education analytics

TECHNICAL REPORT ON THE WISCONSIN VALUE-ADDED MODEL:

ACADEMIC YEAR 2022-23

// November 2023

110 E Main Street, Ste. 1000
Madison, WI 53703

608.466.4966

edanalytics.org

CONTENTS

INTRODUCTION

This report describes the value-added growth model and score calculation methodology used by Education Analytics to estimate the contribution of Wisconsin public schools to growth in student knowledge. The 2022-23 model uses assessment data from the Wisconsin Forward Exam, ACT Aspire (in use during 2020-21 and 2021-2022), PreACT Secure (replaced ACT Aspire in 2022-23), and the ACT.

The report is divided into three sections. The first section describes the data sets used to produce the value-added scores. The second section describes the model used to calculate value-added scores for schools in Wisconsin. The third section presents some properties of the value-added results.

Conceptually, value-added analysis is the use of statistical techniques to isolate the component of student knowledge gains that is attributable to schools. Other factors that may influence student knowledge gains include prior knowledge and student characteristics. In practice, value-added growth models focus on the improvement students make on assessments from one year to the next. Value-added growth models often consider (control for) differences in student characteristics such as economic disadvantage and disability, to help isolate the impact of schooling.

The model used in Wisconsin controls for the available set of student characteristics to identify the extent to which schools contribute to the improvement of student achievement outcomes. Once school-level value-added scores are calculated, these are averaged to obtain district scores. To calculate the final value-added scores, up to three years of results are combined: 2020-21, 2021-22, and 2022-23.

ANALYSIS DATA SETS

Before estimation can take place, a substantial amount of work is required to assemble the data sets used to produce the value-added estimates. A separate data set is prepared for each grade and subject. In total, 16 data sets are produced, covering grades 4 through 11 for English language arts (ELA) and math.

Each analysis data set includes students who have: (1) a test score in 2022-23 (the posttest) in the grade and subject being considered, (2) test score in 2021-22 in both ELA and math (the pretests) and (3) full academic year (FAY) status in their school or district in the 2022-23 school year.

The model also includes students in private schools participating in voucher school programs (referred to as Private School Choice Programs in Wisconsin). All voucher recipients are included in modeling. In addition, some private schools receiving voucher funds opt to receive a value-added score that includes all attending students, including those students not receiving vouchers.

Student-level variables

POSTTEST AND PRETEST SCORES

The value-added model uses test scores from the 2020-21, 2021-22, and 2022-23 administrations of the Forward Exam, ACT Aspire, PreACT Secure, and the ACT. The Forward Exam was administered to students in grades 3 through 8; the ACT Aspire to students in grades 9 and 10 in 2020-21 and 2021-22; the PreACT Secure to students in grades 9 and 10 in 2022-23; and the ACT to students in grade 11. The model produces school-level estimates for grades 4 through 11 in ELA and math based on performance on the 2022-23 assessment. The ELA model uses the 2022-23 ELA score as the posttest. The math model uses the 2022-23 math score as the posttest. Both models include pretests in both ELA and math, both from the year before the posttest in 2021-22 and, when available, from two years before the posttest in 2020-21.

GENDER, RACE/ETHNICITY, ECONOMIC DISADVANTAGE, AND MIGRANCY

Gender, race/ethnicity, economic disadvantage, and migrancy information are drawn from the Wisconsin Information System for Education data (WISEdata) elements submitted by schools and districts. More specifically, the values for these variables are drawn from the Spring Demographic Snapshot of WISEdata captured on May 23, 2023.[1] In the data set used for value-added modeling, students are assigned the gender, race/ethnicity, economically disadvantaged status, and migrant status reported in the post-test year. Gender categories are male, female, and non-binary. Race categories are American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Black/African American, Hispanic/Latino, White, and multi-racial. The model uses an indicator for economic disadvantage and an indicator for whether students are migrant.

[1] WISEdata is a dynamic data delivery system. Snapshots capture a static version of the data as it was delivered to Wisconsin DPI on a given date. The Spring Demographic Snapshot taken near the end of the school year was for the purpose of supplying demographic characteristics to associate with student assessment results.

ENGLISH LANGUAGE PROFICIENCY CLASSIFICATION

Eight indicators for English-language proficiency (ELP) are included in the dataset. Students with ELP classifications of 1 through 5 are English-language learners in ascending levels of proficiency. Students with an ELP classification of 6 are those who were formerly classified as English learners. Students with an ELP classification of 7 are those who were never English Learners. ELP classification is drawn from the WISEdata Snapshot mentioned above. An eighth indicator is created for students for whom ELP status is unknown.

DISABILITY

The dataset includes five indicators for students with disabilities according to their primary disability code. There are separate indicators for emotional/behavioral disability (EBD), learning or intellectual disability (LD/ID), autism (A), and speech/language disability (SL). All other disability codes are grouped into a single indicator for other disabilities. Disability status is based on a student having an active individualized education program (IEP) or individualized service plan (ISP) between December 1 and June 30 of the 2022-23 school year.

School enrollment

The dataset includes indicators of full academic year (FAY) status at the school and district level for 2022-23. For the purpose of the Wisconsin accountability systems and therefore for the Wisconsin value-added model, FAY is defined as being enrolled from the third Friday of September through completion of statewide testing. Students who are FAY at the school level are included in school-level value-added scores. Some students are FAY in a district but not in a single school because of mobility within the district. These students are included in the district value-added scores but not in the school value-added scores.

Students attending private school

The data set also includes test scores for students participating in one of the Private School Choice programs in Wisconsin. These students receive a voucher to attend private school. All participating private schools receive a value-added score based only on students in Choice programs (i.e., those receiving vouchers). In addition, participating private schools are given the option to receive a second report card and value-added score that includes all students in the school. Such schools are denoted as “opt-in” schools because they opted to receive the second non-compulsory score. Value-added scores for "opt-in" schools that include students not in a Choice program (i.e., students attending private schools but not using vouchers) are computed by re-estimating the value-added model using a data set that includes both students receiving vouchers as well as those not receiving vouchers.

Descriptive statistics of analysis samples

Tables 1 and 2 describe the sample used for the 2022-23 school year. Note that the sample includes students from public schools and private schools participating in one of the Choice programs in Wisconsin. The private school students include students attending schools that opted in to receive a score for all their students regardless of whether an individual student is participating in Choice.

Table 1. Math Sample

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Number of Students	54,355	54,927	54,759	55,506	57,035	55,657	55,616	52,289
Number of Public School Students	51,059	51,702	51,425	52,209	53,751	52,934	52,938	49,790
Number of Students in Choice Programs	2,587	2,563	2,515	2,470	2,450	2,093	2,042	1,799
Number of Private School Students not in Choice Programs	413	405	402	440	424	235	352	405
Total Number of Private School Students	3,000	2,968	2,917	2,910	2,874	2,328	2,394	2,204
Number of Public Schools	1,085	1,033	702	669	667	538	550	558
Number of Private Schools	166	161	158	156	150	73	72	71
Number of Public School District Codes	430	429	431	432	429	391	393	387
Posttest Mean	578.819	603.310	612.794	623.232	641.157	17.178	18.580	19.390
Posttest Standard Deviation	54.503	51.861	56.444	59.704	55.623	4.204	5.101	5.422
Math Pretest Mean	554.107	575.276	601.384	607.102	620.128	639.178	426.136	428.232
Math Pretest Standard Deviation	55.119	56.08	51.645	58.268	60.884	58.037	9.676	10.173
ELA Pretest Mean	550.917	579.487	595.364	603.957	620.677	626.564	425.698	427.176
ELA Pretest Standard Deviation	46.822	51.635	51.354	50.146	55.544	59.209	7.376	7.577
Proportion in ELP Level 1	0.011	0.006	0.006	0.006	0.006	0.004	0.003	0.002

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Proportion in ELP Level 2	0.020	0.009	0.007	0.012	0.011	0.009	0.007	0.006
Proportion in ELP Level 3	0.040	0.034	0.026	0.035	0.031	0.027	0.024	0.020
Proportion in ELP Level 4	0.015	0.030	0.026	0.012	0.010	0.013	0.012	0.011
Proportion in ELP Level 5	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion in ELP Level 6 (former English learners)	0.012	0.028	0.042	0.045	0.052	0.053	0.054	0.061
Proportion in ELP Level 7 (not English learners)	0.892	0.883	0.882	0.880	0.879	0.885	0.893	0.893
Proportion ELP Unknown	0.010	0.010	0.011	0.010	0.009	0.008	0.006	0.007
Proportion Female	0.493	0.488	0.489	0.487	0.483	0.485	0.488	0.490
Proportion Male	0.507	0.512	0.511	0.513	0.516	0.514	0.511	0.509
Proportion Non-Binary	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion Asian	0.046	0.045	0.043	0.041	0.043	0.041	0.040	0.042
Proportion African American	0.090	0.088	0.088	0.088	0.089	0.079	0.061	0.054
Proportion Hispanic	0.139	0.142	0.142	0.142	0.142	0.139	0.128	0.125
Proportion Native American	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.008
Proportion Native Hawaiian or Other Pacific Islander	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion Two or More Races	0.055	0.054	0.049	0.047	0.046	0.044	0.040	0.037
Proportion Special Education: Emotional Behavioral	0.010	0.011	0.013	0.013	0.014	0.013	0.011	0.009
Proportion Special Education: Learning/ Intellectual	0.045	0.048	0.049	0.049	0.048	0.043	0.039	0.036
Proportion Special Education Autism	0.015	0.015	0.014	0.014	0.014	0.012	0.010	0.011
Proportion Special Education: Speech/ Language	0.035	0.024	0.014	0.009	0.006	0.003	0.002	0.002

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Proportion Special Education: Other	0.043	0.041	0.041	0.041	0.042	0.036	0.033	0.031
Proportion Economically Disadvantaged	0.436	0.438	0.427	0.422	0.416	0.390	0.344	0.322
Proportion Migrant	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table 2. English Language Arts (ELA) Sample

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Number of Students	54,381	54,950	54,780	55,542	57,061	55,531	55,523	51,651
Number of Public School Students	51,081	51,714	51,439	52,243	53,783	52,817	52,848	49,178
Number of Students in Choice Programs	2,591	2,572	2,526	2,470	2,445	2,089	2,041	1,789
Number of Private School Students not in Choice Programs	413	406	402	440	424	235	352	405
Total Number of Private School Students	3,004	2,978	2,928	2,910	2,869	2,324	2,393	2,194
Number of Public Schools	1,085	1,033	702	669	667	538	550	558
Number of Private Schools	166	161	159	156	150	73	71	70
Number of Public School District Codes	430	429	431	432	429	391	393	387
Posttest Mean	585.410	594.355	606.010	624.485	631.040	16.213	17.698	18.829
Posttest Standard Deviation	52.491	50.880	52.027	54.268	62.761	5.228	5.866	5.616
ELA Pretest Mean	550.903	579.461	595.351	603.927	620.655	626.727	425.711	427.298
ELA Pretest Standard Deviation	46.829	51.653	51.359	50.169	55.563	59.117	7.371	7.505
Math Pretest Mean	554.090	575.246	601.368	607.047	620.096	639.319	426.151	428.357
Math Pretest Standard Deviation	55.123	56.102	51.664	58.315	60.912	57.960	9.673	10.134
Proportion in ELP Level 1	0.011	0.006	0.006	0.006	0.006	0.004	0.003	0.002
Proportion in ELP Level 2	0.020	0.009	0.007	0.012	0.011	0.009	0.007	0.005
Proportion in ELP Level 3	0.040	0.034	0.026	0.035	0.031	0.027	0.024	0.020

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Proportion in ELP Level 4	0.015	0.030	0.026	0.012	0.010	0.013	0.012	0.011
Proportion in ELP Level 5	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion in ELP Level 6 (former English learners)	0.012	0.028	0.042	0.045	0.052	0.053	0.054	0.061
Proportion in ELP Level 7 (not English learners)	0.892	0.883	0.882	0.880	0.879	0.885	0.893	0.894
Proportion ELP Unknown	0.010	0.010	0.011	0.010	0.009	0.008	0.006	0.007
Proportion Female	0.493	0.488	0.489	0.487	0.483	0.485	0.488	0.493
Proportion Male	0.507	0.512	0.511	0.513	0.516	0.514	0.511	0.506
Proportion Non-Binary	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion Asian	0.046	0.045	0.043	0.041	0.043	0.041	0.041	0.042
Proportion African American	0.090	0.088	0.088	0.088	0.089	0.079	0.061	0.053
Proportion Hispanic	0.139	0.143	0.142	0.142	0.142	0.138	0.127	0.125
Proportion Native American	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.008
Proportion Native Hawaiian or Other Pacific Islander	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Proportion Two or More Races	0.055	0.054	0.049	0.047	0.046	0.044	0.040	0.037
Proportion Special Education: Emotional Behavioral	0.010	0.011	0.013	0.013	0.014	0.012	0.011	0.008
Proportion Special Education: Learning/ Intellectual	0.045	0.049	0.049	0.049	0.048	0.043	0.039	0.035
Proportion Special Education: Autism	0.015	0.015	0.014	0.014	0.014	0.012	0.010	0.010
Proportion Special Education: Speech/Language	0.035	0.024	0.014	0.009	0.006	0.003	0.002	0.002
Proportion Special Education: Other	0.043	0.041	0.041	0.041	0.042	0.036	0.033	0.030
Proportion Economically Disadvantaged	0.436	0.438	0.427	0.422	0.416	0.39	0.344	0.319

Variable	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Proportion Migrant	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Transformation of assessment variables

All test scores are transformed to a rank-based z-statistic scale with means equal to zero and standard deviations equal to one in each grade and subject. Thus, in the value-added model, all test scores were measured relative to the state means, and in units of the statewide standard deviations of test scores in given grades and subjects. The rank-based z-statistic transformation ranks scores and then assigns to them a z-statistic score (or z-score) based on the value associated with that rank in the normal distribution.

The transformation to the rank-based z-statistic scale is made for several reasons. Normalizing the test scores makes it easier to interpret the coefficients in the value-added model, given that they are now measured in units of the standard deviation across students in scores on the posttest assessment. In addition, using a rank-based normalization reshapes the distribution of test scores to be the same (standard normal) for all grades and subjects. This is likely to reduce the extent of nonlinearity in the relationships between the posttest and pretests, particularly when the posttest and pretests are different assessments.

This transformation is extended to the conditional standard errors of measurement (CSEMs) of the pretests, which are used to account for pretest measurement error when estimating the value-added regression (see the "Value-added regression" section below). Transforming the CSEMs involves multiple steps, implemented separately for each scale score value. First, we repeatedly simulate measurement error around the scale score, creating a set of repeated, simulated scale scores with measurement error. These simulated scores are drawn from a normal distribution with a mean at the value of the original scale score and a standard deviation at the CSEM associated with the original scale score. Next, we transform the simulated scale scores using the same transformation as that which was used to transform the original scale scores. Last, we compute the standard deviation of the transformed simulated scores. This computes the CSEM on the rank-based z-statistic scale. There is no need to extend this transformation to the CSEM of the posttest. This is because posttest measurement error simply becomes incorporated into the error term of the regression and does not need to be accounted for when estimating the value-added regression.

Across years, grades, and subjects, the reliability of the pretest scores on the rank-based z-statistic scale ranges from 0.88 to 0.93 on the Forward Exam and from 0.84 to 0.93 on the ACT Aspire. The reliability of an assessment is the extent to which differences in test scores reflect true differences in student knowledge rather than just random variation. These reliability values suggest that the vast majority of the variance of these test scores reflect actual differences in student knowledge of the content area.

VALUE-ADDED MODEL

For the Wisconsin school-level model, 2022-23 value-added is estimated in mathematics and English language arts (ELA) in grades 4 through 11 using the Forward Exam (4-8), the ACT Aspire in grades 9 and 10 in 2020-21 and 2021-22, the PreACT Secure in grades 9 and 10 in 2022-23, and the ACT in grade 11. School single-year value-added scores reflect student growth from Spring 2022 to Spring 2023. Schools' value-added scores are averaged to obtain a district-level value-added score, using the number of students attributed to each school as weights.[2] The single-year value-added scores for 2022-23 are averaged with the two most recent prior value-added scores (the single-year scores from 2021-22 and the skip-year scores from 2020-21) to produce a multi-year average score that smooths year-to-year variability in value-added scores. The skip-year approach employed in 2020-21 is described in Appendix A.

The model, in brief

The value-added growth model is defined by six equations: a "best linear predictor" value-added model defined in terms of true student posttest and pretest achievement (i.e., student achievement in the absence of test measurement error) and five measurement error models for observed posttest and pretest scores:

Student achievement: $y_{3i} = \zeta + \lambda_2 y_{2i} + \lambda_2^{alt} y_{2i}^{alt} + \lambda_1 y_{1i} + \lambda_1^{alt} y_{1i}^{alt} + \beta' X_i + \alpha' S_i + e_i$ (1)

Posttest measurement error: $Y_{3i} = y_{3i} + v_{3i}$ (2)

Same-subject, once-lagged pretest measurement error: $Y_{2i} = y_{2i} + v_{2i}$ (3)

Other-subject, once-lagged pretest measurement error: $Y_{2i}^{alt} = y_{2i}^{alt} + v_{2i}^{alt}$ (4)

Same-subject, twice-lagged pretest measurement error: $Y_{1i} = y_{1i} + v_{1i}$ (5)

Other-subject, twice-lagged pretest measurement error: $Y_{1i}^{alt} = y_{1i}^{alt} + v_{1i}^{alt}$ (6)

where:

- the subscript i denotes each individual student;
- y_{3i} is true posttest achievement;

[2] Note that students who changed schools within a given district within a year are included in the district's score but not in a school score (see School Enrollment section).

- y_{2i} and y_{2i}^{alt} are true pretest achievement, one year before posttest achievement, in the same subject and in the other subject (math in the ELA model, ELA in the math model), with slope parameters λ_2 and λ_2^{alt};
- y_{1i} and y_{1i}^{alt} are true pretest achievement, two years before posttest achievement, in the same subject and in the other subject (math in the ELA model, ELA in the math model), with slope parameters λ_1 and λ_1^{alt};
- X_i is a vector of characteristics of student i, with slope parameter vector β;
- S_i is a vector of indicators for school;
- α is a vector of school effects;
- e_i is the error in predicting posttest achievement given the explanatory variables included in the model;
- Y_{3i} is observed posttest scores;
- v_{3i} is measurement error in posttest scores;
- Y_{2i} and Y_{2i}^{alt} are observed pretest scores, one year before the posttest, for the same subject and alternate subject, respectively;
- v_{2i} and v_{2i}^{alt} are measurement error in pretest scores, one year before the posttest, for the same subject and alternate subject, respectively;
- Y_{1i} and Y_{1i}^{alt} are observed pretest scores, two years before the posttest, for the same subject and alternate subject, respectively; and
- v_{1i} and v_{1i}^{alt} are measurement error in pretest scores, two years before the posttest, for the same subject and alternate subject, respectively.

Substituting the measurement error equations (2) through (6) into the student achievement equation (1) yields an equation defined in terms of student test scores:

Observed test scores: $Y_{3i} = \zeta + \lambda_2 Y_{2i} + \lambda_2^{alt} Y_{2i}^{alt} + \lambda_1 Y_{1i} + \lambda_1^{alt} Y_{1i}^{alt} + \beta' X_i + \alpha' S_i + \varepsilon_i$ (7)

where the error term ε_i includes both the original error component and the measurement error components:

Error in observed test scores: $\varepsilon_i = e_i + v_{3i} - \lambda_2 v_{2i} - \lambda_2^{alt} v_{2i}^{alt} - \lambda_1 v_{1i} - \lambda_1^{alt} v_{1i}^{alt}$ (8)

Estimating the observed test scores equation (7) without controlling for pretest measurement error yields biased estimates of all parameters, including the value-added effects. This bias stems from the fact that measurement error in pretest scores causes the error term (8), which includes the measurement error components v_{2i}, v_{2i}^{alt}, v_{1i}, and v_{1i}^{alt}, to be correlated with observed pretest scores. The desired parameters, as defined in equation (1), can be estimated consistently if external information is available on the variance of measurement error for the pretests; approaches for consistent estimation in the presence of measurement error are described in detail in Fuller (1987). Information about the variance of test measurement error is obtained from the conditional standard errors of measurement (CSEMs) provided alongside the test scores.

In contrast to measurement error in the pretest score variables, measurement error in the posttest does not cause any distortions in commonly used regression approaches and can safely be overlooked. This is because we do not expect posttest measurement error v_{3i} to be correlated with observed pretest scores or any of the other right-hand-side variables in the regression equation (7). We do not expect any such correlation because there is no reason to think that a student's good or bad luck on the posttest administration should have anything to do with their past test scores, their demographic characteristics, or their school assignment. Given the absence of such a correlation, the presence of posttest measurement error v_{3i} in the regression error term in (8) will not bias coefficient estimates if it is overlooked. In fact, from the perspective of estimation technique, we can think of posttest measurement error v_{3i} as operating no differently from the structural error e_i.

Value-added regression

As mentioned, the value-added growth model is estimated using a least-squares regression approach that corrects for measurement error in the pretest score variables. It estimates the coefficients λ, β, and α by regressing posttest scores on the pretest scores, student characteristic variables, and a full set of school fixed effects. This regression is estimated using an approach that accounts for measurement error in the pretests Y_{2i}, Y_{2i}^{alt}, Y_{1i}, and Y_{1i}^{alt}. Recall from equation (8) above that v_{2i}, v_{2i}^{alt}, v_{1i}, and v_{1i}^{alt}, the measurement error components of the pretests, are part of the error term ε_i. As a result, estimating the regression using ordinary least squares (without controlling for pretest measurement error) will lead to biased estimates. The regression approach employed accounts for measurement error by removing the variance in the pretests that is attributable to measurement error. To illustrate the measurement error corrected regression, re-cast the above value-added regression equation into matrix form:

$$Y_t = Y_{t-\ell}\lambda + W\delta + \varepsilon$$

where Y_t is an N × 1 matrix of post-test scores, $Y_{t-\ell}$ is an N × 4 matrix of same-subject and other-subject pre-test scores Y_{2i}, Y_{2i}^{alt}, Y_{1i}, and Y_{1i}^{alt}; λ is a 4 × 1 matrix made up of λ_2, λ_2^{alt}, λ_1, and λ_1^{alt}; W is an N × K matrix of the X demographic variables and S school indicators, δ is a K × 1 matrix of the β and α coefficients, and ε is an N × 1 matrix of error terms. The biased ordinary-least-squares estimates of the coefficients in λ and δ are equal to:

$$\begin{bmatrix} \hat{\lambda}_{OLS} \\ \hat{\delta}_{OLS} \end{bmatrix} = \begin{bmatrix} Y'_{t-\ell}Y_{t-\ell} & Y'_{t-\ell}W \\ W'Y_{t-\ell} & W'W \end{bmatrix}^{-1} \begin{bmatrix} Y'_{t-\ell}Y_t \\ W'Y_t \end{bmatrix}$$

The measurement-error-corrected estimates of the coefficients in λ and δ are equal to:

$$\begin{bmatrix} \hat{\lambda}_{EIV} \\ \hat{\delta}_{EIV} \end{bmatrix} = \begin{bmatrix} Y'_{t-\ell}Y_{t-\ell} - \left(\frac{N-K-4}{N}\right)\sum_{i=1}^{N} V_{it-\ell} & Y'_{t-\ell}W \\ W'Y_{t-\ell} & W'W \end{bmatrix}^{-1} \begin{bmatrix} Y'_{t-\ell}Y_t \\ W'Y_t \end{bmatrix}$$

where $V_{it-\ell}$ is a 4×4 variance-covariance matrix of the errors of measurement of the variables in $Y_{t-\ell}$ for student i. This model is described in section 2.2 of Fuller (1987).

The variables in the model

In addition to posttest and pretest scores, the student characteristic variables included in the value-added model (the X variables in equation 1) include gender, race/ethnicity, ELP category (indicators reflecting each of ELP levels 1-5, former ELP, not ELP, and ELP unknown), an indicator for economic disadvantage, disability status (indicators for emotional/behavioral, learning/intellectual, autism, speech/language, and all others), and a migrant status indicator. No higher order terms or interactions of terms are used in the model. Refer to the section “Student-level variables” for a more complete description of the categories that make up each student characteristic variable.

Frequency of lowest observed scale scores

In some grades, a disproportionate number of students received Forward Exam math scores at the lowest observable scale score (LOSS). We present the proportion of students with scores at the LOSS in Table 3. The substantive number of students at the LOSS was a primary reason for converting scale scores using the rank-based z-statistic transformation for use in the value-added model. This conversion sets scores at the LOSS (and all other levels) to values corresponding to a normal distribution of student achievement across the state, which keeps scores at the LOSS from entering the model as outliers that may distort the value-added regression.

Table 3. Percentage of Students at Test Floor (Lowest Observable Scale Score, LOSS) for Pre- and Posttests

	Posttest Grade	Test Subject	Percent at Posttest Floor	Percent at Math Pretest Floor	Percent at ELA Pretest Floor
Included in Growth Analysis Data Set	4	ELA	<0.1	1.1	<0.1
		Mathematics	1.3	1.1	<0.1
	5	ELA	<0.1	2.5	<0.1
		Mathematics	1.6	2.5	<0.1
	6	ELA	<0.1	2.3	<0.1
		Mathematics	1.4	2.3	<0.1

	7	ELA	<0.1	2.2	<0.1
		Mathematics	1.4	2.2	<0.1
	8	ELA	0.1	2.2	<0.1
		Mathematics	1.6	2.2	<0.1
	9	ELA	<0.1	2.6	0.1
		Mathematics	0.1	2.6	0.1
	10	ELA	<0.1	<0.1	<0.1
		Mathematics	0.1	<0.1	<0.1
	11	ELA	<0.1	<0.1	<0.1
		Mathematics	<0.1	<0.1	<0.1

Incorporating students with only two years of scores

The estimation approach above produces school value-added results based on the growth of students with test scores in all three years (2020-21-, 2021-22, and 2022-23). To include students with test scores in 2022-23 and 2021-22 but not in 2020-21, we estimate a value-added model that is identical to that described above except that it does not include the pretest variables y_{1i} and y_{1i}^{alt}. We then produce, for each student, a growth residual equal to an estimate of $\alpha' S_i + \varepsilon_i$, using the coefficients from the complete model that includes y_{1i} and y_{1i}^{alt} when the pretest scores Y_{1i} and Y_{1i}^{alt} are available, and using the coefficients from the model that does not include y_{1i} and y_{1i}^{alt} when the pretest scores Y_{1i} and Y_{1i}^{alt} are not available. This growth residual is demeaned by subtracting its mean across students by grade and subject and regressed on a full set of school indicators S_i using ordinary least squares. This produces school value-added results for each school by grade and subject.

Aggregation to multiple-grade value-added

The value-added regression to obtain school value-added results is performed separately for each grade and subject combination. For schools that have results for more than one grade level, these estimates are averaged across grades, using the number of students attributed to the school and grade as weights, to produce multiple-grade value-added results. Before aggregation, value-added results are normalized by subject and grade, so they are on a similar scale (i.e., with a mean of 0 and a standard deviation of 1). This normalization is done by dividing the results by an estimate of the standard deviation of within-grade value-added. This aggregation is made separately at the elementary/middle (grades 4-8) and high school (grades 9-11) levels.

Shrinkage of value-added

At all levels (school, district, student group), the value-added results are 'shrunk' to obtain value-added scores, using an Empirical Bayes multivariate shrinkage technique described in Longford (1999). This procedure brings value-added scores based on smaller sample sizes closer to the state average, so that schools with fewer students are not overrepresented among the highest- and lowest-value-added score cases simply due to randomness. It also reduces year-by-year variation in value-added scores for schools with small student populations.

This multivariate shrinkage approach begins with single-year value-added results for the 2022-23 and 2021-22 school years. Let $\hat{\alpha}_{kt}$ be the value-added result for school *k* in year *t*. We can group the value-added results for a given school *k* into a *T* x 1 column vector $\hat{\alpha}_k$, where *T* is the number of years in which value-added is estimated for school *k*. (In this application, *T* will usually be 2, although it will equal 1 in schools in which value-added is estimated in 2022-23 but not 2021-2022 or vice versa.) Also let α_{kt} be the true value-added (which is unmeasured, and equal to what the value-added result would be in the absence of sampling error) for school *k* in year *t*, which can be grouped by school into a *T* x 1 column vector α_k. Let the variance of α_k be the *T* x *T* matrix $Var[\alpha_k] = \Omega$, which reflects the within-year variance and across-year covariance of true value-added across schools. Also let the variance of $\hat{\alpha}_k$ conditional on α_k be the *T* x *T* matrix $Var[\hat{\alpha}_k|\alpha_k] = \Sigma_{kk}$, which reflects the within-year variance and across-year covariance of sampling error in $\hat{\alpha}_k$.

We produce shrunk value-added scores using the following equation:

$$\alpha_k^* = \Omega[\Omega + \Sigma_{kk}]^{-1}\hat{\alpha}_k$$

Where α_k^* is a *T* x 1 column vector of shrunk value-added scores for school *k* over the *T* years in which value-added is measured for school *k*. The shrunk value-added scores α_k^* are single-year scores that do not overrepresent small schools among the highest- and lowest-value-added score cases and also exhibit less year-to-year variability than before shrinkage. These scores are later aggregated across years into a weighted three-year average score (see the "Multi-year aggregation" section below).

The expected mean squared error of the shrunk value-added scores α_k^* is equal to:

$$EMSE_k = \Omega - \Omega[\Omega + \Sigma_{kk}]^{-1}\Omega$$

In practice, we use estimates of Ω and Σ_{kk} to estimate α_k^* and its expected mean squared error. The estimate of the matrix Σ_{kk} is the estimated variance-covariance matrix of the value-added results in $\hat{\alpha}_k$. Let $\hat{\sigma}_{t\tau kk}$ be the entry of this matrix in the row corresponding to $\hat{\alpha}_{kt}$ and the column corresponding to $\hat{\alpha}_{k\tau}$. The diagonal entries of this matrix are the squares of the estimated standard errors of the value-added results in $\hat{\alpha}_k$.

The diagonal entries of Ω, which are equal to the variance of α_{kt} across schools in a given year *t* and which we denote ω_{tt}, are estimated by computing the variance across schools *k* within year *t* of the value-added results $\hat{\alpha}_{kt}$, then subtracting from that the average across schools *k* within year *t* of $\hat{\sigma}_{ttkk}$, the estimated squared standard error of $\hat{\alpha}_{kt}$. This estimates the variance of the true school value-added for each year *t*, excluding variance due to randomness in the value-added results. The square root of this variance measure is also used for normalizing value-added results by grade before aggregation to school-level value-added results across all grades. The off-diagonal entries of Ω, which we denote $\omega_{t\tau}$ and are equal to the covariance of α_{kt} and $\alpha_{k\tau}$ across schools between years *t* and τ, is estimated by computing the covariance of the value-added results $\hat{\alpha}_{kt}$ and $\hat{\alpha}_{k\tau}$, and then subtracting from that the average error covariance estimate $\hat{\sigma}_{t\tau kk}$.

Student group value-added

Value-added is also estimated for student groups defined by certain student characteristics. Specifically, we calculate value-added scores for:

- the seven race/ethnicity groups;
- students with and without disabilities;
- economically disadvantaged and non-economically disadvantaged students;
- English-language learners[3] and non-English-language learners;
- students who were proficient (and not proficient) in the same subject in the previous year; and
- students who are in (and not in) a target group made up of students who were the lowest scoring 25% within their school in the same subject in the previous year.

To produce the student group value-added scores by school for all groups other than those defined by proficiency or by target group membership, we first estimate school-level value-added results for each student group. These are estimated by computing the sum of the school effects and the residual, $\alpha' S_i + \varepsilon_i$, for each student, and then averaging this sum across students for each student group in the school. We then shrink each of these results for 2022-23 jointly with the schools' results for that student group in 2021-22 using a multivariate shrinkage approach that considers correlations in school- and student group-level value-added across groups and years. This produces single-year student group results that exhibit less year-to-year variability and reduce the extent to which across-subgroup variability reflects random variation. After shrinkage, the student group results are re-centered for consistency so that the average of school value-added scores across the school's student groups (weighted by

[3] The English-language learners group for purposes of student group value-added estimation includes students who reached English language proficiency in the last four years.

the number of students in each group) is equal to the school's value-added score for all students. These are later aggregated across years into a weighted three-year average student group score (see the "Multi-year aggregation" section below).

Value-added by proficiency status

To produce proficient and not proficient student group results by school, we regress the sum of the school effects and residual, $\alpha' S_i + \varepsilon_i$, on same-subject pretest scores from the year immediately prior within each school. This regression is estimated in a way that accounts for measurement error in pretest scores, using approaches described in section 2.5 of Fuller (1987); this is a modification of the approach used to estimate the main value-added regression (described in the "Value-added regression" section above) that is better suited to the smaller, within-school samples. This regression produces a separate intercept and slope for each school. The intercept estimates the school's effect on a student with the average z-score for that school and subject. The slope estimates the relationship between student pretest scores and test score gains at that school.

Next, we shrink these intercepts and slopes for 2022-23 jointly with intercepts and slopes for 2021-22 using a multivariate shrinkage approach that considers correlations of the intercepts and slopes both with each other and over time. This produces shrunk, single-year intercepts and slopes that exhibit less random variability, including from year to year. After shrinkage, the intercepts are re-centered to achieve consistency between school value-added scores for all students and proficient/not-proficient student group value-added scores. Re-centering sets value-added for a student with the average z-score at the school equal to the school's all students value-added score for that year. School intercepts and slopes (after shrinkage and re-centering) are then used to produce value-added scores for each year for a representative non-proficient student and a representative proficient student. The representative non-proficient student is defined as one with a pretest z-score of -0.67, which corresponds to the average z-score for non-proficient students across grades and subjects statewide in 2017-18. The representative proficient student is defined as one with a pretest z-score of +0.86, which corresponds to the average z-score for proficient students across grades and subjects statewide in 2017-18. The 2017-18 averages continue to be used so that the interpretation of the proficient and not proficient student group results remained consistent over time. The single-year value-added scores for proficient and not proficient students are aggregated into three-year weighted average value-added scores for proficient and not proficient students (see the "Multi-year aggregation" section below).

Value-added for the target group

To produce target group value-added scores by school, we do not simply apply the approach used for the other categorical student groups such as English-language learner or students with disabilities. Doing so would produce scores that are biased upward in the target group (which is lower-scoring) and downward in the non-target group (which is higher-scoring). This is because the pretest scores used to assign students to the target group are inevitably measured with some degree of error. Due to this pretest measurement error, some students will have been erroneously assigned to the target group because their pretest scores are lower than what their true achievement would indicate. Since pretest measurement error does not affect posttest scores, these students are likely, in the absence of any adjustment, to have higher value-added estimates than their true knowledge gains would indicate. Similarly, because of pretest measurement error, some students will have been erroneously assigned to the non-target group. These students are likely, in the absence of any adjustment, to have lower value-added estimates than their true knowledge gains would indicate.

In 2020-21 and 2021-22, we approached this bias by initially producing value-added results in the same way that we did for the other categorical student groups but then making an adjustment before shrinkage that subtracted from the value-added results an estimate of the bias. In 2022-23, we approached this bias in a new way that produced growth residuals that could be averaged across students to produce target group value-added results without further adjustment. This was implemented by re-estimating the value-added regression with the observed pretests replaced by predictions of true pretest achievement and with the school fixed effects replaced by interactions between the school fixed effects and an indicator for target group status. The regression is estimated by two-stage least squares, with the predictions of true pretest achievement instrumented with the observed pretest scores. We describe these approaches in Appendix B.

After adjusting for bias using either approach, we subtracted, from each school's target group value-added result, the average of target group value-added results across all schools. Likewise, we subtracted, from each school's non-target group, the average of non-target group value-added results across all schools. This is because we were concerned, even after adjustments for pretest measurement error, that the average difference between the target and non-target group measures statewide did not necessarily reflect a statewide difference in school effectiveness between the two groups. In particular, we were concerned that it could reflect a possible non-linearity in the relationship between the posttest and the pretests. It is possible that, even in the absence of a target group being identified, the relationship between the posttest and the pretest may "bend" in a way that growth appears to become faster or slower statewide among lower-achieving students. This could appear to be a statewide effect of target group membership, even if it is only a side effect of the shape of the posttest-pretest relationship under "business as usual" conditions. Subtracting out the statewide mean of the target and non-target group value-added results eliminated this statewide difference and avoided the possibility of attribution of a difference that may not be causal.

The target and non-target group value-added results were shrunk using a bivariate shrinkage approach that accounts for the correlation of growth within schools between the target and non-target group. This step was implemented to reduce the impact of differences in growth between the target and non-target groups that were the result of random variability. This is the same approach as employed in shrinkage of value-added results for other student groups, such as disability, with the exception that the target group results were shrunk separately year by year rather than jointly across years. The exception was made so that a school's growth measure for the target group for a given year specifically reflected the growth of students in that target group in that year, in part to recognize schools that made substantial improvements from one year to the next in growth among students in the target group. The shrunk value-added scores were then re-centered within school to ensure that the average of school value-added scores across the target and non-target groups (weighted by the number of students in the two groups) aligns with the school's value-added score for all students. The value-added scores for the target and non-target groups are later aggregated across years into a weighted three-year average score for each group (see the "Multi-year aggregation" section below).

We computed district-level results for the target and non-target groups by averaging the school-level results for each group across schools within the district. We did not include in district-level value-added results for the target and non-target groups students who were not enrolled in a single school for the full academic year. This is because the target group is defined by students' prior achievement level relative to other students within their school.

Final stage for estimation of school and district value-added scores

MULTI-YEAR AGGREGATION

Final school value-added scores are calculated as a weighted three-year average of scores for 2020-21, 2021-22, and 2022-23. The weights used are equal to the number of students included in the school's value-added score for a given year, multiplied by 1.5 for 2022-23, 1.0 for 2021-22, and 0.5 for 2020-21. The multi-year averaged value-added score includes the 2020-21 and/or 2021-22 value-added scores only if there are at least twenty students associated with that specific year's value-added scores. All value-added scores, including the student group scores, are reported as a multi-year average using the weighting described above.

The multi-year average value-added scores are rescaled, based on the number of years included, to have a variance like that of a single-year value-added score. The rescaling is done because an average of value-added scores over multiple years will tend to have a lower variance across schools than a single-year value-added score, and the more years are included in the average, the lower the variance will tend to become. In the absence of rescaling, the highest and lowest value-added scores will be disproportionately among schools in which the average includes only one or two years rather than three. To implement the rescaling, we first compute the standard deviations of four different aggregations of overall value-added scores: a single-year score from 2022-23; a two-year weighted average from 2021-22 and 2022-23; a two-year weighted average from 2020-21 and 2022-23; and a three-year weighted average from 2020-21, 2021-22, and 2022-23. Then, we divide all of a given school's multi-year value-added scores by the standard deviation for the kind of aggregation of the school's overall value-added score, and then multiply it by the standard deviation for single-year overall scores from 2022-23. This rescaling puts the variances of the multi-year averages of overall value-added scores at approximately the same level, regardless of which years are included in the average. It also makes the same adjustment to all of a given school's multi-year averages, whether they are of overall value-added or of student group value-added.

It is important to note that the 2020-21 value-added scores that enter into the multi-year average are estimated using a “skip-year” approach that accounts for there being two years rather than one between the posttest (administered in 2020-21) and the pretest (administered in 2018-19). The skip-year value-added method is described in Appendix A.

CALCULATING DISTRICT-LEVEL SCORES

Final district value-added scores are obtained by averaging the multi-year value-added scores for the schools in each district, with weights determined by the number of students included in each school’s single-year value-added score for 2022-23. This same procedure is used for the all students and the student group value-added scores for the district. As mentioned earlier, the district scores include students if they were FAY at the district even if they were not FAY at any of the district’s schools. Thus, students who moved from one school in a district to another school in the district are included. These students are incorporated into the district score using a placeholder school for each district for students who were FAY in the district but not FAY in any school in the district.

PROPERTIES OF THE VALUE-ADDED RESULTS

Coefficients on student-level variables in the model

The coefficients estimated in the value-added model are presented in Tables 4 and 5. To interpret these coefficients, note that both pretest and posttest scores are measured using z-scores; therefore, all coefficients are measured in the posttest standard deviation scale. For example, note that the coefficient on female gender is -0.017 in grade 5 math. This implies that male students improved by about 0.017 standard deviations more on the grade 5 math test than otherwise similar female students.

It is important to keep in mind the standard errors of the coefficients when interpreting them. A span of 1.96 standard errors in both the positive and negative directions provides a 95 percent confidence range for a coefficient. Continuing with the example of the coefficient on female gender in grade 5 math, note that the standard error of this coefficient estimate is 0.005. This means that, while our best estimate of the difference in growth between female and male students is -0.017 standard deviations of 5th grade achievement, a 95 percent confidence interval for the difference ranges from -0.027 to -0.007 standard deviations.

Table 4. Coefficients on Student-Level Variables, 2022-23 Math

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11	
Variable	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**
Math Pretest (lag 1)	0.836	0.006	0.705	0.013	0.621	0.013	0.672	0.014	0.535	0.015	0.421	0.014	0.499	0.011	0.519	0.011
ELA Pretest (lag 1)	0.052	0.006	0.043	0.011	0.074	0.011	0.080	0.011	0.150	0.012	0.038	0.012	0.082	0.009	0.035	0.012
Math Pretest (lag 2)	n/a	n/a	0.170	0.013	0.234	0.013	0.218	0.013	0.348	0.015	0.484	0.014	0.347	0.012	0.366	0.011
ELA Pretest (lag 2)	n/a	n/a	0.006	0.011	-0.009	0.010	-0.022	0.010	-0.112	0.012	-0.058	0.013	-0.005	0.009	0.009	0.012
ELP Level 1	0.098	0.068	-0.053	0.079	0.139	0.071	-0.088	0.077	-0.091	0.082	0.256	0.118	0.406	0.186	0.201	0.151
ELP Level 2	0.073	0.066	-0.067	0.071	0.088	0.065	-0.033	0.071	-0.076	0.077	0.066	0.110	0.219	0.175	0.361	0.130
ELP Level 3	0.069	0.065	-0.007	0.068	0.119	0.061	0.029	0.069	-0.083	0.074	-0.018	0.107	0.066	0.172	0.292	0.125
ELP Level 4	0.126	0.066	0.006	0.068	0.145	0.061	0.050	0.071	-0.076	0.076	-0.169	0.108	0.058	0.173	0.237	0.126
ELP Level 5	0.166	0.121	-0.012	0.094	0.188	0.082	0.071	0.111	-0.176	0.129	-0.021	0.218	-0.232	0.249	0.088	0.196
ELP Level 6	0.131	0.067	0.025	0.068	0.165	0.061	0.045	0.069	-0.119	0.074	-0.098	0.106	0.081	0.172	0.268	0.124
ELP Level 7	0.082	0.065	-0.034	0.067	0.119	0.060	0.021	0.069	-0.119	0.073	-0.055	0.106	0.106	0.172	0.288	0.124
Female	-0.077	0.005	-0.017	0.005	0.017	0.004	-0.034	0.004	0.057	0.005	-0.096	0.006	-0.140	0.005	-0.115	0.005
Non-Binary	0.203	0.159	-0.347	0.203	-0.174	0.162	-0.068	0.091	0.088	0.075	-0.049	0.105	-0.084	0.079	-0.036	0.085
Asian	0.033	0.012	0.120	0.013	0.055	0.012	0.033	0.013	0.102	0.013	0.030	0.016	-0.024	0.015	-0.007	0.014
African American	-0.071	0.011	-0.003	0.012	-0.019	0.011	-0.051	0.012	0.022	0.012	-0.010	0.014	-0.032	0.014	-0.024	0.015
Hispanic	-0.030	0.008	-0.007	0.009	-0.022	0.008	-0.009	0.008	0.002	0.009	-0.025	0.011	-0.025	0.010	-0.018	0.010
American Indian or Alaskan Native	-0.015	0.025	-0.041	0.026	-0.001	0.023	0.023	0.024	0.013	0.025	-0.005	0.032	0.040	0.030	-0.069	0.029
Native Hawaiian or Other Pacific Islander	-0.040	0.073	0.067	0.081	-0.060	0.075	0.048	0.090	0.100	0.092	0.261	0.099	-0.101	0.095	-0.103	0.081
Two or More Races	-0.012	0.009	0.011	0.010	-0.010	0.010	-0.008	0.010	0.010	0.011	-0.033	0.013	0.004	0.013	-0.001	0.013

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11	
Special Education EBD	-0.100	0.021	-0.124	0.020	-0.095	0.018	-0.052	0.018	-0.112	0.019	0.114	0.023	0.066	0.023	0.124	0.027
Special Education LD/ID	-0.044	0.011	-0.085	0.010	-0.085	0.010	-0.039	0.010	-0.044	0.011	0.038	0.014	0.081	0.013	0.039	0.013
Special Education A	-0.065	0.017	-0.025	0.018	-0.099	0.018	-0.014	0.018	-0.004	0.019	0.050	0.024	-0.034	0.025	0.019	0.023
Special Education SL	0.006	0.011	0.006	0.013	0.002	0.017	0.012	0.022	-0.004	0.029	-0.100	0.046	0.029	0.050	-0.067	0.051
Special Education Other	-0.077	0.011	-0.070	0.011	-0.092	0.011	-0.049	0.011	-0.061	0.011	0.054	0.015	0.043	0.014	0.015	0.014
Economic Disadvantage	-0.030	0.005	-0.024	0.005	-0.023	0.005	0.001	0.005	-0.024	0.005	-0.010	0.006	-0.025	0.006	-0.048	0.006
Migrancy Status	-0.145	0.181	0.046	0.211	-0.052	0.176	0.030	0.159	-0.022	0.269	-0.083	0.214	0.279	0.302	0.283	0.243

Table 5. Coefficients on Student-Level Variables, 2022-23 ELA

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11	
Variable	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**
Math Pretest (lag 1)	0.160	0.007	0.109	0.014	0.072	0.016	0.167	0.016	0.169	0.016	0.130	0.014	0.082	0.011	0.046	0.010
ELA Pretest (lag 1)	0.723	0.007	0.567	0.012	0.620	0.013	0.549	0.012	0.589	0.013	0.384	0.012	0.547	0.009	0.502	0.011
Math Pretest (lag 2)	n/a	n/a	-0.014	0.014	-0.029	0.015	-0.073	0.014	-0.037	0.016	-0.050	0.015	-0.005	0.011	0.045	0.010
ELA Pretest (lag 2)	n/a	n/a	0.251	0.012	0.221	0.012	0.282	0.011	0.200	0.013	0.431	0.013	0.325	0.009	0.330	0.011
ELP Level 1	0.098	0.077	-0.139	0.087	0.071	0.084	-0.181	0.085	-0.048	0.086	0.020	0.118	0.184	0.184	0.063	0.137
ELP Level 2	0.069	0.075	-0.033	0.079	0.070	0.076	-0.103	0.078	0.014	0.081	-0.165	0.111	0.113	0.173	0.102	0.117
ELP Level 3	0.151	0.074	0.003	0.075	0.087	0.072	-0.057	0.076	-0.009	0.078	-0.317	0.107	-0.081	0.170	0.052	0.112
ELP Level 4	0.205	0.076	0.037	0.075	0.100	0.072	-0.008	0.077	0.009	0.080	-0.352	0.108	-0.107	0.171	0.023	0.113
ELP Level 5	0.386	0.138	0.111	0.104	0.229	0.097	-0.099	0.121	-0.049	0.136	-0.421	0.219	-0.158	0.245	0.152	0.176
ELP Level 6	0.235	0.076	0.028	0.075	0.121	0.072	-0.040	0.076	-0.030	0.078	-0.324	0.107	-0.130	0.169	0.034	0.112

	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11	
ELP Level 7	0.166	0.074	-0.004	0.074	0.075	0.071	-0.069	0.075	-0.070	0.077	-0.279	0.107	-0.052	0.170	0.037	0.112
Female	0.014	0.005	0.081	0.005	0.081	0.005	0.041	0.005	0.105	0.005	0.061	0.006	0.002	0.005	-0.017	0.005
Non-Binary	0.330	0.181	-0.248	0.225	-0.212	0.192	-0.090	0.099	0.124	0.080	0.149	0.108	0.032	0.078	0.282	0.076
Asian	-0.008	0.014	0.032	0.014	0.099	0.015	0.078	0.014	0.060	0.014	-0.042	0.016	-0.005	0.015	-0.007	0.013
African American	-0.049	0.012	-0.013	0.013	0.001	0.013	-0.001	0.013	-0.006	0.013	-0.094	0.014	-0.042	0.014	-0.027	0.013
Hispanic	-0.004	0.009	0.013	0.009	0.001	0.010	0.000	0.009	-0.016	0.009	-0.059	0.011	-0.024	0.010	-0.005	0.009
American Indian or Alaskan Native	-0.008	0.028	-0.049	0.028	-0.020	0.028	0.025	0.027	-0.018	0.027	-0.024	0.032	0.033	0.029	-0.098	0.026
Native Hawaiian or Other Pacific Islander	0.178	0.083	0.051	0.090	-0.102	0.089	-0.026	0.098	0.040	0.097	0.018	0.101	-0.059	0.094	-0.051	0.073
Two or More Races	0.013	0.011	-0.008	0.011	-0.008	0.011	0.009	0.011	0.005	0.011	-0.044	0.013	0.026	0.013	0.007	0.011
Special Education EBD	-0.058	0.024	-0.107	0.022	-0.100	0.021	-0.024	0.020	-0.088	0.020	0.296	0.024	0.301	0.023	0.012	0.025
Special Education LD/ID	-0.061	0.012	-0.086	0.012	-0.076	0.012	-0.062	0.011	-0.034	0.011	0.097	0.014	0.224	0.013	-0.011	0.012
Special Education A	-0.083	0.019	-0.124	0.020	-0.084	0.021	-0.008	0.020	-0.042	0.020	0.267	0.025	0.286	0.025	-0.053	0.022
Special Education SL	0.004	0.013	-0.021	0.015	0.022	0.020	0.016	0.024	0.056	0.030	0.014	0.046	0.085	0.049	-0.071	0.046
Special Education Other	-0.089	0.012	-0.108	0.013	-0.099	0.013	-0.037	0.012	-0.041	0.012	0.177	0.015	0.220	0.014	-0.056	0.013
Economic Disadvantage	-0.040	0.006	-0.031	0.006	-0.042	0.006	-0.011	0.005	-0.031	0.005	0.000	0.006	-0.004	0.006	-0.050	0.005
Migrancy Status	0.414	0.207	-0.455	0.233	-0.079	0.208	0.331	0.174	0.235	0.284	-0.132	0.215	0.642	0.298	-0.001	0.219

Test of model neutrality: Correlation with average prior attainment

In this test, we calculate correlations between value-added estimates and school-level prior attainment. The value-added estimates employed in these correlations are the multi-year averages of value-added, post-shrinkage, for all students for a given subject. School-level prior attainment is computed as the average of 2021-22 scale score in the same subject in the previous grade within the value-added regression data set. The scale scores are converted to (non-rank-based) z-scores by subtracting out the scale score mean by subject and grade and dividing by the scale score standard deviation by subject and grade. This makes it possible to meaningfully average scale scores across grades when required. In the grade-level correlations, both value-added and prior attainment are measured by grade. In the overall correlation, both value-added and prior attainment are multi-grade measures.

This correlation is a method for validating whether the variables included on the right-hand side of our regression adequately control for school-level factors influencing value-added estimates. The higher the correlation magnitude, the higher the level of “non-neutrality”. We do not necessarily expect these correlations to be zero, particularly if schools with students with higher prior attainment also tend to be more effective at facilitating growth among their students. However, we may be concerned if this correlation is especially high (particularly if greater than 0.5), since this may be a sign that the control variables are not sufficiently "leveling the playing field" between schools with students with lower and higher prior attainment.

Our results show a low correlation at the school-and-grade level and a modest correlation at the school level across all grades between average prior attainment--a measure of average performance in the previous year--and value-added. For example, the correlation between multi-year value-added in English language arts in grade 4 and average 2021-22 scale score in grade 3 in English language arts is 0.040, which is very low. For context, it is often difficult to see correlations with absolute values of less than 0.2 on a scatterplot. On the other hand, the correlation between multi-year, multi-grade value-added in math and average 2021-22 scale score in math is 0.304, which is more substantial, but not so large as to create concerns about the value-added measures. In general, schools were somewhat more likely to have a high value-added score than a low one if their students had high pretest scores rather than low pretest scores.

Table 6. Correlations between Prior Attainment and Value-Added

Subject	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Overall
ELA	0.040	0.148	0.169	0.138	0.155	0.001	0.074	-0.039	0.217
Math	0.067	0.078	0.166	0.034	0.338	0.042	0.252	0.103	0.304

Correlation between Math and ELA value-added

There were substantive positive correlations between the multi-year math and ELA value-added scores, post-shrinkage, within each school. Schools with high value-added scores in math also tended to have high value-added scores in ELA. This implies that schools with a higher-than-average impact in mathematics also had a higher-than-average impact in English language arts. This is not a surprising result, given that we would expect that many of the aspects of a school that facilitate high growth in one subject would also facilitate growth in other subjects.

Table 7. Correlations between Subjects

Subjects	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Overall
2022-23 Math & ELA	0.610	0.586	0.613	0.520	0.450	0.642	0.509	0.539	0.582

CONTACT

For more information, contact the Growth Team at Education Analytics at growth@edanalytics.org.

REFERENCES

Fuller, W. (1987). *Measurement Error Models*, John Wiley and Sons.

Longford, N. T. (1999). Multivariate shrinkage estimation of small area means and proportions. *Journal of the Royal Statistical Society* 162 (Part 2), 227-245.

APPENDIX A: 2020-21 SKIP-YEAR GROWTH

Value-added results in 2022-23 were measured in the typical fashion, with only one year between the posttest (administered in 2022-23) and the pretest (administered in 2021-22). The same is true for value-added results in 2021-22, which measured growth between the 2020-21 and 2021-22 assessments. In both cases, growth between the posttest and pretest assessments reflects the experience of a student from one grade to the next and one year to the next.

However, value-added growth in 2020-21 was unusual because the most recent pretest available was from 2018-19, two years before the posttest. This continues to be relevant for the multi-year average value-added scores in 2022-23, because it contains up to three years of single-year value-added results. Growth between the assessments in 2018-19 and 2020-21 reflects the experience of a student over two consecutive grades over two consecutive years. To take this into account, the school indicators S_i for 2020-21 value-added were designed to indicate the *combination* of schools attended by students in 2019-20 and 2020-21. For example, there may be an indicator for students who attended school A in 2019-20 and school B in 2020-21; another for students who attended school A in 2019-20 and school C in 2020-21; and a third for students who attended school C in both 2019-20 and 2020-21.

Estimating the value-added model with these indicator variables produces effects for each *combination* of schools that appear in the data set. From these, we produced school value-added results for 2020-21 by averaging the estimated effects across all combinations that include a given school, before the shrinkage technique was applied. This average is weighted by the number of students in the data set associated with that combination of schools, multiplied by 1 if the combination is for the same school in both 2019-20 and 2020-21 and by 0.5 if the combination is for two different schools in 2019-20 and 2020-21.

This is best explained with an example. Suppose that we have three indicators that include school D in some way: one for twenty students who attended school D in both 2019-20 and 2020-21; another for two students who attended school D in 2019-20 and school E in 2020-21; and a third for four students who attended school F in 2019-20 and school D in 2020-21. The school value-added results for school D would be a weighted average of the effects for these three combinations, with a weight of 20 x 1 = 20 on the first combination, a weight of 2 x 0.5 = 1 on the second combination, and a weight of 4 x 0.5 = 2 on the third combination.

The grade-level skip-year growth measures for a given school for 2020-21 were aggregated using a weighted average to produce multi-grade skip-year value-added results for that school for 2020-21. The weight used in the weighted average was a weighted count of students that counts students associated with the school in both 2019-20 and 2020-21 with full weight and students associated with the school in only one of the two years with half weight.

The approaches for producing the student group value-added results were also adapted for the skip-year nature of growth in 2020-21. The student group value-added results, other than those for proficiency level, were adapted for skip-year growth by weighting. Recall that student group value-added results, other than those for proficiency, are produced by computing the sum of the school effects and the residual, $\alpha' S_i + \varepsilon_i$, for each student, and then computing the average of this variable by school and student group. In the skip-year case of 2020-21, this average was weighted by whether a student was in the school for both 2019-20 and 2020-21 (in which case the student entered the average with full weight) or for only one of the two years (in which case the student entered the average with half weight).

The student group value-added results for proficient and non-proficient groups were also adapted for skip-year growth by weighting. Recall that these are produced by regressing the sum of the school effects and residual, $\alpha' S_i + \varepsilon_i$, on same-subject, once-lagged pretest score within each school. In the skip-year case of 2020-21, this regression was estimated as a weighted regression, with students who were in the school in both years entering with full weight and students who were in the school in only one of 2019-20 or 2020-21 entering with half weight.

APPENDIX B: TARGET GROUP MODELS

As noted in the section "Value-added for the target group", producing the target group measures using the same approach as that used to produce other student group measures will lead to biases in the absence of any adjustment. We discuss in detail below the approaches used to eliminate this bias.

In 2020-21 and 2021-22, we produced value-added results in the same way as other categorical student groups, but then, before applying shrinkage, made an adjustment that subtracted from the value-added results an estimate of the bias. This estimate of bias was based on the standard error of measurement of the pretest scores and an assumption that pretest score error is normally distributed. The adjustments were calculated as follows:

Adjustment for target group: $adj\_target_k = -\lambda \frac{\sigma^2_{v(k)}}{\sqrt{\sigma^2_{y*(k)} + \sigma^2_{v(k)}}} \frac{\phi(z_k)}{\Phi(z_k)}$ (A2.1)

Adjustment for non-target group: $adj\_nontarget_k = +\lambda \frac{\sigma^2_{v(k)}}{\sqrt{\sigma^2_{y*(k)} + \sigma^2_{v(k)}}} \frac{\phi(z_k)}{(1-\Phi(z_k))}$ (A2.2)

where adj_target_k and $adj\_nontarget_k$ are the bias adjustments added to the target and non-target group results respectively for school *k;* λ is the coefficient on same-subject, once-lagged pretest; $\sigma^2_{y*(k)}$ is an estimate of the variance in school *k* of same-subject, once-lagged pretest adjusted for measurement error; $\sigma^2_{v(k)}$ is an estimate of the variance in school *k* of measurement error in the same-subject, once-lagged pretest; z_k is the cutoff score in school *k* for inclusion in the target group, after transforming to z-score; $\phi(.)$ is the standard normal probability density function; and $\Phi(.)$ is the standard normal cumulative distribution function.

In 2022-23 a different approach was used to deal with this bias. This approach employs predictions of the true pretest variables y_{2i}, y_{2i}^{alt}, y_{1i}, and y_{1i}^{alt} given all the right-hand-side variables in equation (7) in the "Value-added model" section, including pretest scores, student characteristic variables, and school indicator variables. The use of these predictions makes the adjustments employed in previous years unnecessary.

The predictions are produced using the following steps. First, let $Y_{t\text{-}\ell}$ be an N × 4 matrix of pretest scores Y_{2i}, Y_{2i}^{alt}, Y_{1i}, and Y_{1i}^{alt} -- the row of $Y_{t\text{-}\ell}$ corresponding to student *i* is denoted as $Y_{it\text{-}\ell}$. Let *W* be an N × K matrix of the *X* student characteristic variables and *S* school indicator variables. The pretest scores $Y_{t\text{-}\ell}$ are regressed by ordinary least squares on the other variables *W*, producing a predicted component $\hat{Y}_{t-\ell}$ and a residual component $\hat{u}_{t-\ell}$.

Auxiliary regression: $Y_{t-\ell} = W\pi + u_{t-\ell} = X\pi_1 + S\pi_2 + u_{t-\ell}$ (A2.3)

Estimated predicted component: $\hat{Y}_{t-\ell} = W\hat{\pi} = X\hat{\pi}_1 + S\hat{\pi}_2$ (A2.4)

Estimated residual component: $\hat{u}_{t-\ell} = Y_{t-\ell} - \hat{Y}_{t-\ell}$ (A2.5)

The predicted component $\hat{Y}_{t-\ell}$ should not include pretest measurement error at all, given that we do not expect pretest measurement error to be predictable from other variables. However, the residual $\hat{u}_{t-\ell}$ has a substantial component that is pretest measurement error.

Second, we estimate the variance-covariance matrix of the component of the residual ($\hat{u}_{t-\ell}$) that is *not* pretest measurement error as follows:

Estimated error-adjusted covariance: $\Omega_u = (1/N)\hat{u}_{t-\ell}{}'\hat{u}_{t-\ell} - (1/N)\sum_{i=1}^{N} V_{it-\ell}$ (A2.6)

Where $V_{it-\ell}$ is a 4 × 4 diagonal matrix with the squared conditional standard errors of measurement of the elements of $Y_{it\text{-}\ell}$ on the diagonal. The $(1/N)\hat{u}_{t-\ell}{}'\hat{u}_{t-\ell}$ term estimates the total variance of $\hat{u}_{t-\ell}$, while the $(1/N)\sum_{i=1}^{N} V_{it-\ell}$ term estimates the variance of the component of $\hat{u}_{t-\ell}$ that is pretest measurement error. Consequently, the difference between the two terms (Ω_u) estimates the variance of the component of $\hat{u}_{t-\ell}$ that is *not* pretest measurement error.

Third, we predict the non-measurement-error component of $\hat{u}_{it-\ell}$ as follows:

Predicted non-error component: $\tilde{u}_{it-\ell} = \Omega_u\left(\Omega_u + (1/N_{k(i)})\sum_{i\in k(i)} V_{it-\ell}\right)^{-1}\hat{u}_{t-\ell}$, (A2.7)

Where $N_{k(i)}$ is the number of students in the school attended by student *i* and $\sum_{i\in k(i)} V_{it-\ell}$ is the sum of $V_{it-\ell}$ over students in the school attended by student *i.* The prediction $\tilde{u}_{it-\ell}$ is an Empirical Bayes estimate that uses conditional standard errors of measurement

(CSEMs) to account for the fact that the variance--and, consequently, the relevance--of pretest measurement error is likely to be greater in some schools than in others.

Fourth, the prediction of the true pretests y_{2i}, y_{2i}^{alt}, y_{1i}, and y_{1i}^{alt}, which we denote as the vector $\tilde{y}_{it-\ell}$, is computed by adding $\hat{Y}_{it-\ell}$ to $\tilde{u}_{it-\ell}$.

Predicted true pretest achievement: $\tilde{y}_{it-\ell} = \hat{Y}_{it-\ell} + \tilde{u}_{it-\ell}$ (A2.8)

After producing these predicted true pretest scores $\tilde{y}_{it-\ell}$, we estimate an instrumental-variables regression in which the left-hand-side variable is the posttest score Y_{3i}; the right-hand-side variables are the predicted pretest scores $\tilde{y}_{it-\ell}$, the student characteristic variables X_i, and interactions between the school indicator variables S_i and an indicator for target group status; and the predicted pretest scores $\tilde{y}_{it-\ell}$ are instrumented with the pretest scores $Y_{it-\ell}$. This regression is estimated using two-stage least-squares. In the first stage, $\tilde{y}_{it-\ell}$ is regressed on $Y_{it-\ell}$, X_i, and the interactions between S_i and the target group indicator, producing a prediction $\hat{\tilde{y}}_{it-\ell}$. In the second stage, Y_{3i} is regressed on $\hat{\tilde{y}}_{it-\ell}$, X_i, and the interactions between S_i and the target group indicator.

First-stage regression: $\tilde{y}_{it-\ell} = \tau_1{}'Y_{it-\ell} + \tau_2{}'X_i + \tau_3{}'(S_i \times target_i) + \eta_{it-\ell}$ (A2.9)

Prediction from first stage: $\hat{\tilde{y}}_{it-\ell} = \hat{\tau}_1{}'Y_{it-\ell} + \hat{\tau}_2{}'X_i + \hat{\tau}_3{}'(S_i \times target_i)$ (A2.10)

Second-stage regression: $Y_{3i} = \lambda^{*\prime}\hat{\tilde{y}}_{it-\ell} + \beta^{*\prime}X_i + \alpha^{*\prime}(S_i \times target_i) + \epsilon_{it-\ell}$ (A2.11)

We use the coefficients from this regression to create a special growth residual, equal to the posttest score Y_{3i} minus the products of $\tilde{y}_{it-\ell}$ and X_i and their coefficient estimates in the second-stage regression.

Target group growth residual: $q_i{}^* = Y_{3i} - \hat{\lambda}^{*\prime\hat{\tilde{y}}_{it-\ell}} - \hat{\beta}^{*\prime}X_i$ (A2.12)

These growth residuals can be averaged across students by school and target group status to produce target group value-added results at the school level that do not need to be further adjusted for pretest measurement error.