty Rate, 6-17 Year Olds
Poverty Status in the Last 12 Months by Age	B17020H	White alone, not Hispanic or Latino population for whom poverty status is determined	counts of number of individuals living in households above and below the poverty line in various age bins	we use the counts of those living in poverty that are school aged (6-17 years old)	White Poverty Rate, 6-17 Year Olds
Poverty Status in the Last 12 Months by Age	B17020I	Hispanic or Latino population for whom poverty status is determined	counts of number of individuals living in households above and below the poverty line in various age bins	we use the counts of those living in poverty that are school aged (6-17 years old)	Hispanic Poverty Rate, 6-17 Year Olds
Sex by Age by Employment Status for the Population 16 and Over	B23001	Population 25 to 64 years	counts of individuals by age, labor market status and employment status	we use the count of those employed divided by the count of those in the labor market for civilians ages 16-64 to compute an unemployment rate	Unemployment Rate
Sex by Age by Employment Status for the Population 16 and Over	C23002B	Black or African American alone, not Hispanic or Latino population 16 years and over	counts of individuals by age, labor market status and employment status	we use the count of those employed divided by the count of those in the labor market for civilians ages 16-64 to compute an unemployment rate	Black Unemployment Rate
Sex by Age by Employment Status for the Population 16 and Over	C23002H	White alone, not Hispanic or Latino population 16 years and over	counts of individuals by age, labor market status and employment status	we use the count of those employed divided by the count of those in the labor market for civilians ages 16-64 to compute an unemployment rate	White Unemployment Rate
Sex by Age by Employment Status for the Population 16 and Over	C23002I	Hispanic or Latino population 16 years and over	counts of individuals by age, labor market status and employment status	we use the count of those employed divided by the count of those in the labor market for civilians ages 16-64 to compute an unemployment rate	Hispanic Unemployment Rate

Receipt of Food Stamps/SNAP in the past 12 months by poverty status in the past 12 months for households	B22003	Households	counts of households receiving food stamps/SNAP benefits by poverty status	we use the counts of households receiving SNAP divided by the total number of households to compute the SNAP rate	SNAP Rate
Receipt of Food Stamps/SNAP in the past 12 months by poverty status in the past 12 months for households	B22005B	Households with a householder who is Black or African American alone	counts of households receiving food stamps/SNAP benefits by poverty status	we use the counts of households receiving SNAP divided by the total number of households to compute the SNAP rate	Black SNAP Rate
Receipt of Food Stamps/SNAP in the past 12 months by poverty status in the past 12 months for households	B22005H	Households with a householder who is White alone, not Hispanic or Latino	counts of households receiving food stamps/SNAP benefits by poverty status	we use the counts of households receiving SNAP divided by the total number of households to compute the SNAP rate	White SNAP Rate
Receipt of Food Stamps/SNAP in the past 12 months by poverty status in the past 12 months for households	B22005I	Households with a householder who is Hispanic or Latino	counts of households receiving food stamps/SNAP benefits by poverty status	we use the counts of households receiving SNAP divided by the total number of households to compute the SNAP rate	Hispanic SNAP Rate
Household Type	B11001	Households	counts of different types of households	we use the count of family households with a female householder, no husband present divided by the total number of family households	Female Headed Household Rate
Household Type	B11001B	Households with a householder who is Black or African American alone, not Hispanic or Latino	counts of different types of households	we use the count of family households with a female householder, no husband present divided by the total number of family households	Black Female Headed Household Rate
Household Type	B11001H	Households with a householder who is White alone, not Hispanic or Latino	counts of different types of households	we use the count of family households with a female householder, no husband present divided by the total number of family households	White Female Headed Household Rate
Household Type	B11001I	Households with a householder who is Hispanic or Latino	counts of different types of households	we use the count of family households with a female householder, no husband present divided by the total number of family households	Hispanic Female Headed Household Rate

2. Measurement Error, Attenuation Bias and Solutions

Formally, attenuation bias can be specified as follows. As an example, consider the true relationship between race-specific achievement and socioeconomic status we would like to estimate:

$$Y_g = \beta_{0g} + \beta_{1g}(SES_g) + \varepsilon_g \tag{B-2.1}$$

Where *Y* is White or non-White minority achievement in a unit (district, county, or metropolitan area) (g indexes group), and *SES* is the average socioeconomic status of the group. Race specific SES is measured with error and measurement error will be larger in units with relatively smaller sample sizes of non-White minorities. Thus, the data we observe are $W_g = SES_g + \varepsilon_g$. In this case, the bias in $\beta_{1g}$is known as attenuation bias. This bias can by quantified by multiplying by the variable's reliability $\lambda = \frac{var(SES_g)}{var(SES_g)+\sigma_1^2}$, i.e., the true variance of the variable SES_g relative to the true variance plus the variance of the measurement error.

To address attenuation bias, we use regression calibration, which makes use of the fact that the measurement error in SES_g (and consequently $SESGap$) are known from Census data.[24] Regression calibration is a method that replaces the error-prone variable W with its best linear prediction (blp). The best linear predictor of $SESGap$ can be defined as:

$$SESp_g^{blp} = E(SES_g) + \frac{cov(SES_g, W_g)}{var(W_g)}\left(W_g - E(W_g)\right)$$

$$= \mu + \frac{cov(SES_g, SES_g + \varepsilon_g)}{\sigma_{SES_g}^2 + \sigma_g^2}(W_g - \mu)$$

[24] Specifically, the ACS reports margins of error which can be easily converted standard errors for each Census variable. Appendix B3: Computing the sampling variance of sums of ACS variables provides a full description of how standard errors for cross-tabulated Census data are constructed.

$$= \mu + \lambda(W_g - \mu) \tag{B-2.2}$$

Note that SES_g^{blp} is "shrunken" towards the mean value of SES_g as a function of λ which, recall, is equal to the reliability of the variable SES_g and can be estimated as a random effect (or empirical Bayes estimate) from a generalized linear model.

Now, we show that regressing Y_g on SES_g^{blp} results in consistent estimates of β_{1g}.

$$\begin{aligned}\frac{cov\left(Y_g, \mu + \lambda(W_g - \mu)\right)}{var\left(\mu + \lambda(W_g - \mu)\right)} &= \frac{cov(Y_g, \lambda W_g)}{\lambda^2\left(\sigma^2_{SES_g} + \sigma^2_g\right)} \\ &= \frac{cov(Y_g, SES_g)}{\lambda\left(\sigma^2_{SES_g} + \sigma^2_g\right)} \\ &= \frac{cov(Y_g, SES_g)}{\sigma^2_{SES_g}} = \beta_{1g}\end{aligned} \tag{B-2.3}$$

3. Computing the sampling variance of sums of ACS variables

In each unit we are given counts in K cells: $\widehat{n1}_d, \widehat{n2}_d, \dots, \widehat{nK}_d$; we also know total counts t_d; we also have margins of error of the counts

$$MoE(\widehat{n1}_d), MoE(\widehat{n2}_d), \dots, MoE(\widehat{nK}_d).$$

We then compute the sampling variances of the

$$var(\widehat{nk}_d) = \left[\frac{MOE(\widehat{nk}_d)}{1.645}\right]^2$$

from these we compute

$$\widehat{pk}_d = \frac{\widehat{nk}_d}{t_d}$$

and

$$var(\widehat{pk}_d) = \frac{var(\widehat{nk}_d)}{t_d^2}.$$

We do not know the sampling rate in unit d; let's call it r_d. If the estimates come from a simple random sample, we would have

$$var(\widehat{pk}_d)^* = \frac{pk_d(1 - pk_d)}{r_d t_d}$$

The estimated design effect in district d for variable k is then

$$\widehat{Dk}_d = \frac{var(\widehat{pk}_d)}{var(\widehat{pk}_d)^*}$$

We can compute the average design effect in unit d as

$$D_d = \frac{1}{K}\sum_{k=1}^{K} \widehat{Dk}_d$$

Now we compute

$$\hat{P}_d = \frac{1}{t_d}\sum_{k=1}^{K}\widehat{nk}_d = \sum_{k=1}^{K}\widehat{pk}_d$$

We want to know $var(\hat{P}_d)$. If we had a simple random sample, we would have

$$var(\hat{P}_d)^* = \frac{P_d(1-P_d)}{r_d t_d}$$

Given the design effect in unit d, however, we would expect this to be inflated by a factor D_d.

So, we have:

$$\begin{aligned}
var(\hat{P}_d) &= D_d var(\hat{P}_d)^* \\
&= D_d \frac{P_d(1-P_d)}{r_d t_d} \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\widehat{Dk}_d\right]\frac{P_d(1-P_d)}{r_d t_d} \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\frac{var(\widehat{pk}_d)}{var(\widehat{pk}_d)^*}\right]\frac{P_d(1-P_d)}{r_d t_d} \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\frac{r_d t_d var(\widehat{pk}_d)}{pk_d(1-pk_d)}\right]\frac{P_d(1-P_d)}{r_d t_d} \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\frac{var(\widehat{pk}_d)}{pk_d(1-pk_d)}\right]P_d(1-P_d) \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{nk_d}\right]P_d(1-P_d) \\
&= \frac{1}{\tilde{n}_d}P_d(1-P_d)
\end{aligned}$$

where $nk_d = \frac{pk_d(1-pk_d)}{var(\widehat{pk}_d)}$ is the effective sample size in cell k in unit d (the sample size nk_d such that $\frac{pk_d(1-pk_d)}{nk_d} = var(\widehat{pk}_d)$), and $\tilde{n}_d = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{nk_d}\right)^{-1}$ is the harmonic mean of the effective sample sizes across cells within unit d. Note that $\frac{\tilde{n}_d}{t_d} = \tilde{r}_d$ is the harmonic mean of the effective sampling rate across cells within d.

An alternate approach is to assume a common design effect across units

$$var(\hat{P}_d) = D_d var(\hat{P}_d)^*$$

$$= D_d \frac{P_d(1 - P_d)}{r_d t_d}$$

$$= D \frac{P_d(1 - P_d)}{r_d t_d}$$

where $D = \frac{1}{T}\sum_{j=1}^{J} t_j D_j$ is the average design effect across units (weighted by unit size to increase precision). We can write

$$D = \frac{1}{T}\sum_{j=1}^{J} t_j D_j$$

$$= \frac{1}{T}\sum_{j=1}^{J} t_j \left[\frac{1}{K}\sum_{k=1}^{K} \frac{r_j t_j}{nk_j}\right]$$

$$= \sum_{j=1}^{J} \frac{t_j}{T}\frac{r_j}{\tilde{r}_j}$$

So then,

$$var(\hat{P}_d) = D_d var(\hat{P}_d)^*$$

$$= D_d \frac{P_d(1 - P_d)}{r_d t_d}$$

$$= D \frac{P_d(1 - P_d)}{r_d t_d}$$

$$= \left[\sum_{j=1}^{J} \frac{t_j}{T} \frac{r_j}{\tilde{r}_j} \right] \frac{P_d(1 - P_d)}{r_d t_d}$$

$$= \left[\sum_{j=1}^{J} \frac{t_j}{T} \frac{r_j t_d}{\tilde{r}_j t_d} \right] \frac{P_d(1 - P_d)}{r_d t_d}$$

Assume r_j is constant across units and assume the effective sampling rate in unit j is independent of the unit size t_j; then this simplifies to

$$var(\hat{P}_d) = \frac{P_d(1 - P_d)}{t_d \tilde{r}},$$

where

$$\tilde{r} = \left[\sum_{j=1}^{J} \frac{t_j}{T} \frac{1}{\tilde{r}_j} \right]^{-1}$$

is the (weighted) harmonic mean of the effective sampling rates. We can compute $\tilde{r}$ without knowing the actual sampling rates:

$$\tilde{r} = \left[\sum_{j=1}^{J} \frac{t_j}{T} \frac{1}{\frac{1}{t_j} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{var(\widehat{pk}_j)}{pk_d(1 - pk_j)} \right)^{-1}} \right]^{-1}$$

$$= \left[\sum_{j=1}^{J} \frac{t_j^2}{T} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{var(\widehat{pk}_j)}{pk_d(1 - pk_j)} \right) \right]^{-1}$$

To recap, we have two approaches to compute the sampling variance of $\hat{P}_d$:

1. For each unit, compute the harmonic mean of the effective sample size

$$\tilde{n}_d = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{var(\widehat{pk}_d)}{pk_d(1 - pk_d)} \right)^{-1}$$

then

$$Var(\hat{P}_d) = \frac{P_d(1 - P_d)}{\tilde{n}_d}.$$

Or:

2. Compute the weighted harmonic mean of the effective sampling rate across units (using any of these formulas, all identical):

$$\tilde{r} = \left[\sum_{j=1}^{J} \frac{t_j}{T} \frac{1}{\tilde{r}_j}\right]^{-1}$$

$$= \left[\sum_{d=1}^{D} \frac{t_d^2}{T} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{var(\widehat{pk}_d)}{pk_d(1 - pk_d)}\right)\right]^{-1}$$

$$= \left[\frac{1}{(1.645^2)TK} \sum_{d=1}^{J} \sum_{k=1}^{K} \frac{MoE(\widehat{nk}_d)^2}{pk_d(1 - pk_d)}\right]^{-1}$$

then

$$Var(\hat{P}_d) = \frac{P_d(1 - P_d)}{\tilde{r} t_d}.$$

The first approach allows a different design effect in each unit, but the design effect is probably noisily estimated, so will have more noise in the estimated sampling variances. The second assumes a common design effect across units. Our decision criteria for generating sampling variances is as follows:

1. When $K = 1$ and $P_d > 0$, use the sampling variance provided by ACS, i.e., $var(\hat{p}_d) = \frac{var(\hat{n}_d)}{t_d^2}$

2. When $K = 1$ and $P_d = 0$, use the sampling variance method 2, i.e., $Var(\hat{P}_d) = \frac{P_d(1-P_d)}{\tilde{r}t_d}$, where $P_d = \frac{1}{t_d}$.
3. When $K > 1$ and $P_d > 0$, use the sampling variance method 2, i.e., $Var(\hat{P}_d) = \frac{P_d(1-P_d)}{\tilde{r}t_d}$
4. When $K > 1$ and $P_d = 0$, use the sampling variance method 2, i.e., $Var(\hat{P}_d) = \frac{P_d(1-P_d)}{\tilde{r}t_d}$, where $P_d = \frac{1}{t_d}$.

4. Estimating sampling variance of composite SES measures

Let $\bar{\bar{\mathbf{X}}}_d$ be the vector of 6 variables we use to construct the SES composite in unit d. Let $\mathbf{W}_d$ be the diagonal matrix containing the standard errors of $\widehat{\mathbf{X}}_d$.[25]

Our estimated SES composite (S) in unit d is

$$\hat{S}_d = \bar{\bar{\mathbf{X}}}_d \mathbf{B},$$

where $\mathbf{B}$ is a 6×1 vector of unstandardized coefficients. The sampling variance of $\hat{S}_d$ is

$$var(\hat{S}_d) = \mathbf{B}' \mathbf{V}_d \mathbf{B},$$

where $\mathbf{V}_d$ is the covariance matrix of $\widehat{\mathbf{X}}_d$. We know the diagonal elements of $\mathbf{V}_d$ ($\mathbf{W}_d$); but not the off-diagonals. We need to know $\mathbf{V}_d$ to get the standard error of $\hat{S}_d$. How can we compute $\mathbf{V}_d$?

Define $\mathbf{R}_d$, the correlation matrix describing the correlations of the estimates $\widehat{\mathbf{X}}_d$. If we knew $\mathbf{R}_d$, then we can get

$$\mathbf{V}_d = \mathbf{W}_d \mathbf{R}_d \mathbf{W}_d.$$

The key is getting an estimate of $\mathbf{R}_d$. We can use PUMS data to estimate $\mathbf{R}$ empirically (via bootstrapped samples). We do this as follows:

a. Set $N = 5{,}000$, and $J = 1{,}000$ (or some other values)

b. Pick PUMA k.

c. From all families in PUMA k, draw a random sample of N families.

[25] Note that we get the standard errors of these variables from ACS. The exception is ln(median income), as we get a standard error for median income. Let $\widehat{M}_d$ be the estimated median income in unit d. The Delta method gives us

$$se[\ln(\widehat{M}_d)] \approx \frac{1}{\widehat{M}_d} se(\widehat{M}_d).$$

d. Compute $\widehat{\mathbf{X}}_k$ from the micro-data (so if $\mathbf{X}$ includes ln(median income), then estimate ln(median income) in PUMA k from the sample, and likewise for the 6 variables we include in $\mathbf{X}$).

e. Repeat (c) and (d) J times for PUMA k.

f. Estimate $\widehat{\mathbf{R}}_k^B$ from the J samples

g. Repeat (b)-(f) for all PUMAs $k = 1, \dots, K$.

h. Repeat (b)-(g) for each race/ethnic group r to get $\widehat{\mathbf{R}}_{kr}^B$. We might need to set $N = 1{,}000$ for race-ethnic groups, because race samples are smaller in each PUMA.

Next we examine how $\widehat{\mathbf{R}}_k$ and $\widehat{\mathbf{R}}_{kr}$ vary across PUMAs and race/ethnic groups. If $\widehat{\mathbf{R}}_k$ and $\widehat{\mathbf{R}}_{kr}$ are relatively constant across PUMAs and subgroups, we can just use a single common value of $\widehat{\mathbf{R}}$ for all units and subgroups. We find that they are generally similar, so we use a common $\widehat{\mathbf{R}}$ in all PUMAs.