Evaluating Pandemic Related Impacts on Student Achievement

Arizona Statewide Assessment in English Language Arts and Mathematics

2020–2021 School Year

October 2021

ARIZONA STATEWIDE ASSESSMENT

ARIZONA'S STATEWIDE ACHIEVEMENT ASSESSMENT (AZM2)

ENGLISH LANGUAGE ARTS GRADES 3–11

MATHEMATICS GRADES 3–8, GRADE 10

EVALUATING PANDEMIC RELATED IMPACTS ON STUDENT ACHIEVEMENT

OCTOBER 2021

Prepared by Cambium Assessment, Inc. (CAI) in collaboration with
the Arizona Department of Education (ADE)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION: EVALUATING PANDEMIC IMPACTS ON STUDENT ACHIEVEMENT

State summative assessments were cancelled in spring 2020 just prior to the opening of most state testing windows. As a result, states do not have a spring 2020 measure of achievement available against which to measure losses in student achievement due to pandemic related impacts on instruction. While many schools reopened in spring 2020 employing remote instruction, the length of time that schools remained closed and the ability of schools to provide effective remote instruction varied considerably. Schools opened in fall 2020 employing a range of in-person, remote, and hybrid instruction. Although many states sought to again cancel state assessments for 2021, USED mandated that states continue to assess student achievement of state standards, although accountability of districts to student achievement and growth were postponed. Thus, the spring 2021 test administration provides the first opportunity for states to investigate systematically the impacts of pandemic related disruptions in instruction on student achievement.

Evaluation of pandemic-related impacts on student achievement is made difficult because the student population is not consistent between the pre- and post-pandemic test administrations. Students have left the public education system for a number of reasons, including transferring to private schools, home schooling, or they have simply dropped out of the education system. Because the student population has changed between pre- and post-pandemic, direct comparisons of cohort changes in achievement provide an incomplete understanding of pandemic impacts on student achievement. For example, if students who are no longer participating in state assessments were lower achieving pre-pandemic, then any observed declines in student achievement post-pandemic will be underestimated, since achievement declines among already lower achieving students would not be adequately represented.

To better understand the impacts of the pandemic on student achievement, we identified two analysis strategies designed to control for changes in the tested population in order to examine pandemic related impacts on student achievement. In an initial series of analyses, we used matched samples of students across cohorts to control for differences in achievement and demographic subgroup membership between the two cohorts of students (Ho, 2021). In this approach we built a regression model by first regressing student achievement in 2019 onto student achievement and demographic characteristics of those same students in spring 2017. All students available in the 2017-2019 cohort were used to build the model. This regression model represents the pre-pandemic two-year growth. Since this analysis requires merging student records across a two-year span, it is limited to only those students in grades 5-8 in 2019 who were administered state assessments in grades 3-6 in 2017. We then identified students who were tested post-pandemic in spring 2021 (in grades 5-8) who also participated in state testing in spring 2019 (in grades 3-6). We used all students available in the 2019-2021 cohort as given, and found a matched sample in the 2017-2019 cohort. The matching was based on the grade *g-2* scale scores between the two cohorts using 1:1 nearest neighbor sampling method. We applied the regression coefficients to the grade 3-6 scores in 2017 in the matched sample to predict their grade 5-8 scores in 2019. In this way, the pandemic related impacts on student achievement can be evaluated by comparing the observed 2021 grade 5-8 scores to the predicted 2019 grade 5-8 scores between a pair of matched samples. This approach can provide a better estimate of pandemic related impacts on student achievement for the general education population overall, as well as for demographic subgroups.

It is also possible to investigate the expected performance of students who did not participate in spring 2021. In this approach, characteristics of students in the pre-pandemic cohort are used to predict non-participation in the spring 2021 sample. With the regression coefficients in hand, the prediction model can be applied to the 2017 test records of a sample of students matched to the non-participating students in spring 2021 to predict their 2019 performance, allowing us to estimate the expected level of performance of students who did not participate in the 2021 test administration based on their pre-pandemic performance. While this approach cannot address how those students

may have been impacted by the pandemic, it may provide a picture of the pre-pandemic performance of those non-participating students.

While we drew the matched samples the same way we did in Ho's approach, we conducted the matched sample analyses in a way that is slightly different than Ho's. We drew the matched samples the same way we did in Ho's approach. Rather than using all available students in the 2017-2019 cohorts, the regression model was first constructed for the pre-pandemic matched sample by regressing the 2019 scores on the 2017 scores. Assuming that the two-year growth relationship provides a consistent expectation for growth across cohorts, the regression coefficients were applied to the 2019 scores to predict the 2021 scores of the post-pandemic sample, assuming no pandemic effects on instruction. We were then able to evaluate the pandemic-related impact on achievement by comparing the observed 2021 scores with the expected 2021 scores. We note that the two approaches produced nearly identical results. The difference in the predicted average scale score is within one scale score point, which is mostly due to rounding.

In addition to the matched sample cohort analysis, we also sought to investigate the relationships between prior achievement on subsequent achievement more directly and how that relationship may have been impacted by the pandemic. As part of this analysis, we also sought to investigate whether subgroup differences in achievement gains were differentially impacted by pandemic-related disruption to instruction.

In this approach, we produced a regression model to predict student achievement at time two from student achievement and demographic subgroup membership at time one simultaneously using both the 2017 and 2019 cohorts of students (i.e., those initially tested in 2017 in grades 3-6, and those initially tested in those same grades in 2019). All students available in the 2017 and 2019 cohorts were used to build this regression model. This approach allows us to evaluate whether the relationships between prior and subsequent achievement differ across cohorts, as well as whether the relationships between demographic subgroups on subsequent achievement differ across cohorts, indicating differential impacts of the pandemic by subgroup.

EXAMINING PANDEMIC IMPACTS ON STUDENT ACHIEVEMENT IN MATCHED SAMPLES OF STUDENT COHORTS

Following the approach described by Ho (2021), we began by identifying matched samples of students testing pre- and post-pandemic. To identify matched samples of students we applied the following steps.

Step 1. First, we built a regression model by regressing student achievement in 2019 onto student achievement and demographic characteristics of those same students in 2017. The demographic variables include gender, ethnicity, English learner (LEP) status, special education (SPED) status and low income status. All variables were entered into the regression equation simultaneously:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

where *Y* is the 2019 score, $X_1$is the 2017 score, and $X_2 \sim X_n$ refer to the demographic variables. The bidirectional stepwise selection algorithm was used to identify the final predictors. The adjusted R^2 values ranged from 0.67 to 0.69 for the regression models predicting ELA achievement, and R^2 values ranged from 0.66 to 0.70 for regression models predicting mathematics achievement. The estimated regression weights for each model are provided in Appendix A.

Step 2. The student testing population has changed between pre- and post-pandemic. To enable an appropriate comparison between the performance of students in 2021 and the performance of their academic peers in 2019, we first identified students who had test scores available in both 2019 and 2021 and labeled those students as Sample

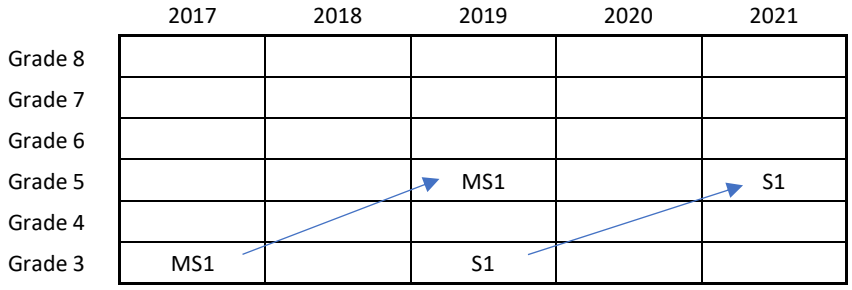
1 (S1). Those students who had 2019 scores, but not 2021 scores, were labeled as Sample 2 (S2). Table 1 shows the number of students and their average 2019 scores in S1 and S2. We found that the students who didn't participate in 2021 tests tended to be lower achieving students in spring 2019.

Table 1. Comparison between Students with Both 2019 and 2021 Scores and Students with Only 2019 Scores

Test (2019-2021)	Students with 2019 & 2021 Scores (S1)		Students with Only 2019 Scores (S2)	
	N Count	Average 2019 Score	N Count	Average 2019 Score
G3E-G5E	65909	2506	16873	2499
G4E-G6E	68555	2525	18221	2517
G5E-G7E	69151	2544	21012	2535
G6E-G8E	69609	2547	20646	2539
G3M-G5M	67141	3529	16053	3517
G4M-G6M	69697	3559	17251	3546
G5M-G7M	70553	3590	19727	3578
G6M-G8M	71011	3619	19381	3606

Step 3. In the third step we sought to identify matched samples of students in S1 from the cohort of students participating in the spring 2017 and spring 2019 test administrations. In this approach, we began by identifying students with the same scale base year scale scores between cohorts. For example, for a student who was administered a grade 3 ELA test in 2019, we drew a student with same grade 3 ELA score from among the 2017 grade 3 ELA test takers. The matching was conducted based on the grade *g-2* scale score using the 1:1 nearest neighbor sampling method. This matched sample is labeled as MS1. Figure 1 illustrates the longitudinal and cross-sectional mapping of S1 and MS1 from 2017 to 2021. The balance in the demographic variables between the matched samples was checked following application of the matching procedure. The tables in Appendix B provide a comparison of the achievement and demographic characteristics between S1 and MS1. The tables present the sample size, the mean, standard deviation, minimum, maximum, skewness and kurtosis of the score distribution and the proportion of students classified in each demographic category. The comparison indicates that the score distribution and demographic composition of the matched samples is quite similar, and that the matching procedure was effective.

Figure 1. The Longitudinal and Cross-Sectional Mapping of S1 and MS1 from 2017-2021



Step 4. The regression coefficients obtained from Step 1 were then applied to the 2017 grade *g-2* scores in MS1 to predict their 2019 grade *g* outcomes.

Step 5. Now the fair comparison can be made between the observed grade *g* scores in 2021 scores and the estimated grade *g* scores of their matching peers in 2019.

MATCHED SAMPLE COHORT COMPARISONS

Table 2 presents the mean and standard deviation of the grade *g* scores in both cohorts, the decline in scale score points, the effect size of the decline and the percent of students met proficiency in both samples. Cohen's *d* is used as the effect size to measure the difference between the two means.

$$Cohen's\ d = \frac{\overline{x}_1 - \overline{x}_2}{s} \text{ and } s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Where $\overline{x}_1$ and $\overline{x}_2$ are the means of the two samples; and s_1 and s_2 are the standard deviations of the two samples.

Table 2. The Comparison between the Observed 2021 Performance and the Estimated 2019 Performance

Test	2019			2021			Difference (2019-2021)		
	Mean	SD	% Met Proficiency	Mean	SD	% Met Proficiency	Scale Score	Effect Size of Scale Score Difference	% Met Proficiency
G5E	2546	30	54	2535	38	45	11	0.32	9
G6E	2548	27	43	2541	33	37	7	0.23	6
G7E	2558	30	49	2549	34	37	9	0.28	12
G8E	2562	31	38	2556	36	35	6	0.18	3
G5M	3590	31	46	3572	43	32	18	0.48	14
G6M	3620	36	42	3603	44	30	17	0.42	12
G7M	3638	34	35	3627	42	30	11	0.29	5
G8M	3660	33	35	3649	40	26	11	0.30	9

In general, we see significant level of learning loss in all grades and subjects when comparing pre- and post-pandemic cohorts. Comparing the effect size between the two subjects in one grade, we can see that the decline is larger in mathematics than in ELA assessment. In both and ELA and mathematics assessments, the decline in scale scores, and associated effect sizes, indicate that the pandemic impact was more pronounced for students in lower grades. We note that the pandemic related impact on proficiency rates may be underestimated due to the shrinkage in the variance of the predicted values from the regression model. In this case, the difference between the average scale scores is a better estimate of the pandemic impact.

MATCHED SAMPLE COMPARISON OF 2021 NON-PARTICIPATING STUDENTS

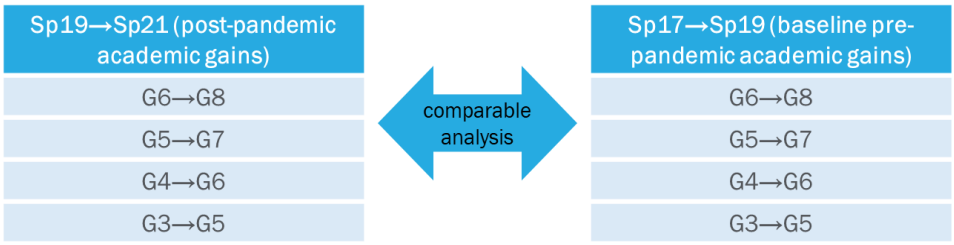
The matched sample method also allows the estimation of the performance of the students who didn't participate in the spring 2021 assessments. To estimate ability of students who did not participate in the spring 2021, we drew a matched sample (MS2) from the 2017 test records to match S2. Tables in Appendix C compare the scores and demographic characteristics between the matched samples S2 and MS2. Then we applied the regression coefficients from Step 1 to MS2 to predict their 2019 performance. Table 3 shows the predicted 2019 score of MS2, which represents expected pre-pandemic performance of those non-participating students and the comparison to the performance of the participating students. The expected average score for the non-participating students is much lower compared to the participating students in all tests.

Table 3. The Expected Performance of the Participating and Non-Participating Students in the 2021 Assessments

Test	participating			non-participating		
	Mean	SD	%proficient	Mean	SD	%proficient
G5E	2546	30	54	2538	31	44
G6E	2548	27	43	2541	27	33
G7E	2558	30	49	2551	27	33
G8E	2562	31	38	2554	30	28
G5M	3590	31	46	3581	33	35
G6M	3620	36	42	3608	37	31
G7M	3638	34	35	3628	34	25
G8M	3660	33	35	3650	33	24

COHORT REGRESSION ANALYSES

Longitudinal analyses were conducted to examine the differential gains in student academic achievement across years. The spring 2021 summative subject area test scores were regressed onto prior student area achievement and demographic variables. Since 2020 test data is not available, the spring 2021 scores were regressed to spring 2019 scores, which represents a two-year growth. A baseline of two-year growth was created to detect if there is any difference between the pre-pandemic and the post-pandemic growth. The spring 2017 to spring 2019 score gain, which represents the pre-pandemic growth, was used as the baseline for the two-year growth comparison. The graph below shows the design of the cohort comparison.



To examine if there are any differential cohort effects in the prediction of student academic growth, we combined the testing data from the two cohorts (2017-2019 and 2019-2021) into one dataset. For example, in the grade 3 to grade 5 growth model, the 2021 grade 5 and 2019 grade 5 scores were combined as the dependent variable, and the 2019 grade 3 and 2017 grade 3 scores were combined as an independent variable. A dummy variable was created to represent cohort: 1 for the records in 2019-2021 cohort; 0 for the records in 2017-2019 cohort. The grade *g* score is the dependent variable in the regression model. The grade *g-2* score is included as an independent variable and centered around the mean in the analyses so that the unstandardized intercept coefficient represents the adjusted mean of the grade *g* scores for a reference group. To compare ethnic subgroup performance, we created six dummy variables contrasting white students with each of the other ethnic groups (e.g., Hispanic vs. White, African American vs White, Hawaiian/Pacific Islander vs. White, American Indian vs. White, Multiple Race vs. White, Asian vs. White). Gender was coded 1 for female. Student with Limited English Proficiency status (LEP), students with special education status (SPED), and students with Low-Income status (Low Income) were coded as 1 to contrast with students who were not identified with those needs and were coded as 0.

In addition, the dummy coded cohort variable and the interaction between the cohort variable and each of the predictors were also included in the regression model as predictors. This cohort regression model allows us to examine whether there is any differential growth between the two cohorts and which demographic groups might

have been differentially impacted. The multiple regressions to test main effect and interaction effect of students' growth between pre-pandemic and post-pandemic is presented below.

$$Y = \beta_{00} + \beta_{10} \times \text{Previous score} + \beta_{01} \times \text{Female} + \beta_{02} \times \text{LEP} + \beta_{03} \times \text{SPED} + \beta_{04} \times \text{Low Income} + \beta_{05} \times \text{Hispanic} + \beta_{06} \times \text{African American} + \beta_{07} \times \text{Pacific Islander} + \beta_{08} \times \text{American Indian} + \beta_{09} \times \text{Multiple Ethnics} + \beta_{010} \times \text{Asian} +$$

$$\beta_{20} \times \text{Cohort} + \beta_{21} \times (\text{Cohort} \times \text{Previous score}) + \beta_{22} \times (\text{Cohort} \times \text{Female}) + \beta_{23} \times (\text{Cohort} \times \text{LEP}) + \beta_{24} \times (\text{Cohort} \times \text{SPED}) + \beta_{25} \times (\text{Cohort} \times \text{Low Income}) + \beta_{26} \times (\text{Cohort} \times \text{Hispanic}) + \beta_{27} \times (\text{Cohort} \times \text{African American}) + \beta_{28} \times (\text{Cohort} \times \text{Pacific Islander}) + \beta_{29} \times (\text{Cohort} \times \text{American Indian}) + \beta_{210} \times (\text{Cohort} \times \text{Multiple Ethnics}) + \beta_{211} \times (\text{Cohort} \times \text{Asian}) + e$$

The outcome, Y is the predicted test scores at grade *g*. The random error term, *e* is assumed to be normally distributed.

Table 4 through Table 11 show the regression coefficients estimated for each model, including standardized and unstandardized coefficients, the standard error of the unstandardized coefficient, *p* value, and partial R^2 regardless of significance level. Although many individual effects attained conventional levels of statistical significance due to large sample sizes, we focus here only on highly significant effects ($p < 0.0001$) and non-zero partial R^2 that are associated with more practically significant effect sizes and that may point to trends across grade-level and/or subject-area assessments.

The previous year variable is the most important variable in these predictive models. Its partial R^2 is the largest among all predictors, and accounts for the greatest amount of explained variation. The positive effect of previous score (β_{10}) indicates that the students with higher average levels of test scores in the previous year have higher average levels of test scores in the current year. The two-year growth across demographic subgroups is shown under the "Intercept" section. By looking at the standardized coefficient estimates and partial R^2 across growth models, results indicate that females (β_{01}) generally performed better than males for ELA across grades. Limited English proficient (LEP) students (β_{02}), special education status (SPED) students (β_{03}), and low-income status students (β_{04}) all performed less well than the general education population across grades in both ELA and mathematics. With respect to ethnicity, Hispanic students (β_{05}), African American students (β_{06}), and American Indian students (β_{08}) generally performed less well than white students across grades in both ELA and mathematics. Asian students (β_{010}) generally performed better than white students across grades in mathematics and in higher grades in ELA.

Differential growth between the pre-pandemic cohort and the post-pandemic cohort is presented under the "Cohort" section. The cohort variable is significant in all predicting models except in ELA grade 5-7 and ELA grade 6-8 models. The negative coefficient estimate (β_{20}) indicates that the post-pandemic two-year growth is smaller than the pre-pandemic two-year growth. As the grade goes up, the standardized estimate of (β_{20}) changes from -0.06 to -0.03 in ELA and from -0.10 to -0.04 in mathematics. This suggests that, when moving to higher grades, the loss in student growth in the post-pandemic cohort becomes smaller in both ELA and mathematics. The cohort models also include the interaction terms between cohort and demographic subgroups. No significant differential growth between two cohorts for any of demographic subgroups (β_{22} through β_{211}) is observed. To conclude, the cohort regression analyses show that the student learning growth is negatively impacted by the pandemic. The post-pandemic gain is significantly smaller compared to the pre-pandemic gain. The negative impact is more pronounced in lower grades both ELA and mathematics. But no evidence supports that any specific subgroup has been differentially impacted.

Table 4. Regression Coefficients for Differential Gain Across Subgroups: ELA Grade 3 to Grade 5

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	2546.38	0.17	<.0001	0.00	.
Previous score (β_{10})	0.83	0.00	<.0001	0.70	0.36
Female *vs.* Male (β_{01})	2.96	0.15	<.0001	0.04	0.01
LEP *vs*. Non-LEP (β_{02})	-13.61	0.43	<.0001	-0.10	0.09
Special Education *vs*. Non-SPED (β_{03})	-10.52	0.30	<.0001	-0.09	0.11
Low income *vs.* Non-Low Income (β_{04})	-1.90	0.16	<.0001	-0.02	0.04
Hispanic *vs.* White (β_{05})	-3.62	0.18	<.0001	-0.05	0.02
African American *vs.* White (β_{06})	-5.58	0.38	<.0001	-0.03	0.01
Pacific Islander *vs.* White (β_{07})	-4.29	1.33	0.0013	-0.01	0.00
American Indian *vs.* White (β_{08})	-10.03	0.42	<.0001	-0.05	0.02
Multiple *vs.* White (β_{09})	-1.43	0.44	0.0012	-0.01	0.00
Asian *vs.* White (β_{010})	3.58	0.53	<.0001	0.02	0.00
Cohort (β_{20})	-4.41	0.24	<.0001	-0.06	0.01
Cohort × Previous score (β_{21})	0.06	0.00	<.0001	0.03	0.00
Cohort × Female (β_{22})	-0.10	0.23	0.6578	0.00	0.00
Cohort × LEP (β_{23})	1.38	0.44	0.0017	0.01	0.00
Cohort × SPED (β_{24})	3.22	0.37	<.0001	0.02	0.00
Cohort × Low Income (β_{25})	-1.75	0.24	<.0001	-0.02	0.00
Cohort × Hispanic (β_{26})	-3.53	0.27	<.0001	-0.04	0.00
Cohort × African American (β_{27})	-1.93	0.55	0.0005	-0.01	0.00
Cohort × Pacific Islander (β_{28})	2.00	2.03	0.3246	0.00	0.00
Cohort × American Indian (β_{29})	-2.10	0.66	0.0015	-0.01	0.00
Cohort × Multiple (β_{210})	-0.98	0.60	0.1016	0.00	0.00
Cohort × Asian (β_{211})	-0.55	0.70	0.4308	0.00	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 5. Regression Coefficients for Differential Gain Across Subgroups: ELA Grade 4 to Grade 6

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	2547.82	0.14	<.0001	0.00	.
Previous score (β_{10})	0.80	0.00	<.0001	0.79	0.38
Female *vs.* Male (β_{01})	2.44	0.13	<.0001	0.04	0.01
LEP *vs*. Non-LEP (β_{02})	-9.08	0.38	<.0001	-0.07	0.07
Special Education *vs*. Non-SPED (β_{03})	-9.23	0.27	<.0001	-0.09	0.10
Low income *vs.* Non-Low Income (β_{04})	-1.77	0.14	<.0001	-0.03	0.04
Hispanic *vs.* White (β_{05})	-2.17	0.16	<.0001	-0.03	0.02
African American *vs.* White (β_{06})	-4.19	0.33	<.0001	-0.03	0.01
Pacific Islander *vs.* White (β_{07})	-1.72	1.09	0.1134	0.00	0.00
American Indian *vs.* White (β_{08})	-6.68	0.36	<.0001	-0.04	0.02
Multiple *vs.* White (β_{09})	-1.17	0.38	0.0023	-0.01	0.00
Asian *vs.* White (β_{010})	6.01	0.45	<.0001	0.03	0.00
Cohort (β_{20})	-3.07	0.20	<.0001	-0.05	0.01
Cohort × Previous score (β_{21})	-0.01	0.00	0.1551	0.00	0.00
Cohort × Female (β_{22})	-0.63	0.20	0.0012	-0.01	0.00
Cohort × LEP (β_{23})	2.17	0.39	<.0001	0.01	0.00
Cohort × SPED (β_{24})	3.14	0.32	<.0001	0.02	0.00
Cohort × Low Income (β_{25})	-1.00	0.21	<.0001	-0.01	0.00
Cohort × Hispanic (β_{26})	-1.90	0.23	<.0001	-0.02	0.00
Cohort × African American (β_{27})	-0.18	0.48	0.7029	0.00	0.00
Cohort × Pacific Islander (β_{28})	-2.63	1.73	0.1288	0.00	0.00
Cohort × American Indian (β_{29})	-1.47	0.56	0.0089	-0.01	0.00
Cohort × Multiple (β_{210})	-0.90	0.52	0.0864	0.00	0.00
Cohort × Asian (β_{211})	-0.77	0.60	0.1990	0.00	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 6. Regression Coefficients for Differential Gain Across Subgroups: ELA Grade 5 to Grade 7

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	2553.35	0.15	<.0001	0.00	.
Previous score (β_{10})	0.80	0.00	<.0001	0.82	0.38
Female *vs.* Male (β_{01})	3.86	0.14	<.0001	0.06	0.02
LEP *vs*. Non-LEP (β_{02})	-7.99	0.43	<.0001	-0.06	0.07
Special Education *vs*. Non-SPED (β_{03})	-7.62	0.30	<.0001	-0.07	0.11
Low income *vs.* Non-Low Income (β_{04})	-1.37	0.15	<.0001	-0.02	0.04
Hispanic *vs.* White (β_{05})	-1.99	0.17	<.0001	-0.03	0.02
African American *vs.* White (β_{06})	-3.07	0.35	<.0001	-0.02	0.01
Pacific Islander *vs.* White (β_{07})	0.91	1.13	0.4239	0.00	0.00
American Indian *vs.* White (β_{08})	-5.73	0.39	<.0001	-0.03	0.02
Multiple *vs.* White (β_{09})	-0.56	0.42	0.1879	0.00	0.00
Asian *vs.* White (β_{010})	8.54	0.47	<.0001	0.04	0.01
Cohort (β_{20})	-1.87	0.21	<.0001	-0.03	0.00
Cohort × Previous score (β_{21})	-0.07	0.00	<.0001	-0.05	0.00
Cohort × Female (β_{22})	-2.01	0.21	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	1.43	0.42	0.0007	0.01	0.00
Cohort × SPED (β_{24})	3.35	0.36	<.0001	0.02	0.00
Cohort × Low Income (β_{25})	-1.45	0.22	<.0001	-0.02	0.00
Cohort × Hispanic (β_{26})	-1.45	0.24	<.0001	-0.02	0.00
Cohort × African American (β_{27})	-0.21	0.51	0.6802	0.00	0.00
Cohort × Pacific Islander (β_{28})	-2.27	1.84	0.2168	0.00	0.00
Cohort × American Indian (β_{29})	-1.15	0.60	0.0549	0.00	0.00
Cohort × Multiple (β_{210})	-0.14	0.56	0.8062	0.00	0.00
Cohort × Asian (β_{211})	1.65	0.63	0.0086	0.01	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 7. Regression Coefficients for Differential Gain Across Subgroups: ELA Grade 6 to Grade 8

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	2561.02	0.15	<.0001	0.00	.
Previous score (β_{10})	0.86	0.00	<.0001	0.77	0.40
Female *vs.* Male (β_{01})	5.46	0.14	<.0001	0.08	0.02
LEP *vs*. Non-LEP (β_{02})	-9.21	0.49	<.0001	-0.06	0.07
Special Education *vs*. Non-SPED (β_{03})	-9.35	0.32	<.0001	-0.08	0.11
Low income *vs.* Non-Low Income (β_{04})	-2.20	0.16	<.0001	-0.03	0.03
Hispanic *vs.* White (β_{05})	-2.90	0.17	<.0001	-0.04	0.02
African American *vs.* White (β_{06})	-2.52	0.36	<.0001	-0.02	0.01
Pacific Islander *vs.* White (β_{07})	-0.68	1.26	0.5931	0.00	0.00
American Indian *vs.* White (β_{08})	-8.42	0.40	<.0001	-0.04	0.02
Multiple *vs.* White (β_{09})	-0.49	0.45	0.2780	0.00	0.00
Asian *vs.* White (β_{010})	5.96	0.49	<.0001	0.03	0.01
Cohort (β_{20})	-2.91	0.22	<.0001	-0.04	0.00
Cohort × Previous score (β_{21})	0.01	0.00	0.0787	0.00	0.00
Cohort × Female (β_{22})	-1.99	0.21	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	-1.10	0.47	0.0192	-0.01	0.00
Cohort × SPED (β_{24})	0.64	0.37	0.0803	0.00	0.00
Cohort × Low Income (β_{25})	0.07	0.23	0.7551	0.00	0.00
Cohort × Hispanic (β_{26})	-0.09	0.25	0.7247	0.00	0.00
Cohort × African American (β_{27})	0.84	0.51	0.1011	0.00	0.00
Cohort × Pacific Islander (β_{28})	-4.63	1.90	0.0148	0.00	0.00
Cohort × American Indian (β_{29})	0.60	0.60	0.3182	0.00	0.00
Cohort × Multiple (β_{210})	0.56	0.59	0.3417	0.00	0.00
Cohort × Asian (β_{211})	2.81	0.64	<.0001	0.01	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 8. Regression Coefficients for Differential Gain Across Subgroups: Mathematics Grade 3 to Grade 5

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	3591.60	0.20	<.0001	0.00	.
Previous score *(β_{10})*	0.73	0.00	<.0001	0.77	0.34
Female *vs.* Male (β_{01})	1.67	0.19	<.0001	0.02	0.00
LEP *vs*. Non-LEP (β_{02})	-10.17	0.45	<.0001	-0.06	0.06
Special Education *vs*. Non-SPED (β_{03})	-10.65	0.34	<.0001	-0.08	0.09
Low income *vs.* Non-Low Income (β_{04})	-3.33	0.20	<.0001	-0.04	0.04
Hispanic *vs.* White (β_{05})	-2.95	0.22	<.0001	-0.03	0.02
African American *vs.* White (β_{06})	-7.37	0.46	<.0001	-0.04	0.01
Pacific Islander *vs.* White (β_{07})	0.14	1.62	0.9325	0.00	0.00
American Indian *vs.* White (β_{08})	-8.23	0.51	<.0001	-0.04	0.02
Multiple *vs.* White (β_{09})	-0.78	0.54	0.1468	0.00	0.00
Asian *vs.* White (β_{010})	9.15	0.67	<.0001	0.03	0.01
Cohort (β_{20})	-8.89	0.29	<.0001	-0.10	0.04
Cohort × Previous score (β_{21})	0.02	0.00	<.0001	0.02	0.00
Cohort × Female (β_{22})	-2.38	0.27	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	0.94	0.53	0.0772	0.00	0.00
Cohort × SPED (β_{24})	6.14	0.45	<.0001	0.03	0.00
Cohort × Low Income (β_{25})	-2.78	0.29	<.0001	-0.02	0.00
Cohort × Hispanic (β_{26})	-9.26	0.32	<.0001	-0.09	0.00
Cohort × African American (β_{27})	-7.61	0.67	<.0001	-0.03	0.00
Cohort × Pacific Islander (β_{28})	-8.81	2.45	0.0003	-0.01	0.00
Cohort × American Indian (β_{29})	-10.18	0.78	<.0001	-0.03	0.00
Cohort × Multiple (β_{210})	-3.37	0.72	<.0001	-0.01	0.00
Cohort × Asian (β_{211})	-2.91	0.85	0.0006	-0.01	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 9. Regression Coefficients for Differential Gain Across Subgroups: Mathematics Grade 4 to Grade 6

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	3620.67	0.20	<.0001	0.00	.
Previous score *(β_{10})*	0.86	0.00	<.0001	0.86	0.38
Female *vs.* Male (β_{01})	1.04	0.18	<.0001	0.01	0.00
LEP *vs*. Non-LEP (β_{02})	-9.92	0.45	<.0001	-0.06	0.06
Special Education *vs*. Non-SPED (β_{03})	-11.47	0.35	<.0001	-0.08	0.09
Low income *vs.* Non-Low Income (β_{04})	-2.40	0.19	<.0001	-0.03	0.04
Hispanic *vs.* White (β_{05})	-4.34	0.21	<.0001	-0.05	0.03
African American *vs.* White (β_{06})	-7.27	0.45	<.0001	-0.04	0.02
Pacific Islander *vs.* White (β_{07})	-1.96	1.49	0.1872	0.00	0.00
American Indian *vs.* White (β_{08})	-8.84	0.48	<.0001	-0.04	0.02
Multiple *vs.* White (β_{09})	-3.17	0.52	<.0001	-0.01	0.00
Asian *vs.* White (β_{010})	7.20	0.65	<.0001	0.03	0.01
Cohort (β_{20})	-7.73	0.28	<.0001	-0.09	0.03
Cohort × Previous score (β_{21})	-0.04	0.00	<.0001	-0.02	0.00
Cohort × Female (β_{22})	-2.50	0.26	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	1.42	0.52	0.0067	0.01	0.00
Cohort × SPED (β_{24})	4.76	0.44	<.0001	0.02	0.00
Cohort × Low Income (β_{25})	-2.33	0.28	<.0001	-0.02	0.00
Cohort × Hispanic (β_{26})	-7.65	0.31	<.0001	-0.07	0.00
Cohort × African American (β_{27})	-5.76	0.65	<.0001	-0.02	0.00
Cohort × Pacific Islander (β_{28})	-9.72	2.35	<.0001	-0.01	0.00
Cohort × American Indian (β_{29})	-9.62	0.75	<.0001	-0.02	0.00
Cohort × Multiple (β_{210})	-1.74	0.71	0.0145	-0.01	0.00
Cohort × Asian (β_{211})	-2.48	0.81	0.0023	-0.01	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 10. Regression Coefficients for Differential Gain Across Subgroups: Mathematics Grade 5 to Grade 7

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	3643.27	0.18	<.0001	0.00	.
Previous score (β_{10})	0.81	0.00	<.0001	0.82	0.39
Female *vs.* Male (β_{01})	-1.40	0.17	<.0001	-0.02	0.00
LEP *vs*. Non-LEP (β_{02})	-13.65	0.47	<.0001	-0.08	0.07
Special Education *vs*. Non-SPED (β_{03})	-11.09	0.36	<.0001	-0.08	0.09
Low income *vs.* Non-Low Income (β_{04})	-2.85	0.18	<.0001	-0.03	0.04
Hispanic *vs.* White (β_{05})	-7.00	0.20	<.0001	-0.08	0.03
African American *vs.* White (β_{06})	-7.82	0.43	<.0001	-0.04	0.02
Pacific Islander *vs.* White (β_{07})	-2.32	1.38	0.0933	0.00	0.00
American Indian *vs.* White (β_{08})	-11.26	0.46	<.0001	-0.05	0.02
Multiple *vs.* White (β_{09})	-2.62	0.51	<.0001	-0.01	0.00
Asian *vs.* White (β_{010})	5.32	0.62	<.0001	0.02	0.01
Cohort (β_{20})	-6.33	0.26	<.0001	-0.07	0.01
Cohort × Previous score (β_{21})	0.01	0.00	0.0393	0.00	0.00
Cohort × Female (β_{22})	-2.13	0.25	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	1.47	0.51	0.0037	0.01	0.00
Cohort × SPED (β_{24})	2.72	0.43	<.0001	0.01	0.00
Cohort × Low Income (β_{25})	-1.41	0.27	<.0001	-0.01	0.00
Cohort × Hispanic (β_{26})	-2.83	0.29	<.0001	-0.03	0.00
Cohort × African American (β_{27})	-2.57	0.61	<.0001	-0.01	0.00
Cohort × Pacific Islander (β_{28})	-4.79	2.22	0.0315	0.00	0.00
Cohort × American Indian (β_{29})	-4.11	0.72	<.0001	-0.01	0.00
Cohort × Multiple (β_{210})	-1.06	0.68	0.1206	0.00	0.00
Cohort × Asian (β_{211})	0.63	0.77	0.4110	0.00	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

Table 11. Regression Coefficients for Differential Gain Across Subgroups: Mathematics Grade 6 to Grade 8

Effect	Unstandardized Coefficient	SE	*p* value	Standardized Coefficient	R^2
Intercept (β_{00})	3657.31	0.19	<.0001	0.00	.
Previous score (β_{10})	0.82	0.00	<.0001	0.88	0.43
Female *vs.* Male (β_{01})	1.01	0.18	<.0001	0.01	0.00
LEP *vs*. Non-LEP (β_{02})	-8.26	0.51	<.0001	-0.05	0.05
Special Education *vs*. Non-SPED (β_{03})	-9.00	0.36	<.0001	-0.07	0.08
Low income *vs.* Non-Low Income (β_{04})	-1.38	0.19	<.0001	-0.02	0.03
Hispanic *vs.* White (β_{05})	-2.70	0.21	<.0001	-0.03	0.03
African American *vs.* White (β_{06})	-2.20	0.43	<.0001	-0.01	0.01
Pacific Islander *vs.* White (β_{07})	1.42	1.52	0.3529	0.00	0.00
American Indian *vs.* White (β_{08})	-5.64	0.46	<.0001	-0.03	0.02
Multiple *vs.* White (β_{09})	-0.58	0.55	0.2928	0.00	0.00
Asian *vs.* White (β_{010})	5.84	0.67	<.0001	0.02	0.01
Cohort (β_{20})	-3.58	0.26	<.0001	-0.04	0.01
Cohort × Previous score (β_{21})	-0.02	0.00	<.0001	-0.01	0.00
Cohort × Female (β_{22})	-2.10	0.25	<.0001	-0.02	0.00
Cohort × LEP (β_{23})	-0.54	0.54	0.3120	0.00	0.00
Cohort × SPED (β_{24})	3.32	0.42	<.0001	0.02	0.00
Cohort × Low Income (β_{25})	-1.62	0.27	<.0001	-0.02	0.00
Cohort × Hispanic (β_{26})	-3.37	0.29	<.0001	-0.04	0.00
Cohort × African American (β_{27})	-3.50	0.60	<.0001	-0.01	0.00
Cohort × Pacific Islander (β_{28})	-5.93	2.22	0.0077	-0.01	0.00
Cohort × American Indian (β_{29})	-4.59	0.69	<.0001	-0.01	0.00
Cohort × Multiple (β_{210})	0.05	0.70	0.9443	0.00	0.00
Cohort × Asian (β_{211})	3.09	0.82	0.0002	0.01	0.00

Note: SE=Standard Error of Unstandardized Coefficient. R^2=Partial R squared. For the effect of special groups, the coefficient represents the difference compared to their contrast group; SPED = Special Education Status *vs*. Non-SPED. LEP=Limited English Proficiency *vs*. Non-LEP, Low Income = Low Income *vs.* Non-Low Income. For the effect of ethnic groups, the coefficient represents differential growth rate compared to White students

REFERENCES

Ho, A. (2021, February 26). *Three test-score metrics that all states should report in the COVID-19-affected spring of 2021.* Harvard Graduate School of Education. https://scholar.harvard.edu/ files/andrewho/files/

APPENDICES

APPENDIX A: REGRESSION MODELS USED TO PREDICT THE SPRING 2019 SCORES USING SPRING 2017 SCORES AND DEMOGRAPHIC VARIABLES

Test	Predictor	Regression Coefficients	Std. Error	t value	Pr(>|t|)
G3R-G5R	(Intercept)	404.977	6.836	59.240	0.000
	score	0.855	0.003	315.091	0.000
	GNDR	3.076	0.153	20.162	0.000
	LEP	-5.450	0.267	-20.386	0.000
	SPED	-8.311	0.255	-32.634	0.000
	LowIncome	-2.216	0.162	-13.676	0.000
	Asian	2.153	0.477	4.518	0.000
	AA	-5.522	0.371	-14.882	0.000
	Pacific	-3.084	1.312	-2.350	0.019
	Hispanic	-3.258	0.181	-17.997	0.000
	AmIndian	-9.575	0.382	-25.046	0.000
	Multi	-0.800	0.443	-1.806	0.071
G4R-G6R	(Intercept)	146.227	1.391	105.149	0.000
	score	0.796	0.003	231.187	0.000
	GNDR	6.078	0.492	12.364	0.000
	LEP	-0.577	0.996	-0.580	0.562
	SPED	-13.947	0.800	-17.428	0.000
	LowIncome	-9.332	0.564	-16.538	0.000
	Asian	9.873	2.079	4.749	0.000
	AA	-10.183	2.231	-4.564	0.000
	Pacific	-8.584	2.016	-4.258	0.000
	Hispanic	-6.465	0.893	-7.242	0.000
	AmIndian	-15.950	2.508	-6.361	0.000
	Multi	0.140	1.544	0.091	0.928
G5R-G7R	(Intercept)	598.497	6.360	94.109	0.000
	score	0.771	0.002	309.451	0.000
	GNDR	3.881	0.142	27.272	0.000
	LEP	-2.720	0.280	-9.706	0.000
	SPED	-6.062	0.236	-25.651	0.000
	LowIncome	-1.593	0.152	-10.459	0.000
	Asian	9.301	0.440	21.162	0.000
	AA	-2.961	0.344	-8.598	0.000
	Pacific	-0.310	1.190	-0.260	0.795
	Hispanic	-2.015	0.168	-12.011	0.000
	AmIndian	-5.322	0.359	-14.813	0.000

Test	Predictor	Regression Coefficients	Std. Error	t value	Pr(>|t|)
	Multi	-1.043	0.429	-2.433	0.015
G6R-G8R	(Intercept)	387.660	6.856	56.542	0.000
	score	0.854	0.003	318.792	0.000
	GNDR	5.493	0.146	37.694	0.000
	LEP	-2.992	0.324	-9.226	0.000
	SPED	-6.040	0.247	-24.455	0.000
	LowIncome	-1.832	0.156	-11.712	0.000
	Asian	5.928	0.437	13.574	0.000
	AA	-2.584	0.347	-7.441	0.000
	Pacific	-1.127	1.291	-0.873	0.383
	Hispanic	-3.013	0.171	-17.631	0.000
	AmIndian	-7.668	0.366	-20.965	0.000
	Multi	-0.076	0.461	-0.165	0.869
G3M-G5M	(Intercept)	1268.273	8.025	158.050	0.000
	score	0.659	0.002	291.793	0.000
	GNDR	1.895	0.187	10.153	0.000
	LEP	-2.044	0.325	-6.298	0.000
	SPED	-6.466	0.312	-20.751	0.000
	LowIncome	-4.107	0.198	-20.740	0.000
	Asian	11.447	0.585	19.580	0.000
	AA	-7.138	0.454	-15.710	0.000
	Pacific	0.874	1.611	0.543	0.587
	Hispanic	-3.537	0.221	-16.019	0.000
	AmIndian	-7.123	0.467	-15.249	0.000
	Multi	-0.832	0.542	-1.533	0.125
G4M-G6M	(Intercept)	947.895	8.293	114.300	0.000
	score	0.752	0.002	324.782	0.000
	GNDR	1.120	0.180	6.225	0.000
	LEP	-2.206	0.317	-6.953	0.000
	SPED	-4.991	0.296	-16.842	0.000
	LowIncome	-3.783	0.192	-19.744	0.000
	Asian	9.981	0.560	17.838	0.000
	AA	-7.346	0.437	-16.826	0.000
	Pacific	-2.421	1.531	-1.582	0.114
	Hispanic	-4.714	0.213	-22.144	0.000
	AmIndian	-7.387	0.450	-16.404	0.000
	Multi	-2.916	0.524	-5.562	0.000
G5M-G7M	(Intercept)	983.470	8.248	119.239	0.000
	score	0.741	0.002	324.731	0.000
	GNDR	-1.264	0.175	-7.235	0.000
	LEP	-4.165	0.340	-12.246	0.000
	SPED	-3.872	0.291	-13.318	0.000

Test	Predictor	Regression Coefficients	Std. Error	t value	Pr(>|t|)
	LowIncome	-3.404	0.187	-18.217	0.000
	Asian	8.698	0.549	15.830	0.000
	AA	-7.773	0.424	-18.348	0.000
	Pacific	-3.700	1.469	-2.518	0.012
	Hispanic	-7.459	0.205	-36.430	0.000
	AmIndian	-10.328	0.439	-23.541	0.000
	Multi	-2.627	0.528	-4.978	0.000
G6M-G8M	(Intercept)	1024.564	9.084	112.794	0.000
	score	0.729	0.002	291.774	0.000
	GNDR	1.052	0.183	5.754	0.000
	LEP	0.122	0.396	0.307	0.759
	SPED	-1.837	0.299	-6.140	0.000
	LowIncome	-1.593	0.194	-8.221	0.000
	Asian	9.362	0.638	14.668	0.000
	AA	-1.945	0.428	-4.539	0.000
	Pacific	-0.307	1.607	-0.191	0.849
	Hispanic	-3.154	0.213	-14.826	0.000
	AmIndian	-4.512	0.442	-10.218	0.000
	Multi	-0.599	0.585	-1.024	0.306

APPENDIX B: COMPARISON OF THE MATCHED SAMPLES—STUDENTS WHO PARTICIPATED BOTH SPRING 2019 AND SPRING 2021 ADMINISTRATIONS VS. MATCHING PEERS IN SPRING 2017 ADMINISTRATION

Test	Variables	2019 Sample	Matched 2017 Sample
G3R-G5R	sample_size	65909	65909
	score_mean	2509.41	2509.28
	score_sd	30.92	30.92
	score_min	2395	2395
	score_max	2605	2605
	score_skewness	0.17	0.18
	score_kurtosis	2.66	2.72
	Male	0.49	0.5
	Asian	0.03	0.03
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.47	0.46
	American_Indian	0.03	0.05
	MultiRacial	0.04	0.03
	ELL	0.08	0.09
	Special_Education	0.12	0.11
	LowIncome	0.42	0.41
G4R-G6R	sample_size	68555	68555
	score_mean	2524.68	2524.26
	score_sd	32.18	32.01
	score_min	2400	2400
	score_max	2610	2610
	score_skewness	0.1	0.09
	score_kurtosis	2.7	2.67
	Male	0.49	0.49
	Asian	0.03	0.03
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.47	0.45
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.03
	ELL	0.08	0.09
	Special_Education	0.12	0.12
	LowIncome	0.43	0.42
G5R-G7R	sample_size	69151	69151

Test	Variables	2019 Sample	Matched 2017 Sample
	score_mean	2543.52	2542.98
	score_sd	36.86	35.22
	score_min	2419	2419
	score_max	2629	2629
	score_skewness	-0.11	-0.12
	score_kurtosis	2.45	2.59
	Male	0.49	0.49
	Asian	0.03	0.03
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.48	0.43
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.03
	ELL	0.09	0.07
	Special_Education	0.12	0.12
	LowIncome	0.43	0.41
G6R-G8R	sample_size	69609	69609
	score_mean	2547.02	2546.17
	score_sd	32.44	31.52
	score_min	2431	2431
	score_max	2641	2641
	score_skewness	0.19	0.09
	score_kurtosis	2.68	2.49
	Male	0.49	0.5
	Asian	0.03	0.03
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.47	0.46
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.03
	ELL	0.08	0.06
	Special_Education	0.12	0.11
	LowIncome	0.40	0.38
G3M-G5M	sample_size	67141	67141
	score_mean	3528.91	3528.98
	score_sd	43.82	44.66
	score_min	3395	3395
	score_max	3605	3605
	score_skewness	-0.26	-0.23
	score_kurtosis	2.55	2.48
	Male	0.49	0.49

Test	Variables	2019 Sample	Matched 2017 Sample
	Asian	0.03	0.02
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.47	0.47
	American_Indian	0.03	0.05
	MultiRacial	0.04	0.03
	ELL	0.08	0.09
	Special_Education	0.12	0.11
	LowIncome	0.42	0.41
G4M-G6M	sample_size	69697	69697
	score_mean	3559.21	3559.13
	score_sd	44.83	45.65
	score_min	3435	3435
	score_max	3645	3645
	score_skewness	-0.25	-0.2
	score_kurtosis	2.66	2.66
	Male	0.49	0.49
	Asian	0.03	0.03
	African_American	0.05	0.04
	Pacific	0	0
	Hispanic	0.47	0.47
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.03
	ELL	0.09	0.09
	Special_Education	0.12	0.12
	LowIncome	0.43	0.42
G5M-G7M	sample_size	70553	70553
	score_mean	3589.97	3589.86
	score_sd	42.28	41.46
	score_min	3478	3478
	score_max	3688	3688
	score_skewness	0.01	-0.03
	score_kurtosis	2.63	2.72
	Male	0.49	0.49
	Asian	0.03	0.02
	African_American	0.05	0.04
	Pacific	0	0
	Hispanic	0.48	0.47
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.03
	ELL	0.09	0.07

Test	Variables	2019 Sample	Matched 2017 Sample
	Special_Education	0.12	0.11
	LowIncome	0.43	0.40
G6M-G8M	sample_size	71011	71011
	score_mean	3619.26	3618.75
	score_sd	43.94	43.82
	score_min	3512	3512
	score_max	3722	3722
	score_skewness	0.11	0.13
	score_kurtosis	2.54	2.49
	Male	0.49	0.49
	Asian	0.03	0.03
	African_American	0.05	0.05
	Pacific	0	0
	Hispanic	0.47	0.46
	American_Indian	0.03	0.05
	MultiRacial	0.03	0.02
	ELL	0.08	0.06
	Special_Education	0.12	0.11
	LowIncome	0.4	0.39

APPENDIX C: COMPARISON OF THE MATCHED SAMPLES—STUDENTS WHO PARTICIPATED SPRING 2019 BUT MISSED SPRING 2021 ADMINISTRATION VS. MATCHING PEERS IN SPRING 2017 ADMINISTRATION

Test	Variables	2019 Sample	Matched 2017 Sample
G3R-G5R	sample_size	16873	16873
	score_mean	2499.28	2499.62
	score_sd	32.35	32.51
	score_min	2395	2395
	score_max	2605	2605
	score_skewness	0.32	0.33
	score_kurtosis	2.66	2.66
	Male	0.49	0.44
	Asian	0.03	0.03
	African_American	0.07	0.05
	Pacific	0	0
	Hispanic	0.41	0.45
	American_Indian	0.12	0.09
	MultiRacial	0.04	0.03
	ELL	0.09	0.11
	Special_Education	0.15	0.16
	LowIncome	0.41	0.43
G4R-G6R	sample_size	18221	18221
	score_mean	2516.75	2516.99
	score_sd	32.29	32.17
	score_min	2400	2400
	score_max	2610	2610
	score_skewness	0.25	0.25
	score_kurtosis	2.74	2.74
	Male	0.48	0.45
	Asian	0.02	0.03
	African_American	0.07	0.05
	Pacific	0	0
	Hispanic	0.42	0.44
	American_Indian	0.12	0.11
	MultiRacial	0.04	0.03
	ELL	0.09	0.10
	Special_Education	0.15	0.17
	LowIncome	0.41	0.43
G5R-G7R	sample_size	21012	21012
	score_mean	2534.59	2534.94

Test	Variables	2019 Sample	Matched 2017 Sample
	score_sd	38.24	38.31
	score_min	2419	2419
	score_max	2629	2629
	score_skewness	0.04	0.05
	score_kurtosis	2.31	2.3
	Male	0.49	0.48
	Asian	0.02	0.03
	African_American	0.07	0.04
	Pacific	0	0
	Hispanic	0.44	0.43
	American_Indian	0.11	0.09
	MultiRacial	0.04	0.03
	ELL	0.1	0.08
	Special_Education	0.15	0.17
	LowIncome	0.43	0.41
G6R-G8R	sample_size	20646	20646
	score_mean	2538.91	2539.21
	score_sd	32.35	32.35
	score_min	2444	2441
	score_max	2641	2641
	score_skewness	0.32	0.31
	score_kurtosis	2.71	2.7
	Male	0.5	0.49
	Asian	0.02	0.03
	African_American	0.07	0.05
	Pacific	0.01	0
	Hispanic	0.43	0.44
	American_Indian	0.11	0.09
	MultiRacial	0.04	0.03
	ELL	0.09	0.08
	Special_Education	0.14	0.15
	LowIncome	0.4	0.39
G3M-G5M	sample_size	16053	16053
	score_mean	3517.12	3517.6
	score_sd	46.48	47.16
	score_min	3395	3395
	score_max	3605	3605
	score_skewness	-0.06	-0.07
	score_kurtosis	2.37	2.35
	Male	0.49	0.44
	Asian	0.03	0.02
	African_American	0.07	0.05

Test	Variables	2019 Sample	Matched 2017 Sample
	Pacific	0	0
	Hispanic	0.4	0.41
	American_Indian	0.12	0.19
	MultiRacial	0.04	0.03
	ELL	0.08	0.10
	Special_Education	0.14	0.15
	LowIncome	0.4	0.41
G4M-G6M	sample_size	17251	17251
	score_mean	3545.8	3546.13
	score_sd	46.59	47.15
	score_min	3435	3435
	score_max	3645	3645
	score_skewness	-0.08	-0.07
	score_kurtosis	2.52	2.59
	Male	0.48	0.47
	Asian	0.02	0.02
	African_American	0.07	0.05
	Pacific	0.01	0
	Hispanic	0.41	0.42
	American_Indian	0.12	0.11
	MultiRacial	0.04	0.03
	ELL	0.09	0.11
	Special_Education	0.15	0.16
	LowIncome	0.4	0.41
G5M-G7M	sample_size	19727	19727
	score_mean	3577.99	3578.56
	score_sd	43	42.81
	score_min	3478	3478
	score_max	3688	3688
	score_skewness	0.14	0.09
	score_kurtosis	2.61	2.6
	Male	0.49	0.44
	Asian	0.02	0.02
	African_American	0.07	0.05
	Pacific	0	0
	Hispanic	0.43	0.44
	American_Indian	0.12	0.09
	MultiRacial	0.04	0.03
	ELL	0.09	0.09
	Special_Education	0.14	0.15
	LowIncome	0.43	0.43
G6M-G8M	sample_size	19381	19381

Test	Variables	2019 Sample	Matched 2017 Sample
	score_mean	3605.93	3606.7
	score_sd	43.75	42.99
	score_min	3512	3512
	score_max	3722	3722
	score_skewness	0.29	0.32
	score_kurtosis	2.65	2.66
	Male	0.5	0.49
	Asian	0.02	0.02
	African_American	0.07	0.05
	Pacific	0.01	0
	Hispanic	0.42	0.43
	American_Indian	0.11	0.09
	MultiRacial	0.04	0.03
	ELL	0.09	0.07
	Special_Education	0.14	0.14
	LowIncome	0.4	0.41