The Educational Opportunity Project
at Stanford University

Stanford Education Data Archive Technical Documentation

SEDA2023

January 2024

Erin M. Fahle, Stanford University

Sean F. Reardon, Stanford University

Benjamin R. Shear, University of Colorado Boulder

Andrew D. Ho, Harvard University

Jie Min, Stanford University

Demetra Kalogrides, Stanford University

Suggested citation:

Fahle, E. M., Reardon, S. F., Shear, B. R., Ho, A. D, Min, J. & Kalogrides, D. (2024). Stanford Education Data Archive Technical Documentation: SEDA2023. Retrieved from https://purl.stanford.edu/xt779fj2637.

Table of Contents

What is SEDA2023?

The Stanford Education Data Archive (SEDA) is created by the Educational Opportunity Project (EOP) at Stanford University (https:\\edopportunity.org). The EOP aims to generate and share data and research that can help scholars, policymakers, educators, and parents learn how to improve educational opportunities for all children. SEDA is the flagship data product of the EOP; it showcases how state accountability test data can be used to study educational opportunity in the U.S.

SEDA2023 is a special release of the Stanford Education Data Archive. It is part of a larger partnership with Harvard University's Center for Education Policy Research. This release is designed to provide insight into how school district average achievement in 2023, three years after the onset of the COVID-19 pandemic, compares to achievement in 2022, two years after the onset of the pandemic, and to achievement in 2019, the year prior to the pandemic. Estimates of test score changes between 2019 and 2022 can shed light on the extent of pandemic-related losses, while estimates of test score changes between 2022 and 2023 can provide insight into the extent of academic recovery that occurred once schools returned to normal operations in fall 2022.

Acknowledgements

The construction of SEDA2023 was supported by a grant from the Bill and Melinda Gates Foundation. Our research using the SEDA2023 data was supported by the Carnegie Corporation of New York, Bloomberg Philanthropies, and Kenneth C. Griffin. The state assessment data used in constructing the SEDA files were gathered from individual states. The National Center for Education Statistics (NCES) and the National Assessment Governing Board provided data on student achievement by state which we used to rescale state proficiency data. The findings and opinions expressed in our research and reported here are those of the authors alone; they do not represent the views of the U.S. Department of Education, NCES, individual states, or any of the aforementioned funding agencies.

We would like to thank: Pete Claar at SchoolDigger (https://www.schooldigger.com/) for his assistance gathering data from state websites; Sadie Richardson, Jiyeon Shim, and Julia Paris for cleaning the 2022 and 2023 state data; Jim Saliba for assembling the 2019 ED*Facts* data; Dan Dewey for gathering information on test score changes; and Tom Kane for his partnership in this effort.

Data Use Agreement

Prior to downloading the data, users must sign the data use agreement (https://edopportunity.org/get-the-data/).

Source Data

State Accountability Test Data

The state accountability data used to construct SEDA2023 test score estimates come from two primary sources: (1) the ED*Facts* data system; and (2) state-reported accountability data.

ED*Facts*. 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. States have the flexibility to select or design a test 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. *No individual student-level data are reported*.

ED*Facts* data on school assessment outcomes are available for eleven consecutive school years from 2008-09 through 2018-19 in grades 3 to 8, and one grade in high school, in RLA and math.[1] The student subgroups include race/ethnicity, gender, and socioeconomic disadvantage, among others. The raw ED*Facts* data used in SEDA include no suppressed cells, nor do they have a minimum cell size for reporting. The data are reported by school, subject, grade, year, and subgroup and include schools in every state. For SEDA2023, we use the 2015-16 through 2018-19 school-level proficiency data in grades 3-8 for all students, as well as the Black, Hispanic, White, economically disadvantaged (ECD), and not economically disadvantaged (not-ECD) subgroups.

One of the requirements for the inclusion of a given state-subject in SEDA2023 is the existence of 2018-19 ED*Facts* data.[2] All state-subjects are represented in the 2018-19 data, except West Virginia RLA, which was removed by ED*Facts* due to data concerns (*State Assessments in Reading/Language Arts and Mathematics- School Year 2018-19* ED*Facts Data Documentation*, n.d.).

State-reported accountability data. Because ED*Facts* has not released 2021-22 or 2022-23 school year data at this time, we rely on state-reported data for these years. Most states publicly report their school and/or district proficiency data as part of federal accountability. We collect these data in multiple

[1] Data are also available for the 2020-21 school year; however, due to limited testing in that year due to the COVID-19 pandemic, data are not comparable to other years. Data for 2021-22 and 2022-23 were not available at the time of publication of SEDA2023.
[2] We do not require complete data in earlier years (2015-16 through 2017-18).

ways. For most states, we use data scraped from state public websites by SchoolDigger or the EOP team. Two states provided data upon request that was not released on their website to the EOP team for this purpose.

Notably, not all states reported usable data. To be used in our estimation process, the 2022 and 2023 data must: (a) include *at least three* proficiency categories; and (b) be disaggregated by school or school district,[3] subgroup, subject, grade, and year. Additionally, to include 2023 data in our process, we require that the state test and proficiency cutscores did not change between 2022 and 2023.[4] A list of states and reasons for exclusion from SEDA2023 appears in **Table 1**.

The usable data that states report also varied in content. There are three patterns of data reporting:

(1) <u>Number of students scoring in each proficiency category</u>. These are the necessary student-level data and were used "as is."
(2) <u>Total number of students tested and the percent scoring in each proficiency category</u>. From these data, we derived the counts in each category (with some rounding error).
(3) <u>Only the percent scoring in each proficiency category</u>. We estimate the total number of test takers (in each school district, subgroup, subject, grade, and year) using the 2022 Common Core of Data (CCD)[5] and the 2019 ED*Facts* data.[6] We then derived an estimated count scoring in each category. In the downloadable datafiles, we flag all cells where we estimated counts.

Additionally, not all states reported data for all subgroups. States also used different suppression rules to protect student privacy. Some states suppress entire rows of data that do not meet reporting thresholds typically based on sample size, while others use partial suppression (suppressing some cells of data within a row). The extent and type of suppression affected our methodology for estimating district average scores, and sometimes prohibited our ability to produce estimates for individual districts. **Table 2**

---

[3] In all but one state-year, we used data disaggregated by school districts. In Pennsylvania, we used data disaggregated by schools because school district files were unavailable in 2022.
[4] Because there was no NAEP assessment in 2023, we rely on the NAEP linkage in 2022 to construct comparable 2023 estimates across states. For this approach to be accurate, states must not have changed their tests or proficiency cutscores between 2022 and 2023. We describe this process and corresponding assumptions more under Step 3.
[5] At the time of data construction, the 2023 CCD was not yet available.
[6] Specifically, we estimate the number of test takers in a district-subgroup-subject-grade as the number of enrolled students in the district-subgroup-subject-grade from the CCD 2022 multiplied by the participation rate for that district-subgroup-subject-grade from the 2019 ED*Facts* data. If the number of enrolled students is missing for a given district-subgroup-grade in 2021-2022, then imputed values are used.

also includes information on the subgroups reported in the data, whether counts were estimated using CCD data, and whether the state used partial suppression.

National Assessment of Educational Progress Data

Because different states use different tests and proficiency thresholds, the test score estimates derived from the above data sources are not readily comparable across states, grade, or years. Therefore, we also draw on the National Assessment of Educational Progress Data (NAEP) 2015, 2017, 2019 and 2022 national and state assessment data in 4th and 8th grade math and reading to link the estimates to a scale that is comparable among states and over grades and years. Specifically, we use the NAEP Expanded Population Estimates.

Construction

Construction of the SEDA2023 test score estimates occurs in a series of steps. These steps are largely similar to those described in the SEDA 4.1 Technical Documentation. Here we provide an overview of the process and highlight where the SEDA2023 process deviates from the SEDA 4.1 process.

Notation

- Test score estimates are denoted as $\hat{y}$ (unit-subgroup-subject-year-grade or "long form" estimates) or $\hat{Y}$ (unit-subgroup-subject-year or "annual" estimates). Their standard errors are denoted as $se(\hat{y})$ and $se(\hat{Y})$, respectively.
- A <u>subscript</u> indicates the aggregation of the estimate. We use the following subscripts:
  - u = unit (generic)
    - d = administrative school district
    - f = state
  - s = subgroup
    - all = all students
    - blk = Black
    - hsp = Hispanic
    - wht = White
    - ecd = economically disadvantaged
    - nec = not economically disadvantaged
  - y = year
  - b = subject
  - g = grade
- A <u>superscript</u> indicates the scale of the estimate. The metric is generically designated as $\boldsymbol{x}$. There are four scales. The first scale is only used in construction. The latter two scales are reported:
  - cs = cohort scale metric
  - ys = year scale metric
  - gys = grade within year scale metric

- An asterisk (*) indicates an Empirical Bayes (shrunken) estimate. The absence of an asterisk indicates a least squares (OL) estimate.

Step 1: Creating the 2019 School-to-District Crosswalk

This step links each public school to an administrative school district (e.g., NCES leaid) in 2019. A school-to-leaid crosswalk is not needed in 2022 or 2023 as the proficiency data are already reported by administrative districts.[7,8]

This crosswalk process deviates from that used for SEDA 4.1, which links schools to geographic school districts. Administrative districts differ from the geographic districts in two ways. First, for geographic school districts in SEDA 4.1, we "reassign" charter schools, magnet schools, and schools operated by secondary districts to the district in which they are physically located (regardless of the entity that operates the schools). Second, we exclude schools classified as "Special Education" from geographic districts and combine them into statewide special education districts. For SEDA2023 administrative districts, we do not reassign schools; charter, magnet, secondary, and special education schools are attached to the traditional public or charter district that operates them. For more information on geographic districts, we refer to you to the SEDA 4.1 Technical Documentation.

The use of administrative districts (rather than geographic districts) is preferred for SEDA2023 for two reasons. First, one of the aims of SEDA2023 is to help school districts understand their learning recovery needs. Administrative districts have authority to set policy for their schools, as such it is most useful for the estimates to reflect only the schools under their operation. Second, to construct geographic school districts, we need data for individual charter schools. While many states report such data in 2022, data for many schools is suppressed due to the small numbers of students taking assessments. Because of this we cannot reliably construct geographic school district estimates for the 2022 school year.

Step 2: Data Cleaning

This step removes data not used in the cutscore estimation process and prepares data for use in the following step.

<u>State-subject-grade-year removals.</u> There are two primary reasons why a state-subject-grade was removed from the data prior to estimation:

---

[7] For Pennsylvania in 2022 where we used school-level data, we use the NCES leaids reported in the 2022 CCD data to aggregate the schools to districts.

[8] While many states report data by school in 2022 and 2023, there is too much suppression to reliably construct district estimates from the school level data in most states. Because of this, it is possible that the set of schools assigned to a given district in 2019 is not identical to that in 2022 or 2023. We acknowledge this limitation.

1. Test participation in a state-subject-grade-year was too low. We use a threshold of 94%[9] state-wide participation and remove any state-subject-grade-year where the 2019 participation rate falls below this threshold. Notably, the auxiliary participation rate data used to determine state-level removals is not available in 2022 or 2023, so the 2019 state participation was used to remove state-subject-grade-year in all years.[10]
2. More than 5% of students took a test that is not the primary grade-level state assessment used for accountability. This occurs for two reasons. First, in some states students may take an end-of-course assessment. This is common in 8th grade math, when a subset of students takes the Algebra I test in place of the 8th grade math assessment. Second, a Spanish-language version of an assessment may be used in some subjects and grades. Spanish- and English-language versions of the assessments are not always equivalent, particularly when assessing reading and language skills. Because our estimation methodology relies on the fact that all students took a common test (within a state-subject-grade-year), we remove cases where fewer than 95% of students took the primary state accountability assessment for a given state-year-subject-grade. We identified these cases using two types of information:
    a. State-reported information on the number of students taking each type of assessment. If more than 5% of students were reported to take the non-primary test in a subject-grade-year cell, it was removed.
    b. State-level data on the number of students tested by grade-subject. If the reported number tested in a grade-subject-year cell was substantially lower than other grade subjects in the same year (suggesting a lower testing rate), it was removed.

A list of the states-subject-grade-years excluded from data construction are shown in **Table 3**.

Data cleaning. For the 2019 data, we combine performance data for regular and alternate assessments. For 2022 and 2023, we use data as reported, which may or may not include alternative assessments.

[9] We use 94% (rather than 95% as done in SEDA 4.1) to preserve as much sample as possible. This affects a single state in our data, Oregon, where the participation rate in three subject-grades is between 94% and 95% in 2019. By including these cells with slightly lower participation, we can produce more estimates that meet our reliability thresholds.
[10] The exception to this is Colorado, which was removed entirely from SEDA2023 due to low state-reported test participation in 2022.

Step 3: Estimating and Linking Cutscores

In this step, we use Heteroskedastic Ordered Probit (HETOP) models or the inverse cumulative standard normal distribution function to estimate the state-grade-subject-year cutscores. Cutscore estimation is done using data for all students (not separately by subgroup). For all state-subject-grades in 2019, we use the HETOP model to estimate the cutscores. In 2022, we use the HETOP model when the state-subject-grade data meets the following requirements: (1) counts are not estimated; and (2) the district-level data represents at least 95% of the students in the state, subject, and grade (e.g., there is limited suppression). When either of these requirements is not met, we calculate the cutscores from the state proficiency count data using the inverse cumulative standard normal distribution function.

We then link the 2016-2022 estimated cutscores to the NAEP scale and standardize the NAEP-linked cutscores relative to a reference cohort of students using the same process as SEDA 4.1. The resulting cutscores are in the Cohort Standardized (CS) scale and are comparable across states and years under the linking assumptions reviewed in the 4.1 documentation. Additional detail on the HETOP estimation and NAEP linking process can be found in Reardon, Shear, Castellano, and Ho (2017) and Reardon, Kalogrides, and Ho (2021).

We must use a different process in 2023, because we do not have 2023 NAEP data. Instead, for 2023, we use the 2022 linked proficiency thresholds. For this approach to enable accurate comparisons of 2022-2023 test score changes among districts in *the same state*, states' test score scales and proficiency thresholds must be comparable from 2022 to 2023. We exclude 2023 data for states where we found evidence of changes in state assessments. For this approach to enable accurate comparisons of 2022-2023 test score changes among districts *in different states*, we also assume that state test score trends from 2022 to 2023 are comparable to unmeasured NAEP trends from 2022 to 2023.

Step 4: Selecting and Preparing Data for Mean Estimation

This step selects data for unit-subgroup-subject-grade-year cases that will be used in estimation. For 2019, the same rules are used in SEDA2023 and SEDA 4.1. These are as follows:

1. <u>The participation rate is less than 95%.</u> 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.
2. <u>Incomplete data reported by student demographic subgroups</u>. There are a small number of cases where the total number of test scores reported by race is less than 95% of the total reported test scores for all students. We are concerned about subgroup data quality in these cases.

3. More than 40% of students take alternate assessments. Measurement error may affect district-subgroup-subject-grade-year cases where students take alternate assessments. These assessments typically differ from the regular assessment and generally have different proficiency thresholds or meanings. The threshold for exclusion is 40%.
4. 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.
5. District-subgroup-subject-grade-year cells that do not meet the minimum statistical estimation requirements. When all cells for a district are insufficient (e.g., have all observations in a single middle category; have all observations in only 2 adjacent categories; have only 2 proficiency categories (one cutscore); or have all observations in only the top and bottom categories) or small (have fewer than 100 test scores) we do not have sufficient information to produce mean and standard deviation estimates.

For 2022 and 2023, the information to make exclusions based on participation, representation, and alternate assessments is not available. Thus, we only make removals based on 4 and 5, above.

Step 5: Estimating Means

District Mean Estimates. In this step, we use the pooled HETOP model to estimate district-subgroup-subject-grade-year means and standard deviations, along with their standard errors. The input data to these models are the cutscores from Step 3 and the data prepared in Step 4. We estimate two pooled HETOP models in 2009-2019 data (one for math and one for RLA) and two pooled HETOP models for 2022- 2023 data (one for math and one for RLA). As noted above, in both 2022 and 2023, we used the linked 2022 cutscores. The resulting district-subgroup-subject-grade mean estimates are on the CS scale, and are comparable across states, within subjects and grades (under the assumptions noted in Step 3). For more details on the pooled HETOP model, see the SEDA 4.1 Technical Documentation and Shear and Reardon (2021).

Deviating from SEDA 4.1, we also construct two standard errors for the estimates—one that does not account for the NAEP uncertainty and one that does account for the NAEP uncertainty.

State Mean Estimates. To construct state-all estimates in 2015-2022, we use the NAEP data. We do not interpolate the missing grades (3, 5, 6, 7) and years (2016, 2018) prior to the pooling estimation step. Because we use the NAEP data for these estimates, the standard errors do not require any adjustment—they already account for NAEP uncertainty.

For the state-all estimates in 2023 and the state-subgroup estimates in all years, we use the process described above for districts. For these estimates, we produce both unadjusted and adjusted standard errors at this step.

Step 6: Creating Reporting Scales

This step creates the two scales reported in SEDA2023: the Year Standardized (YS) and the Grade Year Standardized (GYS) scales.

To create the YS scale, we standardize the estimates to the 2019 national average in each grade and subject. In this scale, each unit is equivalent to a 2019 national standard deviation in the same subject and grade.

To create the GYS scale, we first approximate the average amount student test scores grow in a grade on NAEP using the 4^{th} and 8^{th} grade estimates by subject in 2019. We calculate the amount the test scores changed between 4th and 8th grade as the average score in 8th grade in 2019 minus the average score in 4th grade in 2019. Then, to get an estimate of per-grade differences, we divide that value by 4. We scale the data using these parameters, such that in the GYS scale each unit is interpretable as 1 grade level referenced to the 2019 national population.

Step 7A: Pooling Mean Estimates – OL Annual Estimates

We use a new pooling strategy in SEDA2023. The goal of the first step of pooling is to construct subject-specific OL estimates, in grade 5.5, for each year, called "OL annual estimates." These OL annual estimates and their unadjusted standard errors will serve as the input data for the models in Step 7B where we construct EB estimates.

Our methods for producing OL annual estimates differ slightly for states and districts; we describe each in turn below.

State Estimates

For the state-all annual estimates in 2015, 2017, 2019, and 2022, we take a weighted average of the 4^{th} and 8^{th} grade state NAEP estimates (rescaled to the GYS or YS scale, described above) to get an estimate of the average test score in grade 5.5. We then interpolate the 2016 and 2018 estimates as follows:

$$\hat{Y}^{x}_{fsby} = \frac{\left(\hat{Y}^{x}_{fsb(y-1)} + \hat{Y}^{x}_{fsb(y+1)}\right)}{2}$$

$$se\left(\hat{Y}^{x}_{fsby}\right) = \frac{\sqrt{\left(se\left(\hat{Y}^{x}_{fsb(y-1)}\right)^{2} + se\left(\hat{Y}^{x}_{fsb(y+1)}\right)^{2}\right)}}{2}$$

For the state-all annual estimate in 2023 and the state-subgroup annual estimates in all years, we estimate a model of the following form, separately by subject and subgroup, using the HETOP state-subgroup-subject-grade-year estimates, $\hat{y}^{x}_{fsbyg}$, and their standard errors, $se(\hat{y}^{x}_{fsbyg})$, from Step 5. Models are estimated separately by subgroup, subject, and scale. For parsimony, we have omitted subject and subgroup subscripts in the notation.

$$\hat{y}^{x}_{(fy)g} = \beta_{0(fy)} + \beta_{1(fy)}(grade - 5.5) + \varepsilon_{(fy)g}$$
$$\beta_{0(fy)} = \gamma_{00} + u_{0(fy)}$$
$$\beta_{1(fy)} = \gamma_{10} + u_{1(fy)}$$

Where $\varepsilon_{(fy)g} \sim N\left(0, var(\hat{y}^{x}_{(fy)g})\right)$ and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$. From this model, we use the estimates of $\hat{\beta}^{OL}_{0(fy)}$ and $se(\hat{\beta}^{OL}_{0(fy)})$—OL annual estimates in grade 5.5 and their unadjusted standard errors. For these estimates, we also create an adjusted standard error, which accounts for the NAEP linking error. For the 2023 estimates, we adjust the standard error using the 2022 NAEP SE.

At the end of this step, we have OL annual estimates for 2015-2023 for state-all and state-subgroups. For simplicity, we use the notation $\hat{Y}^{x}_{fsby}$ and $se(\hat{Y}^{x}_{fsby})$ to identify the OL annual estimates and their standard errors from both approaches.

District Estimates

We use district-subgroup-subject-year-grade mean test score estimates and their unadjusted standard errors in 2015 through 2023 from Step 5 to estimate the following model separately for each district-subgroup-subject. Models are estimated separately by subgroup, subject, and scale. For parsimony, we have omitted subject and subgroup subscripts in the notation.

$$\hat{y}^{x}_{(dy)g} = \beta_{0(dy)} + \beta_{1(dy)}(grade - 5.5) + \varepsilon_{(dy)g}$$
$$\beta_{0(dy)} = \gamma_{00} + u_{0(dy)}$$
$$\beta_{1(dy)} = \gamma_{10} + u_{1(dy)}$$

In this model, district-subgroup-subject-year-grades are nested in district-subgroup-subject-years; $\varepsilon_{(dy)g} \sim N\left(0, var(\hat{y}^{x}_{(dy)g})\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$. From this model, we use the estimates of $\hat{\beta}_{0(dy)}$ and $se(\hat{\beta}_{0(dy)})$—OL annual estimates in grade 5.5 and their unadjusted standard errors. We also create an adjusted OL SE, accounting for the NAEP linking uncertainty.

District ECD/NEC Estimates Adjustment

Because the proportion of ECD students in a district may change over time, we adjust the ECD and non-ECD OL annual estimates to keep a constant proportion of ECD or NEC students over time. First, we calculate the average proportions of ECD students across our sample time period (2016 to 2019), called $\bar{p}_{ecd}$. We then calculate adjusted ECD and non-ECD means as:

$$\widehat{Ya}^{x}_{d(ecd)y} = \hat{Y}^{x}_{d(all)y} - (1-\bar{p}_{ecd})\left(\hat{Y}^{x}_{d(ecd)y} - \hat{Y}^{x}_{d(nec)y}\right)$$

$$\widehat{Ya}^{x}_{d(nec)y} = \hat{Y}^{x}_{d(all)y} - (\bar{p}_{ecd})\left(\hat{Y}^{x}_{d(nec)y} - \hat{Y}^{x}_{d(ecd)y}\right)$$

This adjustment holds constant the difference or gap between the ECD and non-ECD test score estimates, while adjusting the means to enable better comparisons over time (e.g., from 2019 to 2022) that are not conflated by changing definitions of ECD. We only report the adjusted OL estimates and use them as input data in all subsequent steps.

We make no adjustment to the state-level ECD/NEC data.

Step 7B: Pooling Mean Estimates – EB Annual Estimates

The goal of the second step of pooling is to construct subject-specific EB estimates, in grade 5.5, for each year, called "EB annual estimates." Our methods for producing EB annual estimates differ based on the unit-subgroup; we describe each process in turn below.

State-All Estimates

We fit three separate pooling models to produce the EB estimates. Note that the subgroup subscript has been removed from the notation in this subsection because it focuses on all students (not subgroups). We have also removed the subject subscript for parsimony, as models are estimated separately by subject.

Model 1-S. We first fit a model to get EB estimates of the pre-trend (from 2015 to 2019) in test scores, the change in test scores from 2019 to 2022, and the change in test scores from 2022 to 2023.

$$\hat{Y}^{x}_{fy} = \beta_{0f} + \beta_{1f}(year - 2019)(I(year \leq 2019)) + \beta_{2f}\left(I(year \geq 2022)\right) + \beta_{3f}(year = 2023) + \epsilon_{fy}$$

$$\beta_{0f} = \gamma_{00} + u_{0f}$$

$$\beta_{1f} = \gamma_{10} + u_{1f}$$

$$\beta_{2f} = \gamma_{20} + u_{2f}$$

$$\beta_{3f} = \gamma_{20} + u_{2f}$$

Where $\epsilon_{(fy)g} \sim N\left(0, var\left(\hat{Y}^{x}_{fy}\right)\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$.

From this model, we get an estimate of the fitted trend in scores for 2015-2019 in each state. We also construct the EB estimates of the average scores in 2022 and 2023, as well as their standard errors.

Model 2-S. We then fit a model to get estimates of the EB residuals in each year (2015, 2017 and 2019), which we will use to construct the annual EB estimates in 2015-2019:

$$\hat{Y}^{x}_{fy} = \alpha_{fy} + \varepsilon_{fy}$$
$$\alpha_{fy} = \beta_{0f} + \beta_{1f}(YRC) + r_{fy}$$
$$\beta_{0f} = \gamma_{00} + u_{0f}$$
$$\beta_{1f} = \gamma_{10} + u_{1f}$$

Where $\varepsilon_{fy} \sim N\left(0, var\left(\hat{Y}^{x}_{fy}\right)\right)$; $r_{fy} \sim N(0, \sigma^2)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$.

To construct the EB estimates in 2015, 2017, and 2019, we add these EB residuals to the fitted values in each year from Model 1-S (our best estimate of the fitted trend). To get 2016 and 2018 estimates, we interpolate the EB annual estimates using the same interpolation formulas for the OLs.

Notably, however, the standard errors from this model are not correct. Allowing for an additional error term, r_{ys}, overestimates the imprecision in the EB residuals. We need to estimate one final model to get the standard errors for the EB annual estimates in 2015-2019.

Model 3-S. To get the standard errors in 2015, 2017, and 2019, we fit a model of the form:

$$\hat{Y}^{x}_{fy} = \beta_{0f}(y15) + \beta_{1f}(y17) + \beta_{2f}(y19) + \varepsilon_{fy}$$
$$\beta_{0f} = \gamma_{00} + u_{0f}$$
$$\beta_{1f} = \gamma_{10} + u_{1f}$$
$$\beta_{2f} = \gamma_{20} + u_{1f}$$

Where $\varepsilon_{fy} \sim N\left(0, var\left(\hat{Y}^{x}_{fy}\right)\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$. From this model, we use the standard error of the EB annual estimates in each year. Again, we interpolate the standard errors of EB annual estimates using the same formulas for the OL standard errors (shown above).

At the end of this step, we have estimates in 2015, 2016, 2017, 2018, 2019, 2022, and 2023, and their standard errors, for all students in states with sufficient data, denoted $\widehat{Y^*}^{x}_{fy}$ and $se\left(\widehat{Y^*}^{x}_{fy}\right)$. Note that we do not report 2015 estimates.

District-All Estimates

For the district-all estimates, we use a similar three-model approach to that described for the state-all group above. The deviations are noted here. Again, we have omitted the subgroup and subject subscripts from the notation below for parsimony.

Model 1-D. We fit a model identical to Model 1-S (described above) but using district OL annual estimates from 2016, 2017, 2018, 2019, 2022, and 2023, along with their unadjusted standard errors. We include the proportion of students receiving free and/or reduced-price lunch (from the 2019 CCD) as a predictor on all terms in the model to improve shrinkage. From this model, we get estimates of the pre-trend (from 2016 to 2019), the change from 2019 to 2022, and the change from 2022 to 2023. From these parameters, we also calculate EB estimates of the average test score in 2022 and 2023 and their standard errors.

Notably, unlike states, not all districts have data in all years. For the districts where only 2019, 2022, and/or 2023 data are available, we do not get reliable estimates for the average in 2022 or 2023 from Model 1-D. In these cases, we will use the EB annual estimates for 2019, 2022, and 2023 from Model 3-D below. This affects a very small number of district-subject cases.

<u>Model 2-D.</u> We fit a model identical to Model 2-S (described above) using data for OL annual estimates from 2016, 2017, 2018, and 2019 and their unadjusted standard errors. We include the proportion of students receiving free and/or reduced-price lunch (from the 2019 CCD) as a predictor on all terms in the model. From this model, we get estimates of the EB residuals in each year and construct the EB annual estimates in 2016-2019 by adding these residuals to the fitted values in each year from Model 1-D.

Again, when there is no 2016-2018 data, the EB residual for 2019 from this model is not reliable. We will instead use the EB estimates for 2019 from Model 3-D in these cases.

<u>Model 3-D.</u> We fit a variant of Model 3-S, which includes all years of data. As noted above, we will use this model to both produce the appropriate standard errors for the 2016-2019 EB annual estimates, as well as to construct the EB annual estimates and their standard errors for 2019-2023 in cases where these is limited data and Models 1-D and 2-D do not produce reliable estimates. In these cases, we will use the EB annual estimates for 2019, 2022, and 2023 from this model.

$$\hat{Y}_{dy}^{x} = \beta_{0d}(y16) + \beta_{1d}(y17) + \beta_{2d}(y18) + \beta_{3d}(y19) + \beta_{4d}(y22) + \beta_{5d}(y23) + \varepsilon_{dy}$$
$$\beta_{0d} = \gamma_{00} + \gamma_{01}(frl2019) + u_{0d}$$
$$\beta_{1d} = \gamma_{10} + \gamma_{11}(frl2019) + u_{1d}$$
$$\beta_{2d} = \gamma_{20} + \gamma_{21}(frl2019) + u_{2d}$$
$$\beta_{3d} = \gamma_{30} + \gamma_{31}(frl2019) + u_{3d}$$
$$\beta_{4d} = \gamma_{40} + \gamma_{41}(frl2019) + u_{4d}$$
$$\beta_{5d} = \gamma_{50} + \gamma_{51}(frl2019) + u_{5d}$$

Where $frl2019$ is the proportion of students receiving free and/or reduced-price lunch in the district in 2019 (from CCD); $\varepsilon_{dy} \sim N\left(0, var\left(\hat{Y}_{dy}^{x}\right)\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$. From this model, we get the standard error of the EB annual estimates in each year. We also use the 2019, 2022, and 2023 annual estimates and their standard errors in cases where Models 1-D and 2-D do not produce reliable estimates due to missing data in 2016-2018.

At the end of this step, we have annual EB estimates, $\widehat{Y^*}^x_{dy}$, and their standard errors, $se(\widehat{Y^*}^x_{dy})$, in 2016, 2017, 2018, 2019, 2022, and 2023 for all students in districts with sufficient data.

State-Subgroup Estimates

The state-subgroup EB estimation follows that of the district-all estimation. Estimation is performed separately by subgroup using the state-subgroup OL annual estimates as input.

District-Subgroup Estimates

We take a different approach for the district-subgroup estimates to improve the shrinkage. Rather than model the average subgroup estimates, as we do above for all students, we instead model the differences between the subgroup and the overall (all-student) estimates. We first construct estimates of the differences between the subgroup estimate and the all-student estimates and their standard errors.

$$\widehat{D}^x_{dsy} = \hat{Y}^x_{dsy} - Y^x_{d(all)y}$$

$$se\left(\widehat{D}^x_{dsy}\right) = \sqrt{(1 - 2\bar{p}_s)\left(se\left(\hat{Y}^x_{dsy}\right)^2\right) + se\left(Y^x_{d(all)y}\right)^2}$$

Where $\bar{p}_s$ is the average proportion of students in subgroup, s, across our sample period (2016 to 2019).

We use these difference estimates as the outcomes in the subsequent models. Again, we'll take a 3-model approach. The models differ slightly for the racial subgroups and the ECD/non-ECD subgroups.

Race Estimates

All models are estimated separately by state-subject-subgroup. These subscripts have been omitted for parsimony.

<u>Model 1-R:</u> This model is a corollary to Model 1-D, which provides the EB fitted trend in the differences between the district-subgroup and the district-all group, as well as estimates of the EB difference in 2022 and 2023.[11]

$$\begin{aligned}\widehat{D}^x_{dy} = \beta_{0d}(1-\bar{p}) + \beta_{1d}(year - 2019)(I(year \leq 2019))\,(1-\bar{p}) + \beta_{2d}\big(I(year \geq 2022)(1-\bar{p})\big) \\ + \beta_{3d}(year = 2023)(1-\bar{p}) + \varepsilon_{dy}\end{aligned}$$

$$\beta_{0d} = \gamma_{00} + \gamma_{01}(\bar{p}) + u_{0d}$$

$$\beta_{1d} = \gamma_{10} + \gamma_{01}(\bar{p}) + u_{1d}$$

$$\beta_{2d} = \gamma_{20} + \gamma_{01}(\bar{p}) + u_{2d}$$

$$\beta_{3d} = \gamma_{30} + \gamma_{01}(\bar{p}) + u_{3d}$$

---

[11] In states where no 2023 data are available, we remove the β_{3d} term from the model.

Where $\bar{p}$ is the average proportion of the subgroup in the district; $\varepsilon_{dy} \sim N\left(0, var\left(\widehat{D}_{dy}^{x}\right)\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$.

From this model, we take the EB differences in 2022 and 2023 and add them to the EB district-all estimates in the corresponding year (from the process above).

Model 2-R: This model is a corollary to model 2-D, which enables us to get the 2016-2019 annual EB residuals around the fitted trend for the differences.

$$\widehat{D}_{dy}^{x} = \alpha_{dy}(1 - \bar{p}) + \varepsilon_{dy}$$
$$\alpha_{dy} = \beta_{0d} + \beta_{1d}(YRC) + r_{dy}$$
$$\beta_{0d} = \gamma_{00} + \gamma_{01}(\bar{p}) + u_{0d}$$
$$\beta_{1d} = \gamma_{10} + \gamma_{11}(\bar{p}) + u_{1d}$$

Where $\varepsilon_{dy} \sim N\left(0, var\left(\widehat{D}_{dy}^{x}\right)\right)$; $r_{dy} \sim N(0, \sigma^2)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$.

We add the EB residuals from this model to the fitted values from Model 1-R. This provides an EB estimate of the difference between the subgroup and all estimates in that year. We then add those EB differences to the all estimates in the same year (from the process above).

Model 3-R: This model is a corollary to Model 3-D, which we use to get the standard errors for the EB annual difference estimates.

$$\widehat{D}_{dy}^{x} = \beta_{0d}(y16)(1 - \bar{p}) + \beta_{1d}(y17)(1 - \bar{p}) + \beta_{2d}(y18)(1 - \bar{p}) + \beta_{3d}(y19)(1 - \bar{p}) + \beta_{4d}(y22)(1 - \bar{p}) + \beta_{5d}(y23)(1 - \bar{p}) + \varepsilon_{dy}$$
$$\beta_{0d} = \gamma_{00} + \gamma_{01}(\bar{p}) + u_{0d}$$
$$\beta_{1d} = \gamma_{10} + \gamma_{11}(\bar{p}) + u_{0d} + u_{1d}$$
$$\beta_{2d} = \gamma_{20} + \gamma_{21}(\bar{p}) + u_{0d} + u_{2d}$$
$$\beta_{3d} = \gamma_{30} + \gamma_{31}(\bar{p}) + u_{0d} + u_{3d}$$
$$\beta_{4d} = \gamma_{40} + \gamma_{41}(\bar{p}) + u_{0d} + u_{4d}$$
$$\beta_{5d} = \gamma_{50} + \gamma_{51}(\bar{p}) + u_{0d} + u_{5d}$$

Where $\varepsilon_{dy} \sim N\left(0, var\left(\widehat{D}_{dy}^{x}\right)\right)$; and $\mathbf{U} \sim \mathrm{MVN}(0, \tau^2)$.

To get the standard errors on the annual EBs, we calculate:

$$se\left(\widehat{Y^*}_{dsy}^{x}\right) = \sqrt{se\left(\widehat{D^*}_{dsy}^{x}\right)^2 + se\left(\widehat{Y^*}_{d(all)y}^{x}\right)^2}$$

ECD Estimates

We take the same approach described above for the racial subgroups, but because ECD + non-ECD add to the total students in the sample, we estimate all models using data for both ECD and non-ECD students simultaneously.

Step 8: Constructing Change Estimates

At the end of pooling, we have OL and EB estimates for all years (2016-2023). We construct OL and EB estimates of the 2019-2022, 2022-2023, and 2019-2023 changes from the OL and EB annuals, as follows (formulas show OL state estimates, same applies to EB and district estimates):

$$C(y1)(y2)_{fs}^{x} = \hat{Y}_{fs(y2)}^{x} - \hat{Y}_{fs(y1)}^{x}$$

$$se\left[C(y1)(y2)_{fs}^{x}\right] = \sqrt{se\left(\hat{Y}_{fs(y1)}^{x}\right)^{2} + se\left(\hat{Y}_{fs(y2)}^{x}\right)^{2}}$$

Where $y1$ is equal to 2019 or 2022; and $y2$ is equal to 2022 or 2023.

Step 9: Suppressing and Flagging Data for Release

Our goal is to ensure that the data we release is useful for various education stakeholders. We take caution to not report data that is unreliable and to flag estimates that require additional information for interpretation.

Data suppression. We suppress estimates for three reasons:

(1) The estimates do not reflect at least 20 unique students in at least 2 grades a given school year.

(2) The estimates are too imprecise to be useful. For all subgroups except ECD and non-ECD, the precision threshold we use is 0.33 grade levels. For the ECD and non-ECD subgroups, we keep both groups' estimates when one or both have a change estimate with an OL standard error less than 0.33 grade levels.

(3) The district does not have a corresponding geographic boundary in the 2019 school district shape files. These tend to be specialized administrative districts, like charter school and virtual school districts, and tend to not serve stable populations over time.

Data flags. We flag districts in the data where there were large changes in the overall enrollment and/or in the racial composition using grade 3-8 CCD enrollment data by district-subgroup from fall 2019 and 2022.[12] We identified districts as having large changes in overall enrollment if the proportional change in grade 3-8 CCD enrollment was greater than .20 (20%), calculated as:

$$\frac{|g38enroll_{2022} - g38enroll_{2019}|}{\min(g38enroll_{2022}\,, g38enroll_{2019})} > .20$$

We then identified districts as having large changes in racial composition if the percentage of students of any given racial group changed by more than 5 percentage points, calculated as:

[12] 2023 CCD was unavailable at the time of the release; as such this data flag could not be extended to 2023.

$$\max(|\%asn_{2022} - \%asn_{2019}|, |\%blk_{2022} - \%blk_{2019}|, |\%hsp_{2022} - \%hsp_{2019}|, |\%hpi_{2022} - \%hpi_{2019}|, |\%mtr_{2022} - \%mtr_{2019}|, |\%nam_{2022} - \%nam_{2019}|, |\%wht_{2022} - \%wht_{2019}|) > .05$$

The final flag shown on the website and reported in the data are a combination of these two, such that we flag any places with large changes in the overall enrollment and/or in the racial composition per the above measures. We opted to flag these places to indicate that changes in achievement should be interpreted with caution given the shift in student population.

Data Quality and Validation Checks

Inferences regarding changes in the estimated test scores from 2019 to 2023 hinge on the comparability of the data over time. There are a few potential threats to these inferences:

Data discrepancies. Because state reported proficiency data are the source data for ED*Facts*, the data reported in the two sources should be equivalent (save for differences in data suppression). However, as part of the ED*Facts* data collection process, state proficiency data are vetted and cleaned. As such, it is possible that the 2022 or 2023 state data are of lower or different quality than the 2019 ED*Facts* data. Additionally, there may be differences in how the data are reported, for example, whether alternate assessments are reported in the count data for each district.

To better understand any potential discrepancies in the data, we cleaned the 2019 state reported data using the same rules (described above) as the 2022/2023 state reported data. We then compared the count data and estimates produced from both sources. Differences in estimated means are associated with differences in the underlying count data at the state and/or district levels (i.e., the two sources reported different proportions of students scoring at the same proficiency level). On average, these differences in the estimated means were small (.07 grade levels). Only in a small percentage of cases (6.6%) did estimates differ by more than .2 grade levels. More than three-quarters of cases with differences larger than .2 grade levels are in district-subject-grades with fewer than 100 test-takers; only 3% are in cases with more than 500. Thus, while there are differences in the data, they would impact the estimates we report in limited ways for relatively few test-takers.

Notably, as a result of this analysis, we determined that the data reported in ED*Facts* for Arkansas RLA is not comparable to the data reported by the state in either Reading or English. As such RLA data for Arkansas was removed from website and data files.

NAEP Linking. We replicated a subset of the analyses from Reardon, Kalogrides, and Ho (2021) to demonstrate that the NAEP linking continues to perform well in 2022. However, we do not have a NAEP linkage in 2023 and rely on the linkage from 2022. This requires that states' tests and cutscores did not change between 2022 and 2023. We researched each states' tests and identified cases where it was apparent that there was a change in the state test and/or proficiency cutscores between 2022 and 2023 and excluded these cases.

Comparison to percent proficient data. As a face validity check of the data, for each state-subject we correlated our estimated changes in means (pooled over grades) with changes in the probit transformed reported percent proficient (pooled over grades). Correlations suggest that our estimates tell a story that is largely consistent with that from the reported percent proficient.

Version Changes

Multiple improvements have been made to the data cleaning and estimation process since the SEDA2022 release. The key changes are detailed here:

- Integration of the 2023 state assessment data.
- Switch to using NAEP Expanded Population Estimates for linking.
- Creation of unadjusted and adjusted standard errors for OL estimates.
- New pooling models and shrinkage approach.

Because of these changes, estimates of the 19-22 changes from SEDA2022 may not be identical to those in SEDA2023. We believe that our current estimates improve upon those in SEDA2022 and encourage users to use the latest version of the data.

Covariates

The unit-year covariates provided in SEDA2023 come from the Common Core of Data (CCD). We first impute missing school-level data (using the process described in the SEDA 4.1 documentation) and aggregate the school data to administrative districts.

Definitions

Administrative school district: Administrative school districts operate sets of public and charter schools. The schools operated by each school district are identified using the National Center for Education Statistics (NCES) school and district identifiers. Most commonly, administrative school districts operate local public schools within a given physical boundary; these are what we refer to as "traditional public school districts." There are also specialized administrative districts that do not have a physical boundary, like charter school and virtual school districts.

Subgroup: Subgroups are defined per state accountability reporting requirements; our data include the following subgroups: all, Black, Hispanic, White, poor (economically disadvantaged), and non-poor (not disadvantaged).

Frequently Asked Questions

Why aren't results available for my state or district?

There are several reasons why we may not show data for a particular district.

1. We may not be able to construct an estimate for the district in 2019-2023 if:
    a. Sufficient data for estimation were not reported by the state or district in 2019-2023.
    b. The district changed its leaid between 2019 and 2023. This may occur as a result of a district merger, split, or takeover that occurred during this four-year span.
    c. Fewer than 94% of students in the state or 95% of students in the district participated in testing in the subject in 2019.
    d. More than 40% of students in the district took alternative assessments rather than the regular tests in 2019.
2. We may suppress the data because:
    a. The district is too small and/or has too few grades of data available to allow for the construction of reliable estimates.
    b. The district does not have a geographic boundary; such districts include charter districts and/or specialized local education agencies.
    c. For the ECD/NEC subgroups, if a state or district has only an ECD or NEC estimate, we suppress the other group.

Why did the results for my district change between the SEDA2022 and SEDA2023 releases?

A number of improvements were made in our data inclusions/exclusion rules, as well as in our estimation methods that enabled us to get better estimates of the average achievement in 2019 and 2022, and the change in achievement been those two years. See the notes on version changes above for the key improvements.

Why are you releasing unadjusted and adjusted standard errors and how should they be used?

The unadjusted standard error (which does not account for NAEP linking uncertainty) is appropriate to use when making within-state comparisons (between districts in the same state or between subgroups within the same district). Within a state, the linking error is common to all districts such that within-state comparisons do not rely on the NAEP linkages.

The adjusted standard error (which accounts for NAEP linking uncertainty) is appropriate to use when making comparisons between districts in different states. These comparisons rely on the NAEP linking to be accurate.

How have you accounted for population changes from 2019 to 2022 in the estimates?

We have not made any adjustment to the estimates to account for changes in the population between 2019 and 2022. On the website and in the downloadable files, we provide a flag that indicates if there was a large change in the total enrollment and/or racial composition of the district during the time period. The calculation of this flag is described above under the section *Step 8: Suppressing and Flagging Data for Release*.

The data are not yet available to provide a similar flag for the 2022 to 2023 change.

References

Raudenbush, S.W., & Congdon, R.T. (2021). *HLM 8: Hierarchical linear and nonlinear modeling*. Chapel Hill, NC: Scientific Software International, Inc.

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. (2021). 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, 46*(2), 138–167.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

*State Assessments in Reading/Language Arts and Mathematics- School Year 2018-19 ED*Facts *Data Documentation*, U.S. Department of Education, Washington, DC: EDFacts. Retrieved 04/24/2023 from http://www.ed.gov/edfacts.

Tables & Figures

Table 1. *Overview of State Data Inclusion in SEDA2023*

State	In Data		Reasons for data exclusion by state, subject, and year					
			Math			RLA		
	19-22 Estimates	22-23 Estimates	2019	2022	2023	2019	2022	2023
AL	Y	Y (math only)						Test changed
AK	N	N		Two proficiency categories reported	Two proficiency categories reported		Two proficiency categories reported	Two proficiency categories reported
AZ	Y	N			Change in format of reported data; quality unclear			Change in format of reported data; quality unclear
AR	Y (math only)	Y (math only)					Multiple RLA tests	Multiple RLA tests
CA	Y	Y						
CO	N	N		Participation < 95%	No 2022 data		Participation < 95%	No 2022 data
CT	Y	Y						
DE	N	N		Two proficiency categories reported	Two proficiency categories reported		Two proficiency categories reported	Two proficiency categories reported
DC	Y (RLA only)	N	Participation < 95%	No 2019 data	No 2019 data			Change in format of reported data; quality unclear
FL	Y	N			Test changed			Test changed
GA	Y	Y						
HI	N	N		Two proficiency categories reported	Two proficiency categories reported		Two proficiency categories reported	Two proficiency categories reported
ID	Y	N			Test changed			Test changed
IL	Y	Y						
IN	Y	Y						
IA	N	N		Two proficiency categories reported	Two proficiency categories reported		Two proficiency categories reported	Two proficiency categories reported
KS	Y	Y						
KY	Y	Y						

LA	Y	Y						
ME	N	N		No data	No data		No data	No data
MD	Y	N			Test changed			Test changed
MA	Y	Y						
MI	Y	Y						
MN	Y	N			Test changed			Test changed
MS	Y	Y						
MO	Y	N			No data			No data
MT	N	N		No data	No data		No data	No data
NE	Y	N			Test changed			Test changed
NV	Y	Y						
NH	Y	Y						
NJ	Y	Y						
NM	N	N		No data	No data		No data	No data
NY	N	N	Participation < 95%	No 2019 data	No 2019 data	Participation < 95%	No 2019 data	No 2019 data
NC	Y	Y						
ND	Y	N			No data			No data
OH	Y	Y						
OK	Y	Y						
OR	Y	Y						
PA	Y	Y						
RI	Y	Y						
SC	Y	N			Test changed			Test changed
SD	Y	Y						
TN	Y	Y						
TX	Y	N			Test changed			Test changed
UT	Y	Y						
VT	N	N		No data	No data		No data	No data
VA	Y	Y						
WA	Y	Y						
WV	Y (math only)	Y (math only)				Data suppressed by EDFacts	No 2019 data	No 2019 data
WI	Y	Y						
WY	Y	Y (RLA only)			Test changed			

Table 2. *Source data overview*

State	Subgroups reported	Estimated counts	Partial suppression
AL	All, Black, ECD, Hispanic, White students	Y	N
AR	All students	N	N
AZ	All, Black, Hispanic, White students	N	N
CA	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
CT	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
FL	All students	N	N
GA	All students	N	N
ID	All, Black, ECD, Hispanic, Not-ECD, White students	Y	Y
IL	All, Black, ECD, Hispanic, Not-ECD, White students	Y	N
IN	All students	N	N
KS	All, Black, ECD, Hispanic, White students	Y	N
KY	All, Black, ECD, Hispanic, Not-ECD, White students	Y	N
LA	All, Black, ECD, Hispanic, Not-ECD, White students	Y	N
MA	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
MD	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
MI	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
MN	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
MO	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
MS	All students	N	N
NC	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
ND	All, ECD, Not-ECD	N	N
NE	All, Black, ECD, Hispanic, White students	Y	N
NH	All, Hispanic, White	Y	Y
NJ	All, Black, ECD, Hispanic, Not-ECD, White students	In some cells	N
NV	All students	N	Y
OH	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
OK	All students	N	Y
OR	All, Black, Hispanic, White students	N	N
PA	All students	N	N
RI	All students	N	N
SC	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
SD	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
TN	All, Black, ECD, Hispanic, Not-ECD, White students	N	Y
TX	All, Black, ECD, Hispanic, Not-ECD, White students	N	N

UT	All students	Y	Y
VA	All students	Y	N
WA	All, Black, ECD, Hispanic, Not-ECD, White students	In some cells	N
WI	All, Black, ECD, Hispanic, Not-ECD, White students	N	N
WV	All, Black, White students	Y	N
WY	All, ECD, Hispanic, Not-ECD, White students	Y	N

Table 3. *Data cleaning, state-subject-grade removals*

State	Subject	Grade	Reason removed pre-estimation		
			2019	2022	2023
AZ	Math	8	Primary Testing Rate < 95%		N/A
CO	All Subjects	All Grades	2022 Participation Rate < 94%	2022 Participation Rate < 94%	2022 Participation Rate < 94%
DC	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
FL	Math	7	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
FL	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
MD	Math	7	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
MD	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
MO	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
NJ	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%
NY	All Subjects	All Grades	2019 Participation Rate < 94%	2019 Participation Rate < 94%	2019 Participation Rate < 94%
OH	Math	7	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%
OH	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%
TN	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%
TX	Math	7	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
TX	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
TX	RLA	3	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
TX	RLA	4	Primary Testing Rate < 95%	Primary Testing Rate < 95%	N/A
VA	Math	7	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%
VA	Math	8	Primary Testing Rate < 95%	Primary Testing Rate < 95%	Primary Testing Rate < 95%