Consistency of Subgroup Differences on K-PREP from 2019 to 2021

Final Report

Prepared for: Kentucky Department of Education
Office of Assessment and Accountability
300 Sower Boulevard
Frankfort, KY 40601

Prepared under: Contract #1900004339

Authors: Justin Paulsen
Meng Fan
Emily Dickinson
Arthur Thacker

Date: June 15, 2022

Consistency of Subgroup Differences on K-PREP from 2019 to 2021

Table of Contents

Table of Contents (Continued)

List of Tables

List of Figures

Consistency of Subgroup Differences on K-PREP from 2019 to 2021

Introduction

The COVID-19 pandemic impacted the operations of Kentucky's schools across multiple school years. School buildings were closed, and the education community scrambled to quickly provide non-traditional instruction (NTI) options for students. These efforts continued through the 2019-20 academic year and well into 2020-21. Students returned to in-person instruction in 2021 in stages. Many returned to hybrid or reduced schedule versions of their previous instructional routines. Some students continued to receive only online instruction. Students' return to in-person instruction varied by district across the state. These changes to the instructional experience have likely impacted student learning across the commonwealth.

As part of supporting the Kentucky Department of Education (KDE) in designing and implementing recovery strategies to address any learning loss associated with COVID-19, the Human Resources Research Organization (HumRRO) analyzed the consistency of subgroup differences in the Kentucky Performance Rating for Educational Progress (K-PREP) assessments before and after disruptions in education related to COVID-19. Using data from the 2019 and 2021 administrations, we calculated effect size differences between groups for each year and then compared their direction and magnitude. Large differences in effect sizes would indicate that the COVID event had a differential impact on subgroup performance on K-PREP. These analyses are supplemented by computing prediction models based on prior performance and using subgroup membership as a controlling variable. If subgroup membership differentially impacts prediction from pre- to post-COVID, then we can conclude that the pandemic impacted subgroups differently.

Method

Data

Data for these analyses were provided by KDE. HumRRO received datasets including all Kentucky students with 2016-2017, 2018-19, and 2020-2021 assessment data. The datasets included each student's scale scores for every K-PREP test they attempted, as well as any additional standardized tests administered during the year. The datasets also included a variety of demographic variables including grade, gender, race, and free/reduced lunch program status.

The 2020-2021 K-PREP assessments differed from the 2018-2019 versions. First, new reading and math tests were introduced for Grade 10. Additionally, through a waiver from the U.S. Department of Education, KDE was able to suspend accountability requirements associated with K-PREP. This led to a directive from KDE leadership to develop simple, straightforward assessments limited to no more than 1 hour per subject. Thus, the number of items on each of the assessments was reduced. Across all assessments (except the new Grade 10 Reading and Math tests), test forms were designed to include fewer items than what was typical in previous administrations. On-demand writing tests included only one prompt (as opposed to two).

KDE opted to use a different reporting scale to highlight the uniqueness of the spring 2021 assessment and to dissuade inappropriate comparisons. The reporting scale ranged from 100 to 200 rather than the 100-300 scale used for K-PREP tests in 2018-2019 and previous years. For on-demand writing, a simple sum writing score was reported to students (ranging from 0 to 8).

Table 1 shows the descriptive statistics for the 2018-2019 and 2020-2021 K-PREP assessment scores. As noted above, the scales used for reporting the scores differ across the administration periods, preventing meaningful comparisons. Sample sizes differ meaningfully between comparable tests across school years, ranging from 6,000 to 10,000 more examinees in the 2018-2019 school year than the 2020-2021 school year. Interestingly, the standard deviations for the K-PREP Reading and K-PREP Science tests are close to or larger in 2020-2021 (when ranges were smaller) than in the 2018-2019 (when ranges were larger). The comparable or slightly larger 2020-2021 standard deviations occur when the test score range is smaller which suggests greater variance in student scores on the 2020-2021 assessments than on the 2018-2019 assessments.

Table 1. K-PREP Test Score Descriptive Statistics Across 2018-2019 and 2020-2021 School Years

Test	Grade	2018-2019			2020-2021		
		Mean	SD	N	Mean	SD	N
K-PREP Math	3	209.55	21.06	48,617	157.15	17.34	39,105
	4	209.58	19.22	50,733	151.13	16.37	42,110
	5	211.26	19.93	51,578	150.69	17.77	42,330
	6	209.75	18.41	51,635	149.77	16.57	42,864
	7	210.23	18.06	49,970	147.30	16.31	43,078
	8	208.84	18.29	49,278	147.96	17.31	42,891
	10	NA	NA	NA	139.49	13.59	38,927
K-PREP Reading	3	210.20	18.25	48,449	152.31	18.00	39,103
	4	210.46	15.47	50,568	158.46	14.74	42,095
	5	212.23	15.73	51,398	157.91	16.56	42,357
	6	212.66	15.39	51,486	160.90	15.28	42,896
	7	212.12	15.43	49,837	156.42	15.49	43,146
	8	214.04	15.75	49,125	160.37	16.51	42,977
	10	NA	NA	NA	154.92	15.79	39,164
K-PREP Science	4	202.91	13.03	50,705	147.19	11.72	42,008
	7	201.15	12.91	49,896	139.32	13.51	42,957
	11	202.08	14.03	44,782	142.30	16.02	34,422
K-PREP On-Demand Writing	4	223.38	37.65	51,366	4.32	1.61	42,166
	8	219.85	33.81	49,036	4.66	1.80	42,656
	11	238.12	37.39	44,568	5.02	1.89	33,985

Note. NA= Not assessed in 2018-2019.

Methods

We conducted two different analyses to evaluate the consistency of subgroup differences in K-PREP assessment scores between the two school years. First, we computed effect sizes for the differences between average scores for three different demographic characteristics: gender, race/ethnicity, lunch status (as a proxy for socioeconomic status), and Individualized Education Program (IEP) enrollment. Each effect size was computed using the Cohen's d formula (Lakens, 2013)

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{n_1+n_2-2}}}$$

Given the large sample sizes used in our analyses, Hedge's g, which corrects for population effect size bias, produced essentially the same estimates. Effect sizes are interpreted as the difference in means between groups in terms of standard deviations. Typical benchmarks for interpreting effect sizes suggest small effects range from $d = 0.2 - 0.5$, medium effects range from $d = 0.5 - 0.8$, and large effect sizes are $d > 0.8$ (Cohen, 1988). For our analyses, we are primarily interested in the similarity of effect size between school years. We flag effect sizes that differ by more than 10% from 2018-2019 to 2020-2021 as being meaningfully large. As a sensitivity analysis, we estimated effect sizes for 2017-2018 assessment scores, computed the percent change between 2018 and 2019, and determined whether the extent of change differed from the percent change between 2019 and 2021.

Additionally, we conducted two separate regression-based analyses to compare the relative impact of subgroup membership in 2017-2019 and 2019-2021. We first estimated the relationship between subgroup status and K-PREP assessment score after controlling for the student's previous K-PREP score (two years ago). We used the following formulas for these analyses:

$$KPREPScaleScore2019 \sim Subgroup2019 + KPREPScaleScore2017$$

$$KPREPScaleScore2021 \sim Subgroup2021 + KPREPScaleScore2019$$

We standardized each variable in the regressions to account for the differences in reporting scales. This ensures that coefficients from the two school years' regressions are comparable. This does render the coefficients uninterpretable in the context of the model – standardized coefficients are interpreted as the change in the standard deviations of the outcome associated with a standard deviation in the predictor. When the predictor is a dichotomous variable (e.g., female-male), a standard deviation change is not meaningful. Thus, interpretations of the regression results should focus only on the degree to which the standardized coefficients differ between years.

We also conducted regressions to understand the impact of the subgroup on current year assessment scores beyond previous school year scores. We found the R-squared for the regressions below (comparable regressions were conducted for 2020-2021 school year scores):

$$KPREPScaleScore2019 \sim KPREPScaleScore2017$$

$$KPREPScaleScore2019 \sim Subgroup2019 + KPREPScaleScore2017$$

Comparing the R-squared between the first and second models indicates the degree to which the subgroup status provides predictive power beyond the predictive power of the previous school year scores. We then compare the degree of change in R-squared across years to determine whether subgroup status has a differential impact based on school year. A difference by school year would suggest that subgroup status has a greater or lesser influence on the relationship across the years.

To ensure data quality, we removed any cases where the gender or race/ethnicity subgroup status changed between school years. Because the models depend on data from a prior year,

we could only estimate the subgroup effect using K-PREP Math and K-PREP Reading for Grades 5 through 8 (Grades 3 and 4 are not possible because those cohorts were in Grades 1 and 2 in 2018-2019).

Results

Effect Size Comparisons

Below, we present the effect sizes by grade by year for each of the assessments, considering gender, race and ethnicity, lunch status, and IEP enrollment in turn. The figures below present the effect sizes for each assessment-grade combination by year with dotted lines to indicate minimal (between the yellow lines), small (between the yellow and green lines), medium (between the green and blue lines), and large (above the blue lines) effect size ranges. Effect sizes that differ by at least 10% between years are considered meaningfully different effects. Appendix A provides tables with the sample size (N), mean, and standard deviation for each grade-subgroup effect size comparison within each content area assessment.

Gender

Figure 1 plots the effect sizes by grade by year for each of the assessments. Effect sizes greater than zero indicate females on average scored higher than males. Overall, the effect size differences between males and females tend to be small across math, reading, science, and on-demand writing. The effect size differences tended closer toward parity in the 2020-2021 assessments as demonstrated by bars closer to zero as seen in all of the grades for reading, Grades 6 and 8 in math, Grades 7 and 11 in science, and Grades 5 and 8 in on-demand writing. Other grade-test combinations saw an increase in disparity between groups. While the raw changes in effect size are not large, the scale of these changes from year to year exceeds 10% in every instance except for Grade 11 science and on-demand writing. The change in effect sizes between 2019 and 2021 assessment scores exceeds the change observed between 2018 and 2019 except for on-demand science which saw similar variability in Grades 4 and 7.
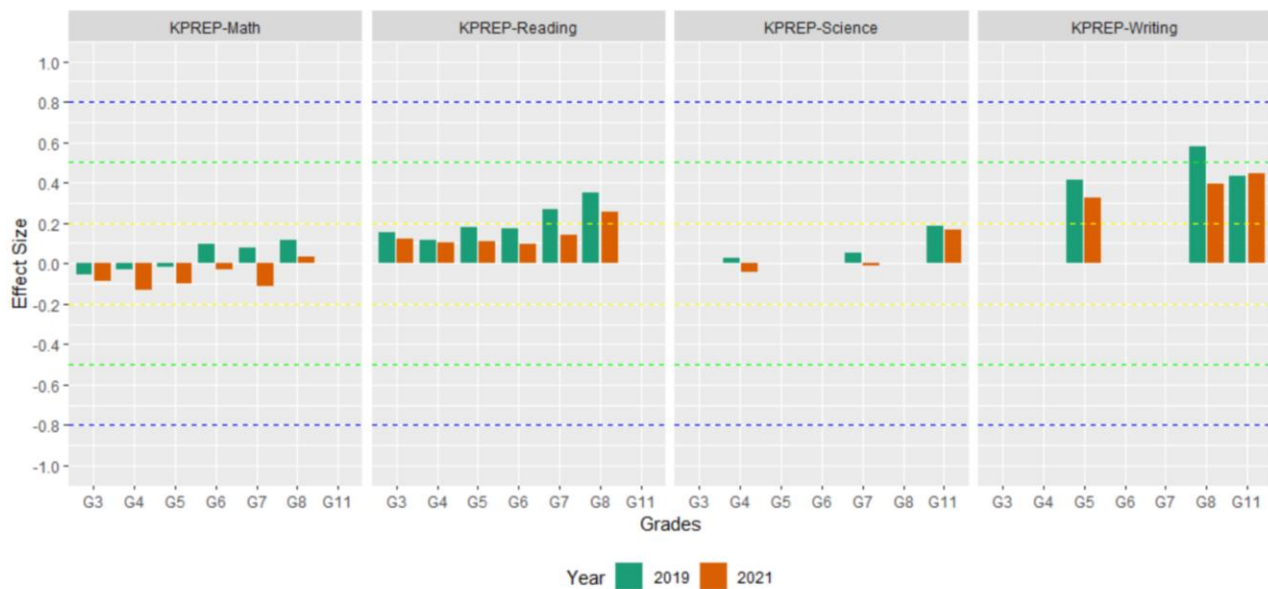

Figure 1. Differences in Test Scores by Gender and Year

Race and Ethnicity

The following six figures present the effect sizes for differences in mean assessment scores between White students and each of the following subgroups: African American, Asian, Hispanic, American Indian, multiracial, and Native Hawaiian, and Pacific Islander (NHPI) students. White is used as the base group in comparisons to highlight how the minority racial/ethnic subgroups differ from the majority subgroup.

Figure 2 below illustrates the effect sizes comparing White and Asian students. Positive effect sizes indicate a higher average test score for Asian than White students. The differences between the student subgroups were mostly small in reading, science, and on-demand writing, and mostly medium in math. For most assessment-grade combinations, the differences in effect sizes between 2019 and 2021 were not meaningful and in inconsistent directions – both increasing disparity (like Grades 6-7 math and Grades 6-8 reading) and decreasing disparity (like Grade 4 math and reading). The Grade 4 math effect size decreased by half from 0.50 in 2019 to 0.25 in 2021. The Grade 11 on-demand writing effect size increased from 0.07 to 0.33. The rate of change between years exceeds 10% for all of the reading and on-demand writing assessment grades and most of the math (3, 4, 6, 7) and science grades (4 and 11). These changes suggest meaningful differences between the two school years, although similar grades in science saw greater than 10% changes across the 2018 and 2019 assessment effect sizes.
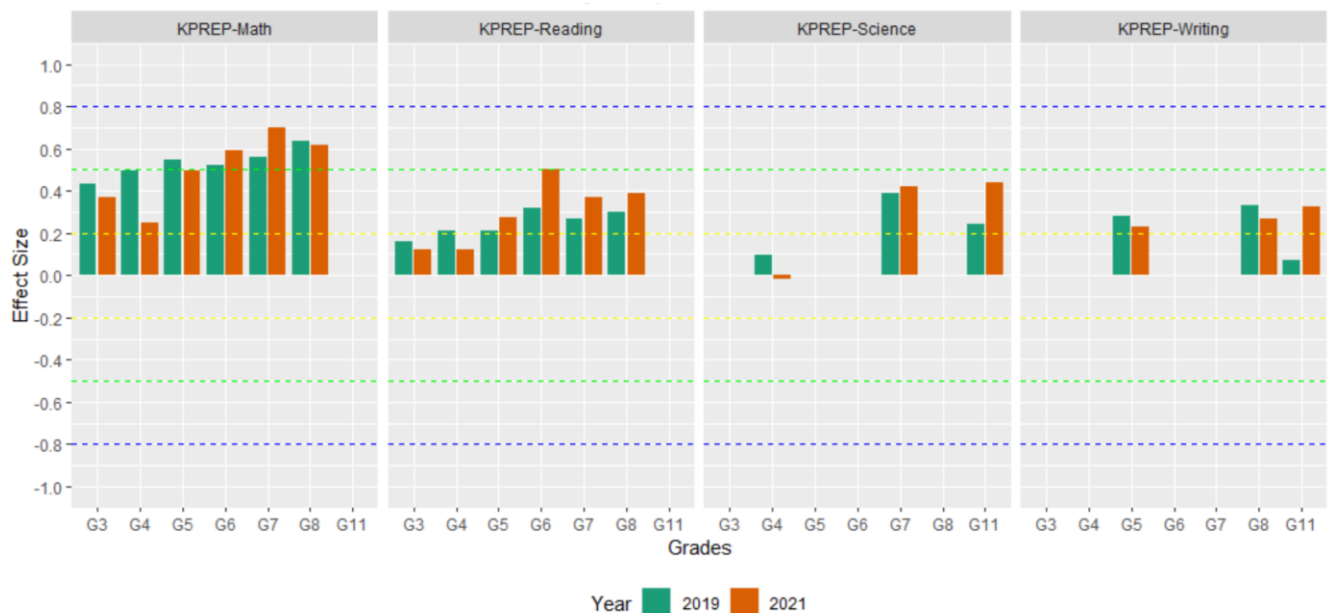
Figure 2. Differences in Test Scores by Race-Ethnicity Group and Year: Asian – White

Figure 3 illustrates effect sizes for differences in average scores between African American and White students, where negative effect sizes indicate lower average test scores among African American students. Across all the assessment-grade combinations, effect sizes tend to be in the medium range, with lower average scores for African American students. In all grades for the reading and science tests and most of the math test grades, the 2021 effect sizes trended marginally toward greater parity. The Grades 3 and 4 reading and Grades 5 and 8 on-demand writing test effect sizes increased in 2021. With the exception of some grade-assessment

combinations (Grade 8 reading, Grades 3-5 math, Grade 11 science, and Grade 5 on-demand writing), the changes in effect sizes were at least 10%. This scale of change is much larger than between 2018 and 2019 when only one grade-assessment combination saw a change of 10% in effect size.

Figure 3. Differences in Test Scores by Race-Ethnicity Group and Year: African American – White

Figure 4 reports the effect size differences between the average test scores for Hispanic and White students. Negative effect sizes indicate that the average test score for Hispanic students is smaller than the average test score for White students. The effect sizes across tests and grades tend to be small to medium, with lower average scores among Hispanic students. Effect sizes in 2021 tended to be slightly closer to parity for all math, reading, and science comparisons. The 2021 effect sizes for on-demand writing Grades 5 and 8 were negative. The changes in effect sizes for all science and on-demand writing were greater than 10%, as were the change in effect sizes for Grades 4, 6, and 8 reading, and Grades 6-8 math. The number of effect size changes of at least 10% in 2018 and 2019 was comparable in math; the other assessments saw no or one such change.
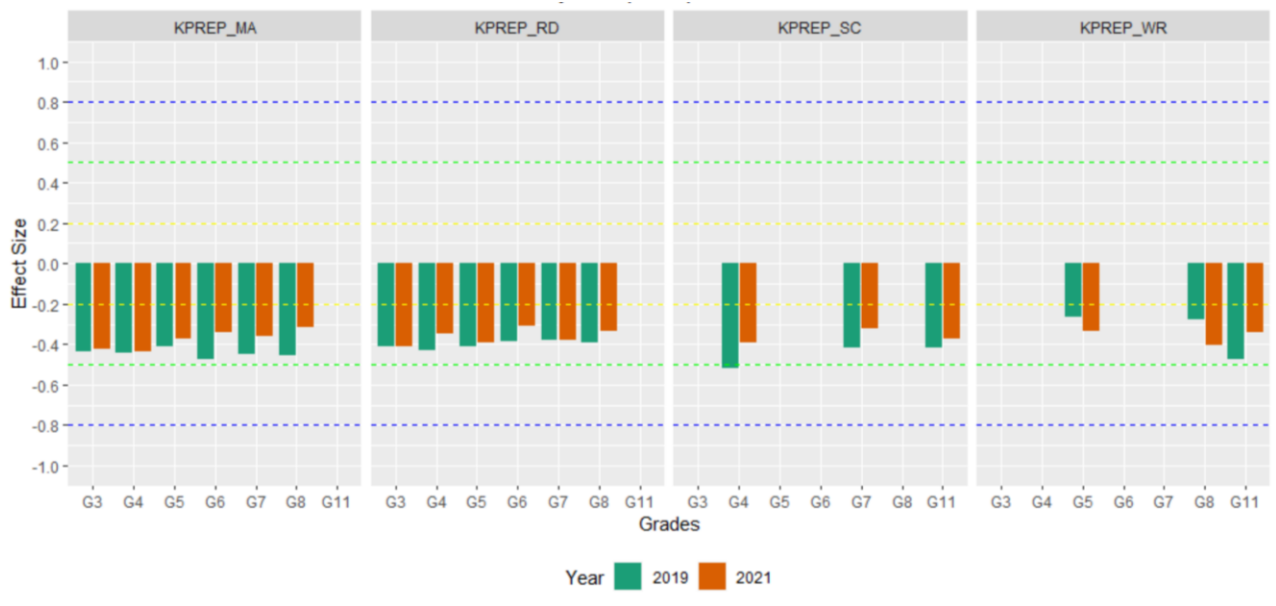
Figure 4. Differences in Test Scores by Race-Ethnicity Group and Year: Hispanic – White

Figure 5 shows the effect sizes between average test scores for American Indian and White students, where positive effect sizes indicate higher average test scores among American Indian students. Effect sizes are mostly within the minimal or small range, some showing higher averages for American Indian students and others for White students. The results show great variance from grade to grade within and across assessments and years. For example, Grade 6 reading in 2021 showed a small positive effect size indicating higher averages for American Indian students, while Grade 7 reading showed a medium negative effect size indicating lower averages for American Indian students compared to White students. This variability is likely related to the small sample sizes of American Indian students, typically between 40 and 60 students in each grade. Individual student outcomes have an outsized influence with small samples. More assessment-grade combinations saw 2021 effect sizes above the comparable 2019 effect sizes, all of which were greater than 10% changes year on year. This is similar to results in 2018 and 2019 when all but one effect size change across the years were at least 10%.

Figure 5. Differences in Test Scores by Race-Ethnicity Group and Year: American Indian – White

Figure 6 presents the effect sizes by year in average score differences between multiracial and White students by grade and content area assessment. Negative effect sizes indicate that the average test score among multiracial students is lower than that for White students. Across all tests and grades, effect sizes for both years fall in the minimal or small and negative, indicating somewhat lower average test scores among multiracial students. Effect sizes in 2021 tended to be slightly closer to parity than in 2019 in most grades and assessments. Most of the changes in effect size between years exceeded 10% (except Grades 4 and 5 reading and math, and Grade 8 on-demand writing). The number of instances of assessment-grade combinations with changes in effect size in 2018 and 2019 was the same for reading and math and less in science and on-demand writing.

Figure 6. Differences in Test Scores by Race-Ethnicity Group and Year: Multiracial – White

Figure 7 illustrates the effect sizes documenting differences in average scores between NHPI students relative to White students across assessments and grades in 2019 and 2021. Negative effect sizes mean that NHPI students had lower average assessment scores than White students. In general, effect sizes fell in the small and medium negative range in both 2019 and 2021. Most assessment-grade combinations' effect sizes across years were similar, some slightly higher in 2019 (e.g., Grades 4 and 5 math) and some slightly higher in 2021 (e.g., Grades 6 and 7 reading). Except for a handful of assessment-grade combinations (Grade 4 reading, Grade 3 math, and Grade 8 on-demand writing), the difference from 2019 to 2021 effect sizes exceeded a 10% change. This scale of changes of at least 10% was observed for all test domains except science between 2018 and 2019. This may be related to the variability in test score averages resulting from small sample sizes.
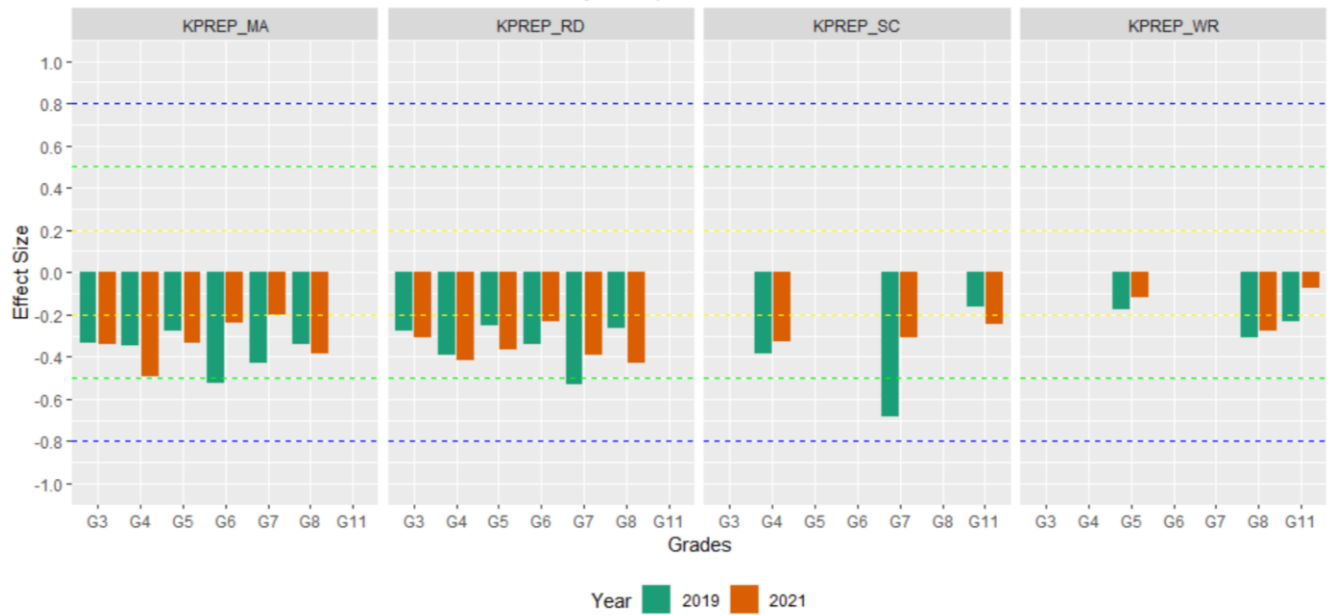
Figure 7. Differences in Test Scores by Race-Ethnicity Group and Year: Native Hawaiian / Pacific Islander – White

Lunch Status

Figure 8 below presents the differences in average test scores between students who qualify for free lunch compared to students who qualify for paid lunch based on family income. The effect size differences between free and paid lunch status students all fell within the medium effect size range. The 2021 effect sizes for the free-paid status comparisons were slightly closer to parity in some instances: math in Grades 6-8, reading in Grades 4-8, all grades in science, and Grade 11 in on-demand writing. Other assessment-grade combinations saw slight increases in effect sizes in 2021 relative to 2019. However, with a handful of exceptions (Grade 7 reading, Grade 3 math, Grade 11 science, and Grades 5 and 8 on-demand writing), the change in effect size between 2019 and 2021 did not exceed 10%. In comparison, none of the changes between 2018 and 2019 exceeded 10%.

Figure 8. Differences in Test Scores by Free-Paid Lunch Status and Year

Figure 9 below presents the differences in average test scores between students who qualify for reduced-priced lunch compared to students who qualify for paid lunch based on family income. The effect sizes all fell in the small effect size range. The 2021 effect sizes for the reduced-price-paid status comparisons were slightly closer to parity in math Grades 5-8, in reading in Grades 4-8, in Grades 7 and 11 in science, and in Grades 5 and 11 in on-demand writing. Other instances saw slight effect size increases. The change in effect size between 2019 and 2021 exceeded 10% except for Grade 4 and 8 reading, Grades 5-7 math, and Grade 5 on-demand writing.

Figure 9. Differences in Test Scores by Reduced-Paid Lunch Status and Year

IEP Status

Figure 10 below demonstrates the differences in average test scores between IEP students compared to their non-IEP peers. Negative effect sizes indicate that the average test score among non-IEP students is higher than students with IEPs. The 2019 effect sizes tend to be lower for most of the assessment-grade combinations except for Grade 11 on-demand writing. The scale of these changes from year to year exceeds 10% in most instances. However, in several exceptions (Grade 11 science, and Grades 5, 8, and 11 on-demand writing), the change in effect size between 2019 and 2021 did not exceed 10%.
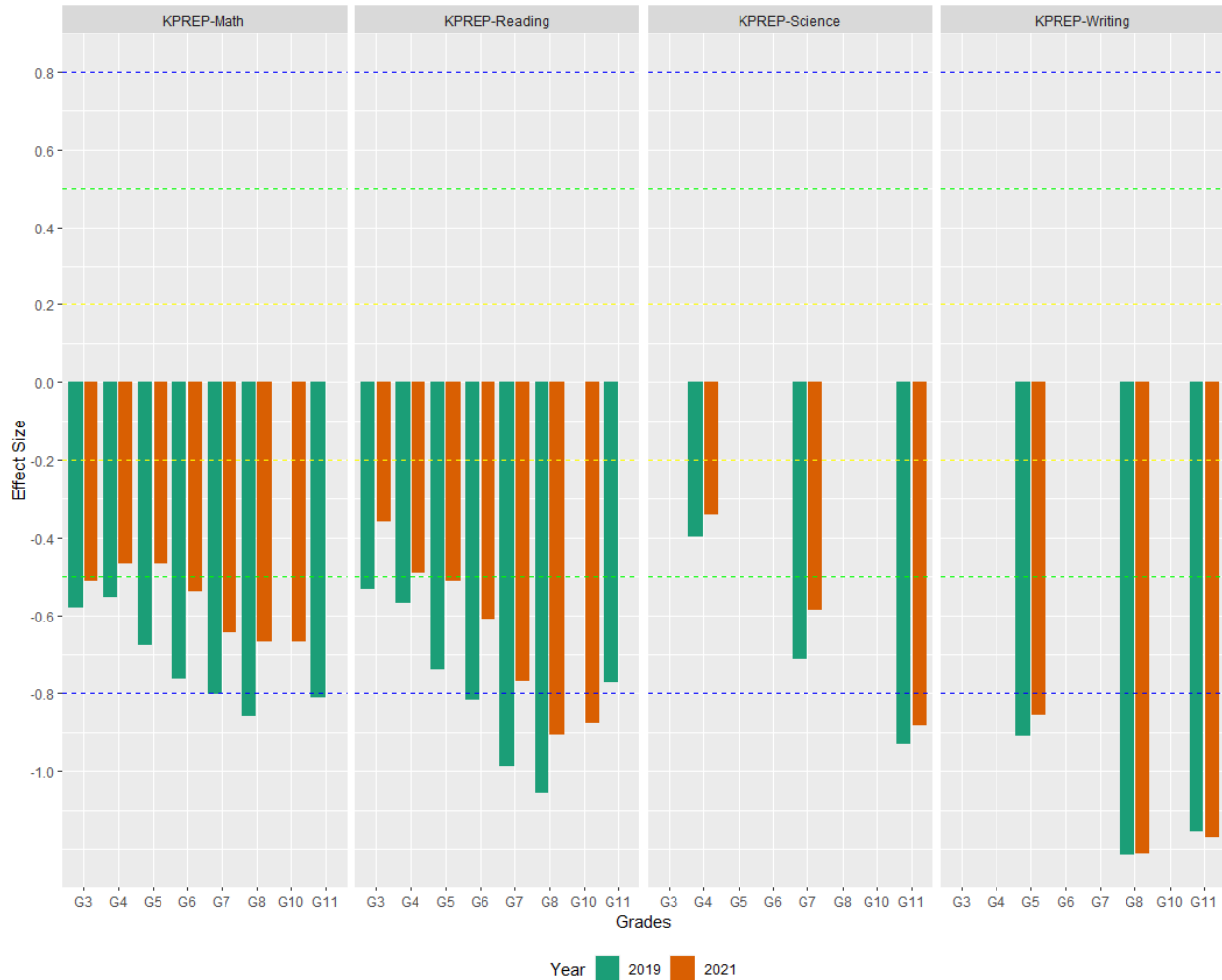
Figure 10. Differences in Test Scores by IEP Status and Year

Regression Analyses

Comparing R-squared Across Prediction Models

As noted above, we estimated regression relationships between current year assessment scores and scores of two years ago without and with group membership for 2017-2019 and 2019-2021. The R-squared values for those models are reported in the tables below. Comparing the R-squared values between the years provides an indication of whether subgroup membership had a consistent effect across years.

Gender

Tables 2 and 3 present the changes in R-squared in models predicting reading and math assessment scores without and with the inclusion of gender. The changes in R-squared with the inclusion of gender to a model containing previous year assessment scores remained below 1% across all years, all subjects, and all grades except Grade 8 reading in 2017-2019. Some

assessment-grade combinations showed increases in the predictive value of including gender (Grades 5, 7, and 8 math), and others showed decreases (Grades 5-8 reading, Grade 6 math).

Table 2. R-squared Changes in Regression Models with Inclusion of Gender – Reading

Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Gender	Change in R-squared
5	2017-2019	0.4484	0.4496	0.26%
5	2019-2021	0.3851	0.3852	0.03%
6	2017-2019	0.4960	0.4991	0.62%
6	2019-2021	0.4211	0.4213	0.05%
7	2017-2019	0.5072	0.5113	0.82%
7	2019-2021	0.4182	0.4184	0.04%
8	2017-2019	0.5157	0.5260	1.99%
8	2019-2021	0.4952	0.4990	0.77%

Table 3. R-squared Changes in Regression Models with Inclusion of Gender – Math

Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Gender	Change in R-squared
5	2017-2019	0.5239	0.5239	0.01%
5	2019-2021	0.4227	0.4236	0.23%
6	2017-2019	0.5265	0.5309	0.82%
6	2019-2021	0.4385	0.4384	-0.03%
7	2017-2019	0.5329	0.5343	0.25%
7	2019-2021	0.4913	0.4941	0.56%
8	2017-2019	0.5664	0.5668	0.07%
8	2019-2021	0.4619	0.4623	0.09%

Race and Ethnicity

Tables 4 and 5 present the changes in R-squared associated without and with the inclusion of each race or ethnic group in predicting reading and math outcomes. The changes in R-squared with the inclusion of race to a model containing the previous year assessment score remained below 1% across most year, subject, and grade combinations The exceptions include African American Grade 5 in reading 2017-2019 regression models and Grades 5-7 in math 2017-2019 regression models. The inclusion of an indicator for Asian students had a slightly larger impact on R-squared in all the 2019-2021 regression models, for both reading and math. The inclusion of an indicator for African American students had a slightly smaller impact on R-squared in all

the 2019-2021 regression models, for both reading and math. The inclusion of an indicator for NHPI resulted in zero change in R-squared for all reading 2019-2021 regression models and grade 8 for the math assessments. The other race or ethnic groups presented a mix of slightly smaller or larger changes in R-squared in 2019-2021 relative to changes in 2017-2019.

Table 4. R-squared Changes in Regression Models with Inclusion of Race-Ethnicity – Reading

Race-Ethnicity	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Race-Ethnicity	Change in R-squared
Asian	5	2017-2019	0.5505	0.5508	0.05%
		2019-2021	0.3689	0.3701	0.32%
	6	2017-2019	0.5528	0.5538	0.18%
		2019-2021	0.4155	0.4194	0.94%
	7	2017-2019	0.5878	0.5880	0.04%
		2019-2021	0.4055	0.4071	0.40%
	8	2017-2019	0.6103	0.6107	0.07%
		2019-2021	0.4819	0.4828	0.19%
African American	5	2017-2019	0.4499	0.4546	1.05%
		2019-2021	0.3854	0.3877	0.58%
	6	2017-2019	0.4965	0.5005	0.80%
		2019-2021	0.4242	0.4245	0.07%
	7	2017-2019	0.5047	0.5092	0.88%
		2019-2021	0.4178	0.4201	0.55%
	8	2017-2019	0.5152	0.5201	0.94%
		2019-2021	0.4931	0.4945	0.28%
Hispanic	5	2017-2019	0.4291	0.4301	0.24%
		2019-2021	0.3677	0.3684	0.19%
	6	2017-2019	0.4749	0.4752	0.05%
		2019-2021	0.4117	0.4117	0.00%
	7	2017-2019	0.4892	0.4894	0.05%
		2019-2021	0.4053	0.4058	0.10%
	8	2017-2019	0.5002	0.5003	0.03%
		2019-2021	0.4807	0.4807	0.00%

Table 4. (Continued)

Race-Ethnicity	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Race-Ethnicity	Change in R-squared
American Indian	5	2017-2019	0.4275	0.4275	0.00%
		2019-2021	0.3668	0.3668	0.00%
	6	2017-2019	0.4733	0.4733	0.00%
		2019-2021	0.4141	0.4141	0.00%
	7	2017-2019	0.4876	0.4876	0.00%
		2019-2021	0.4034	0.4036	0.04%
	8	2017-2019	0.4988	0.4988	0.00%
		2019-2021	0.4775	0.4775	0.00%
Multiracial	5	2017-2019	0.4278	0.4279	0.02%
		2019-2021	0.3688	0.3689	0.03%
	6	2017-2019	0.4761	0.4764	0.06%
		2019-2021	0.4140	0.4140	0.01%
	7	2017-2019	0.4888	0.4890	0.03%
		2019-2021	0.4037	0.4038	0.02%
	8	2017-2019	0.4972	0.4973	0.01%
		2019-2021	0.4784	0.4784	0.00%
Native Hawaiian / Pacific Islander	5	2017-2019	0.4276	0.4276	0.00%
		2019-2021	0.3670	0.3670	0.00%
	6	2017-2019	0.4739	0.4739	0.00%
		2019-2021	0.4137	0.4137	0.00%
	7	2017-2019	0.4876	0.4876	0.00%
		2019-2021	0.4034	0.4034	0.00%
	8	2017-2019	0.4988	0.4988	0.00%
		2019-2021	0.4784	0.4784	0.00%

**Table 5. R-squared Changes in Regression Models with Inclusion of Race-Ethnicity –
Math**

Race-Ethnicity	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Race-Ethnicity	Change in R-squared
Asian	5	2017-2019	0.6171	0.6189	0.28%
		2019-2021	0.4226	0.4239	0.30%
	6	2017-2019	0.6064	0.6072	0.14%
		2019-2021	0.4374	0.4393	0.42%
	7	2017-2019	0.6528	0.6544	0.25%
		2019-2021	0.5004	0.5025	0.41%
	8	2017-2019	0.6546	0.6560	0.22%
		2019-2021	0.4660	0.4679	0.40%
African American	5	2017-2019	0.5202	0.5262	1.16%
		2019-2021	0.4272	0.4318	1.09%
	6	2017-2019	0.5233	0.5304	1.34%
		2019-2021	0.4398	0.4423	0.57%
	7	2017-2019	0.5289	0.5364	1.43%
		2019-2021	0.4963	0.4980	0.34%
	8	2017-2019	0.5602	0.5634	0.57%
		2019-2021	0.4634	0.4641	0.15%
Hispanic	5	2017-2019	0.5079	0.5082	0.06%
		2019-2021	0.4146	0.4150	0.08%
	6	2017-2019	0.5084	0.5093	0.17%
		2019-2021	0.4272	0.4274	0.03%
	7	2017-2019	0.5203	0.5211	0.16%
		2019-2021	0.4842	0.4845	0.06%
	8	2017-2019	0.5501	0.5504	0.05%
		2019-2021	0.4552	0.4553	0.01%

Table 5. (Continued)

Race-Ethnicity	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Race-Ethnicity	Change in R-squared
American Indian	5	2017-2019	0.5057	0.5057	0.01%
		2019-2021	0.4199	0.4199	0.01%
	6	2017-2019	0.5078	0.5079	0.00%
		2019-2021	0.4322	0.4322	0.00%
	7	2017-2019	0.5231	0.5231	0.00%
		2019-2021	0.4945	0.4945	0.00%
	8	2017-2019	0.5475	0.5475	0.00%
		2019-2021	0.4609	0.4609	0.00%
Multiracial	5	2017-2019	0.5062	0.5063	0.02%
		2019-2021	0.4187	0.4191	0.10%
	6	2017-2019	0.5111	0.5120	0.18%
		2019-2021	0.4321	0.4323	0.03%
	7	2017-2019	0.5247	0.5253	0.12%
		2019-2021	0.4945	0.4946	0.04%
	8	2017-2019	0.5477	0.5478	0.01%
		2019-2021	0.4614	0.4614	0.00%
Native Hawaiian / Pacific Islander	5	2017-2019	0.5060	0.5060	0.00%
		2019-2021	0.4198	0.4198	0.01%
	6	2017-2019	0.5080	0.5080	0.00%
		2019-2021	0.4319	0.4320	0.01%
	7	2017-2019	0.5230	0.5230	0.00%
		2019-2021	0.4941	0.4942	0.01%
	8	2017-2019	0.5473	0.5473	0.00%
		2019-2021	0.4608	0.4608	0.00%

Lunch Status

Tables 6 and 7 present the changes in R-squared for models predicting assessment scores, in reading and math, using the previous year score without and with lunch status. The relative impact of adding an indicator for students with reduced lunch status is barely above 1% in both subjects and years and for all grades. The change to R-squared with the inclusion of an indicator for free lunch status has the largest impact of all the group indicators we have

examined – the 2017-2019 regressions saw a change in R-squared from 1.64 to 3.57%; the 2019-2021 regressions saw a change in R-squared ranging from 1.65 to 4.60%. Including the group membership in the 2019-2021 models had a slightly larger effect on math than reading in most grades.

Table 6. R-squared Changes in Regression Models with Inclusion of Lunch Status – Reading

Lunch Status	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Lunch Status	Change in R-squared
Reduced	5	2017-2019	0.4621	0.4630	0.18%
		2019-2021	0.3835	0.3836	0.03%
	6	2017-2019	0.5199	0.5207	0.17%
		2019-2021	0.4387	0.4391	0.10%
	7	2017-2019	0.5279	0.5286	0.13%
		2019-2021	0.4278	0.4280	0.03%
	8	2017-2019	0.5392	0.5396	0.08%
		2019-2021	0.4914	0.4915	0.01%
Free	5	2017-2019	0.4495	0.4639	3.20%
		2019-2021	0.3863	0.4001	3.57%
	6	2017-2019	0.4957	0.5053	1.93%
		2019-2021	0.4211	0.4324	2.68%
	7	2017-2019	0.5066	0.5187	2.39%
		2019-2021	0.4194	0.4303	2.60%
	8	2017-2019	0.5165	0.5273	2.08%
		2019-2021	0.4958	0.5040	1.65%

Table 7. R-squared Changes in Regression Models with Inclusion of Lunch Status – Math

Lunch Status	Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + Lunch Status	Change in R-squared
Reduced	5	2017-2019	0.5620	0.5628	0.13%
		2019-2021	0.4422	0.4425	0.07%
	6	2017-2019	0.5582	0.5593	0.20%
		2019-2021	0.4585	0.4591	0.14%
	7	2017-2019	0.5855	0.5867	0.19%
		2019-2021	0.5383	0.5389	0.11%
	8	2017-2019	0.5988	0.5991	0.06%
		2019-2021	0.4946	0.4947	0.03%
Free	5	2017-2019	0.5237	0.5370	2.54%
		2019-2021	0.4230	0.4424	4.60%
	6	2017-2019	0.5256	0.5377	2.30%
		2019-2021	0.4394	0.4545	3.42%
	7	2017-2019	0.5325	0.5477	2.86%
		2019-2021	0.4918	0.5060	2.90%
	8	2017-2019	0.5679	0.5772	1.64%
		2019-2021	0.4621	0.4731	2.39%

IEP Status

Tables 8 and 9 present the changes in R-squared in models predicting reading and math assessment scores without and with the inclusion of IEP status. The changes in R-squared with the inclusion of IEP status to the model containing the previous year assessment score ranged from 0.00% to 3.32%. Most assessment-grade combinations showed increases in the predictive value of including IEP status except the Grade 5 math regression model which showed no change in R-squared value.

Table 8. R-squared Changes in Regression Models with Inclusion of IEP Status – Reading

Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + IEP	Change in R-squared
5	2017-2019	0.4484	0.4629	3.23%
	2019-2021	0.3851	0.3860	0.24%
6	2017-2019	0.4960	0.5092	2.64%
	2019-2021	0.4211	0.4235	0.57%
7	2017-2019	0.5072	0.5251	3.53%
	2019-2021	0.4182	0.4255	1.74%
8	2017-2019	0.5157	0.5327	3.28%
	2019-2021	0.4952	0.5033	1.64%

Table 9. R-squared Changes in Regression Models with Inclusion of IEP Status – Math

Grade	Regression Years	R-squared – Previous Year Score	R-squared – Previous Year Score + IEP	Change in R-squared
5	2017-2019	0.5239	0.5286	0.90%
	2019-2021	0.4227	0.4227	0.00%
6	2017-2019	0.5265	0.5340	1.42%
	2019-2021	0.4385	0.4393	0.17%
7	2017-2019	0.5329	0.5447	2.21%
	2019-2021	0.4913	0.4929	0.32%
8	2017-2019	0.5664	0.5730	1.17%
	2019-2021	0.4619	0.4631	0.26%

Comparing Group Membership Standardized Regression Coefficients

Regression analyses also produced standardized coefficients indicating the magnitude of the relationship between subgroup membership and assessment score. The results found that in most assessment-grade combinations, neither gender nor racial group differ across the school year results. Below, we do present figures illustrating differences across years across lunch status and IEP status groups. The figure plots the standardized coefficients, with the associated 95% confidence interval, for the given subgroup effect by year, after controlling for the previous year test performance. Confidence intervals across years that intersect indicate a subgroup effect that is similar across years. Confidence intervals that do not intersect indicate a subgroup effect that differs across years. These analyses were only possible for those assessments and grades that had previous school year assessment results (reading and math, Grades 5 through 8). Appendix B provides tables with the full regression model summaries.

Lunch Status

Figure 11 illustrates the standardized coefficients by year by lunch status for each assessment-grade combination, after controlling for previous year assessment score. Negative coefficients indicate that qualifying for free or reduced-price lunch status is associated with a lower predicted test score relative to qualifying for paid lunch status. With the exception of the Grade 5 math coefficients, the predicted coefficients for the reduced-price status relative to the paid lunch status overlap, suggesting the same relationship across school years. On the other hand, the coefficients for the 2021 relationships between free lunch status and test scores is uniformly lower than the 2019 coefficients. This indicates that being in the free lunch status, relative to the paid lunch status, is associated with a statistically significant larger decrease in test scores in 2021 than in 2019.
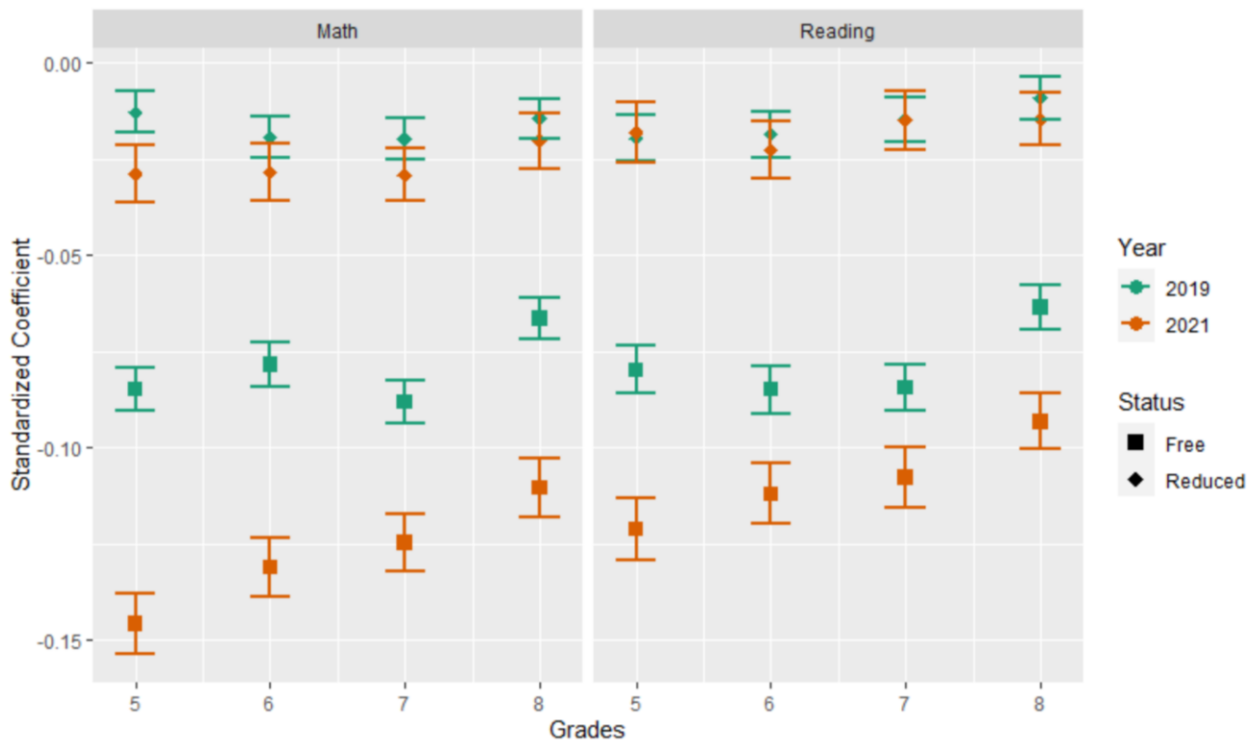
Figure 11. Differences in Standardized Coefficients by Lunch Status and Year

IEP Status

Figure 12 illustrates the standardized coefficients by year by IEP status for each assessment-grade combination, after controlling for the previous year assessment score. Negative coefficients indicate that IEP status is associated with a lower predicted test score. The standardized coefficients for 2021 relationships for both reading and math regression models are higher than the coefficients for 2019 relationships. This indicates that having an IEP is associated with a significantly smaller decrease in test scores in 2021 than in 2019.
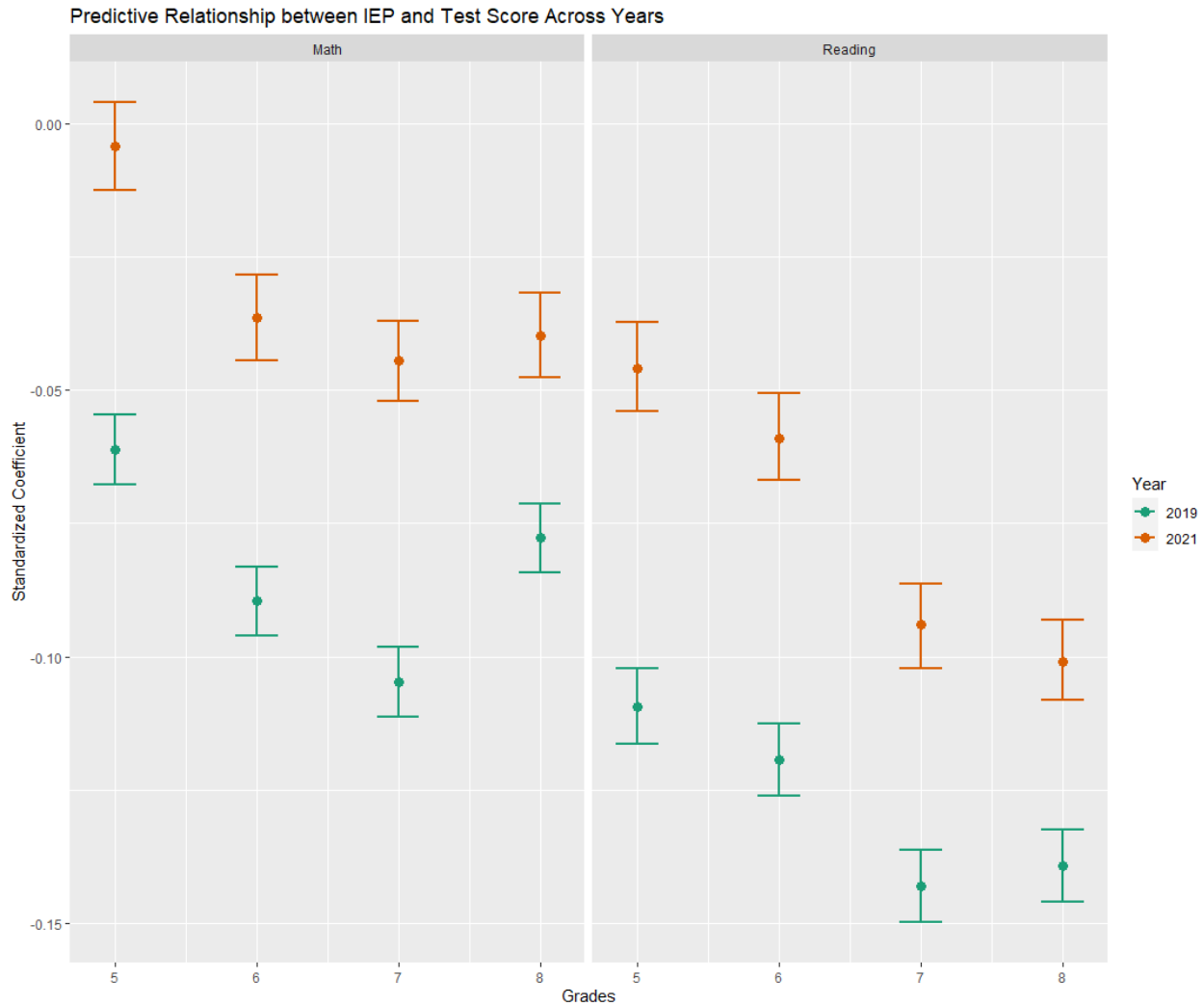
Figure 12. Differences in Standardized Coefficients by IEP Status and Year

Conclusion

To document potential learning loss associated with subgroups' educational experience during the COVID-19 disruption to schooling, we analyzed the consistency of subgroup differences in K-PREP assessment scores between 2019 and 2021. For each year assessment-grade combination, we calculated effect sizes by subgroup. Effect sizes that changed by 10% or more from 2019 to 2021 were considered substantively meaningful; other differences were considered inconsequential. Additionally, we conducted regression analyses to determine whether the impact of subgroup membership was statistically different across years after controlling for the previous year performance (two years prior) on K-PREP assessments. We did this in two ways: (a) examining the change in R-squared with the inclusion of group membership to a model regressing current year assessment scores on previous year assessment scores, and (b) comparing standardized coefficients for subgroup membership from the 2017-2019 and 2019-2021 regressions predicting current year assessment score from previous year score and subgroup membership.

Between males and females, effect sizes changed more than 10% in every grade for math and reading assessments, some toward greater parity (all reading tests saw a smaller gap between average female and male scores), others toward greater difference (average effect sizes for grades 3-5 math favored males more in 2021 than in 2019). Among race and ethnicity subgroups, changes in effect sizes between 2019 and 2021 exceeded 10% and were more variable than in the past for Asian (Grades 3-8 reading, Grades 3-4 and 6-7 Math, Grades 4 and 11 science, all grades on-demand writing), African American (Grades 3-7 reading, Grades 6-8 Math, Grades 4 and 7 science, Grades 8 and 11 on-demand writing), and Hispanic students (Grades 4, 6, and 8 reading, Grades 6-8 Math, all grades on science and on-demand writing) relative to White students. Most of the changes in effect sizes were toward greater parity in the African American-White and Hispanic-White comparisons; there were a mix of effect size changes toward more and less parity for Asian-White subgroup comparisons. Other student subgroups (American Indian, multiracial, and NHPI) experienced greater than 10% change in effect sizes between years, but this level of variability was previously witnessed between 2018 and 2019 suggesting these subgroup data may be more variable. Among lunch status subgroups, there were some changes in effect size exceeding 10% between years for free compared to paid lunch (Grade 7 math, Grade 3 reading, Grade 11 science, Grades 5 and 8 on-demand writing) and reduced compared to paid lunch (Grades 3 and 5-7 math, Grades 3-4 and 8 reading, all grades science, Grades 8 and 11 on-demand writing). Between IEP and non-IEP students, the 2021 effect sizes tend to be lower for most of the assessment-grade combinations except for Grade 11 on-demand writing. Also, effect size changes from year to year exceeded 10% in most instances. However, with several exceptions (Grad 11 science, and Grades 5, 8, and 11 on-demand writing), the change in effect size between 2019 and 2021 did not exceed 10%.

The two different regression analyses both indicate that membership in the free lunch status group had a more negative impact in 2021 than in 2019. The change in R-squared analyses indicated that membership in the free lunch subgroup explained 1.65-4.60% additional variation in the outcomes. While small, it is much larger than any other change in R-squared observed and does suggest meaningful variation explained. Additionally, the standardized coefficient regression analyses suggest some consistent differences in IEP status in the 2021 and 2019 results. Figure 12, for example, shows that in 2021, IEP status had a significantly smaller negative impact on K-PREP Math and K-PREP Reading scores in all grades compared to 2019.

Overall, the results suggest that subgroup assessment results did vary from what was observed in previous years, though not in a consistent direction. Although all students on average demonstrated declines in 2021 (see Thacker, Johnston-Fisher, Jorgensen, & Dickinson, 2022), these declines did not happen so consistently as to maintain score gaps between student subgroups; some increased whereas others decreased. Notably larger subgroup effects may warrant further investigation. Specifically, regression analyses suggest that students from the free lunch group, on average, may have been more negatively impacted by disruptions to in-person instruction than their paid lunch group peers. In contrast, regression analyses indicate that disruptions to in-person instruction may have reduced the achievement gap between IEP and non-IEP students.

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863.

Thacker, A. A., Johnston-Fisher, J., Jorgensen, A. G., and Dickinson, E. R. (2022). Examining Kentucky Students' Performance Trendlines Pre- and Post-COVID-19 Interruption. (HumRRO Report). Alexandria, VA: Human Resources Research Organization.

Appendix A: Effect Size Data

Table A1. Descriptive Statistics Underlying Gender Effect Size Comparisons

Test	Grade	Gender	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Math	3	F	23,514	208.94	20.42	18,835	156.36	17.03
		M	25,100	210.12	21.62	20,251	157.88	17.59
	4	F	24,867	209.28	18.47	20,472	150.05	16.01
		M	25,862	209.87	19.91	21,614	152.16	16.64
	5	F	25,321	211.11	19.13	20,392	149.78	17.06
		M	26,255	211.40	20.68	21,926	151.55	18.37
	6	F	25,377	210.68	18.06	21,021	149.49	16.61
		M	26,256	208.85	18.70	21,826	150.03	16.52
	7	F	24,237	210.94	17.53	20,972	146.33	15.86
		M	25,733	209.56	18.52	22,096	148.22	16.68
	8	F	24,045	209.95	17.76	20,867	148.23	17.10
		M	25,232	207.79	18.73	22,017	147.71	17.50
K-PREP Reading	3	F	23,441	211.64	18.03	18,837	153.43	18.11
		M	25,005	208.86	18.36	20,247	151.26	17.84
	4	F	24,803	211.38	15.31	20,465	159.22	14.86
		M	25,761	209.57	15.56	21,606	157.73	14.59
	5	F	25,236	213.64	15.39	20,406	158.88	16.57
		M	26,160	210.88	15.94	21,939	157.02	16.50
	6	F	25,319	214.03	15.12	21,042	161.63	15.16
		M	26,165	211.34	15.54	21,837	160.18	15.36
	7	F	24,187	214.20	15.07	21,015	157.56	15.34
		M	25,650	210.15	15.51	22,121	155.34	15.55
	8	F	23,964	216.82	15.25	20,912	162.50	15.95
		M	25,160	211.38	15.75	22,058	158.35	16.77
K-PREP Science	4	F	24,859	203.09	12.62	20,414	146.95	11.41
		M	25,842	202.74	13.42	21,571	147.41	12.00
	7	F	24,213	201.52	12.56	20,914	139.26	13.27
		M	25,682	200.81	13.22	22,033	139.37	13.73
	11	F	22,042	203.38	13.50	16,689	143.67	15.17
		M	22,734	200.82	14.41	17,723	141.00	16.68
K-PREP On-Demand Writing	5	F	25,223	231.19	36.12	20,316	4.59	1.60
		M	26,141	215.85	37.55	21,838	4.07	1.58
	8	F	23,927	229.48	30.35	20,769	5.02	1.73
		M	25,108	210.69	34.38	21,880	4.32	1.79
	11	F	21,966	246.14	33.89	16,489	5.45	1.75
		M	22,596	230.33	38.95	17,486	4.62	1.93

Note: F stands for female, M stands for male.

Table A2. Descriptive Statistics Underlying Race and Ethnicity Effect Size Comparisons

Test	Grade	R-E	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Math	3	Asian	910	220.59	27.24	836	165.14	19.36
		AfAm	5,137	198.17	19.13	3,334	146.39	15.30
		Hisp	3,831	202.76	22.28	3,058	151.63	16.07
		IndAm	53	213.00	19.84	45	162.13	16.91
		Mult	2,336	207.80	19.61	2,000	155.07	16.40
		NHPI	67	204.90	19.32	56	153.05	18.91
		White	36,280	211.72	20.35	29,756	158.83	17.07
	4	Asian	852	221.07	27.02	796	156.97	18.80
		AfAm	5,399	198.77	17.04	3,916	141.45	13.41
		Hisp	3,928	203.46	19.43	3,470	145.86	15.03
		IndAm	56	210.71	22.92	52	151.35	17.80
		Mult	2,407	206.78	17.48	2,081	148.72	15.23
		NHPI	66	205.23	19.32	70	144.91	15.99
		White	38,021	211.68	18.63	31,701	152.93	16.26
	5	Asian	883	223.99	26.31	774	161.28	22.10
		AfAm	5,492	200.10	17.27	4,004	140.65	14.01
		Hisp	4,114	205.29	20.98	3,410	145.83	15.66
		IndAm	67	205.70	18.08	48	148.75	17.80
		Mult	2,319	209.28	18.55	2,060	148.61	16.59
		NHPI	80	207.99	25.52	61	146.46	15.54
		Wh	38,621	213.33	19.32	31,961	152.36	17.78
	6	Asian	877	221.25	22.46	728	160.99	20.72
		AfAm	5,559	198.82	15.64	3,909	140.99	14.07
		Hisp	3,909	203.39	19.08	3,405	145.57	14.79
		IndAm	64	209.97	22.03	50	151.30	18.26
		Mult	2,297	206.85	16.96	2,042	147.76	15.55
		NHPI	76	202.46	18.87	54	147.19	15.72
		Wh	38,849	211.88	17.88	32,660	151.13	16.49
	7	Asian	898	222.02	25.40	727	160.16	21.51
		AfAm	5,137	199.58	15.60	3,947	138.48	12.66
		Hisp	3,713	204.25	18.21	3,486	142.84	14.10
		IndAm	74	212.61	18.28	47	143.23	14.83
		Mult	2,025	207.77	16.63	1,917	145.47	15.43
		NHPI	55	204.64	22.10	66	145.42	12.21
		Wh	38,066	212.11	17.49	32,878	148.66	16.32
	8	Asian	894	221.95	24.56	673	159.99	20.42
		AfAm	5,207	198.22	15.93	4,017	139.15	14.47
		Hisp	3,379	202.55	19.73	3,284	143.88	15.62
		IndAm	54	207.54	20.52	55	145.85	19.21
		Mult	1,869	206.33	17.06	1,896	146.20	16.36
		NHPI	69	204.67	19.56	54	142.63	13.93
		Wh	37,803	210.69	17.60	32,905	149.31	17.31

Test	Grade	R-E	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Reading	3	Asian	884	215.16	19.50	829	156.17	19.36
		AfAm	5,105	199.38	17.62	3,344	142.57	16.07
		Hisp	3,726	204.97	16.90	3,037	146.71	16.60
		IndAm	53	213.09	17.59	45	156.33	16.68
		Mult	2,336	208.25	18.01	2,003	150.44	17.32
		NHPI	67	207.31	14.47	56	148.46	17.78
		Wh	36,275	212.27	17.79	29,769	153.99	17.87
	4	Asian	824	215.38	16.97	788	161.66	14.31
		AfAm	5,367	201.40	14.92	3,920	150.80	13.76
		Hisp	3,835	205.79	14.39	3,453	154.83	14.20
		IndAm	56	210.04	14.67	52	156.85	14.54
		Mult	2,406	208.53	15.41	2,081	156.46	14.66
		NHPI	65	206.29	14.22	70	153.80	14.18
		Wh	38,011	212.23	15.04	31,707	159.86	14.55
	5	Asian	853	217.16	17.76	769	163.91	16.99
		AfAm	5,470	203.40	15.26	4,003	149.42	15.36
		Hisp	4,001	207.63	14.92	3,407	153.09	15.33
		IndAm	67	209.66	14.63	48	159.69	13.54
		Mult	2,318	211.23	15.57	2,061	156.56	16.11
		NHPI	78	210.09	17.71	61	153.44	16.66
		Wh	38,609	213.92	15.30	31,996	159.44	16.40
	6	Asian	865	219.10	16.45	723	169.66	16.86
		AfAm	5,523	203.69	14.96	3,915	154.03	13.80
		Hisp	3,809	208.56	15.16	3,405	157.34	14.24
		IndAm	64	211.84	15.71	50	163.28	15.37
		Mult	2,298	210.89	14.97	2,046	159.43	15.06
		NHPI	74	209.26	15.83	54	158.46	14.10
		Wh	38,849	214.31	14.92	32,687	161.99	15.21
	7	Asian	878	217.72	17.35	724	163.39	15.79
		AfAm	5,116	203.15	15.30	3,966	148.26	15.06
		Hisp	3,624	208.01	15.18	3,486	152.00	15.05
		IndAm	74	211.54	16.52	47	149.77	14.66
		Mult	2,023	210.86	15.04	1,923	155.65	15.12
		NHPI	54	205.70	18.13	66	151.86	13.46
		Wh	38,066	213.66	14.93	32,924	157.78	15.18
	8	Asian	880	220.15	16.96	674	168.09	16.71
		AfAm	5,173	204.87	15.90	4,029	151.17	15.90
		Hisp	3,272	209.64	15.46	3,285	156.36	15.72
		IndAm	54	211.41	22.10	56	161.39	17.14
		Mult	1,869	212.25	15.00	1,902	159.22	16.18
		NHPI	67	211.57	16.73	54	154.80	16.87
		Wh	37,807	215.62	15.20	32,970	161.81	16.20

Test	Grade	R-E	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Science	4	Asian	850	205.83	18.17	795	148.15	12.09
		AfAm	5,395	194.52	13.13	3,906	140.95	10.65
		Hisp	3,924	198.18	14.83	3,466	143.86	10.75
		IndAm	56	201.20	12.40	52	147.04	12.57
		Mult	2,405	201.43	12.52	2,072	146.05	11.28
		NHPI	66	199.94	13.39	70	144.59	11.34
		Wh	38,005	204.63	12.09	31,624	148.38	11.65
	7	Asian	895	207.24	16.99	725	146.10	16.05
		AfAm	5,127	194.07	11.87	3,917	132.67	11.32
		Hisp	3,697	197.17	14.35	3,478	136.03	12.25
		IndAm	74	202.84	13.47	47	136.09	12.88
		Mult	2,021	200.18	12.06	1,910	138.39	13.09
		NHPI	56	193.93	15.22	66	136.18	10.23
		Wh	38,023	202.41	12.39	32,804	140.37	13.53
	11	Asian	913	206.81	17.15	670	150.28	18.39
		AfAm	4,628	193.93	12.13	2,669	133.16	15.01
		Hisp	2,773	197.72	14.11	2,175	137.58	15.69
		IndAm	59	200.86	13.10	48	141.67	14.76
		Mult	1,246	200.35	13.48	1,087	141.33	15.42
		NHPI	56	201.18	14.85	49	139.51	15.93
		Wh	35,096	203.44	13.73	27,714	143.40	15.71
K-PREP On-Demand Writing	5	Asian	852	237.08	36.50	765	4.83	1.62
		AfAm	5,463	204.30	37.22	3,970	3.43	1.53
		Hisp	3,996	216.92	35.72	3,385	3.94	1.56
		IndAm	67	215.61	38.78	48	4.29	1.66
		Mult	2,318	219.46	37.42	2,055	4.23	1.52
		NHPI	78	220.26	41.20	60	4.28	1.58
		Wh	38,590	226.71	36.97	31,871	4.47	1.59
	8	Asian	880	233.65	33.62	664	5.31	1.79
		AfAm	5,146	201.50	35.00	3,966	3.59	1.69
		Hisp	3,262	213.58	33.04	3,245	4.13	1.78
		IndAm	53	220.75	35.93	55	4.51	1.97
		Mult	1,861	216.65	33.16	1,887	4.50	1.77
		NHPI	67	212.63	38.18	54	4.35	1.76
		Wh	37,764	222.75	32.79	32,778	4.84	1.75
	11	Asian	906	244.41	40.51	664	5.75	1.85
		AfAm	4,592	217.63	36.68	2,603	4.08	1.83
		Hisp	2,725	224.66	37.89	2,134	4.51	1.92
		IndAm	58	232.64	38.57	47	5.09	1.87
		Mult	1,242	233.75	37.13	1,066	4.84	1.89
		NHPI	56	233.36	34.71	50	5.00	1.80
		Wh	34,978	241.87	36.22	27,411	5.14	1.86

Note: AfAm stands for African American, Hisp stands for Hispanic, IndAm stands for Indian American and Native Alaskan, Mult stands for multiracial, NHPI stands for Native Hawaiian and Pacific Islander, Wh stands for White.

Table A3. Descriptive Statistics Underlying Lunch Status Effect Size Comparisons

Test	Grade	Lunch Status	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Math	3	F	28,973	204.54	19.37	22,656	152.32	16.01
		R	1,995	211.61	20.21	1,276	158.13	16.73
		P	17,648	217.53	21.30	15,173	164.27	16.82
	4	F	30,093	204.85	17.28	24,023	146.62	14.57
		R	2,056	211.16	18.59	1,479	151.31	15.47
		P	18,584	217.06	19.85	16,608	157.65	16.71
	5	F	30,309	206.26	18.09	24,279	145.93	15.56
		R	2,021	212.42	18.77	1,472	151.84	17.43
		P	19,247	219.01	20.31	16,579	157.56	18.56
	6	F	29,802	204.99	16.65	24,207	145.32	14.86
		R	2,147	210.28	17.59	1,446	150.03	15.96
		P	19,686	216.90	18.72	17,211	156.01	16.87
	7	F	28,046	205.11	16.09	23,916	142.68	13.85
		R	2,085	210.55	17.23	1,345	147.37	15.57
		P	19,839	217.44	18.32	17,817	153.49	17.34
	8	F	26,974	203.60	16.33	23,365	143.31	15.42
		R	1,980	209.47	17.51	1,423	148.50	17.00
		P	20,324	215.74	18.54	18,103	153.91	17.80
K-PREP Reading	3	F	28,882	206.16	17.59	22,673	147.93	16.98
		R	1,990	211.90	17.44	1,276	153.37	17.49
		P	17,576	216.65	17.52	15,154	158.77	17.56
	4	F	30,004	206.77	14.75	24,023	154.89	14.05
		R	2,052	211.72	14.61	1,479	159.14	14.51
		P	18,512	216.29	14.87	16,593	163.57	14.22
	5	F	30,205	208.46	15.03	24,303	154.06	15.63
		R	2,015	213.06	14.48	1,472	159.51	16.46
		P	19,177	218.09	15.11	16,582	163.42	16.31
	6	F	29,693	208.89	14.82	24,227	157.16	14.21
		R	2,143	213.30	14.75	1,446	161.44	15.01
		P	19,650	218.28	14.56	17,223	166.11	15.22
	7	F	27,961	207.94	14.87	23,969	152.54	14.82
		R	2,078	212.71	14.33	1,346	157.47	14.67
		P	19,798	217.95	14.39	17,831	161.55	14.92
	8	F	26,871	209.73	15.20	23,408	156.14	15.71
		R	1,976	214.95	14.92	1,427	161.11	15.89
		P	20,278	219.66	14.72	18,142	165.76	15.96

Table A3. (Continued)

Test	Grade	Lunch Status	2018-2019			2020-2021		
			N	Mean	SD	N	Mean	SD
K-PREP Science	4	F	30,073	200.06	12.53	23,952	144.66	11.05
		R	2,056	204.35	12.10	1,476	147.32	11.26
		P	18,576	207.37	12.66	16,580	150.82	11.75
	7	F	28,000	197.82	12.02	23,818	135.96	12.23
		R	2,083	201.83	12.15	1,343	139.98	12.79
		P	19,813	205.79	12.76	17,796	143.75	13.90
	11	F	21,761	198.23	13.05	15,593	138.02	15.21
		R	1,905	202.59	13.18	1,091	143.06	15.77
		P	21,115	206.01	13.99	17,738	146.01	15.80
K-PREP On-Demand Writing	5	F	30,177	215.40	37.21	24,166	3.93	1.55
		R	2,015	224.82	35.76	1,467	4.44	1.58
		P	19,173	235.80	35.05	16,533	4.89	1.55
	8	F	26,800	212.28	33.98	23,183	4.17	1.75
		R	1,974	220.54	31.69	1,417	4.73	1.70
		P	20,262	229.80	31.08	18,056	5.29	1.66
	11	F	21,617	228.98	37.23	15,338	4.53	1.84
		R	1,899	239.24	35.58	1,081	5.07	1.83
		P	21,051	247.42	35.35	17,566	5.45	1.84

Note: F stands for free lunch status, R stands for reduced-price lunch status, and P stands for paid lunch status.

Table A4. Descriptive Statistics Underlying IEP Status Effect Size Comparisons

Test	Grade	Gender	2018-2019			2020-2021		
			N	**Mean**	**SD**	**N**	**Mean**	**SD**
K-PREP Math	3	N	40,889	211.44	20.68	32,711	158.57	17.13
		Y	7,728	199.52	20.19	6,394	149.84	16.53
	4	N	43,220	211.12	19.01	35,662	152.29	16.37
		Y	7,513	200.70	17.96	6,448	144.76	14.82
	5	N	44,453	213.08	19.65	36,192	151.88	17.89
		Y	7,125	199.93	17.82	6,138	143.70	15.33
	6	N	45,267	211.43	18.15	37,258	150.91	16.58
		Y	6,368	197.81	15.65	5,606	142.16	14.30
	7	N	44,306	211.82	17.87	37,710	148.58	16.36
		Y	5,664	197.78	14.31	5,368	138.30	12.79
	8	N	44,149	210.42	18.05	38,004	149.25	17.25
		Y	5,129	195.24	14.27	4,887	137.96	14.24
K-PREP Reading	3	N	40,722	211.72	17.77	32,702	153.35	17.92
		Y	7,727	202.18	18.68	6,401	146.97	17.46
	4	N	43,054	211.73	14.90	35,644	159.55	14.50
		Y	7,514	203.13	16.57	6,451	152.41	14.63
	5	N	44,277	213.79	15.03	36,206	159.12	16.29
		Y	7,121	202.53	16.48	6,151	150.78	16.30
	6	N	45,116	214.17	14.88	37,284	162.09	15.12
		Y	6,370	201.99	14.75	5,612	152.96	13.91
	7	N	44,173	213.77	14.72	37,764	157.86	15.11
		Y	5,664	199.21	14.77	5,382	146.31	14.35
	8	N	43,998	215.69	15.04	38,076	162.01	16.00
		Y	5,127	199.87	14.48	4,901	147.63	14.75
K-PREP Science	4	N	43,199	203.67	12.81	35,574	147.79	11.62
		Y	7,506	198.56	13.45	6,434	143.84	11.70
	7	N	44,240	202.17	12.78	37,619	140.28	13.47
		Y	5,656	193.19	10.96	5,338	132.52	11.75
	11	N	41,169	203.10	13.81	31,327	143.53	15.73
		Y	3,613	190.46	10.97	3,095	129.82	13.42
K-PREP On-Demand Writing	5	N	44,253	227.90	35.44	36,044	4.51	1.55
		Y	7,113	195.29	38.79	6,122	3.19	1.50
	8	N	43,932	223.87	31.47	37,805	4.89	1.70
		Y	5,104	185.31	33.54	4,851	2.86	1.46
	11	N	40,995	241.43	35.97	30,931	5.21	1.82
		Y	3,573	200.20	32.09	3,054	3.11	1.50

Note: N stands for non-IEP students, Y stands for IEP students.

Appendix B: Regression Tables

Table B1. Regression Coefficients for Gender Models

Test	Grade	Predictor	2017-2019			2019-2021		
			Estimate	SE	P	Estimate	SE	p
K-PREP Math	5	Intercept	-0.044	0.003	p < .001	-0.008	0.004	0.03
		Female	-0.007	0.015	0.11	-0.031	0.004	p < .001
		Previous Assessment Score	0.971	0.004	p < .001	0.647	0.004	p < .001
	6	Intercept	-0.059	0.003	p < .001	-0.015	0.004	p < .001
		Female	0.065	0.003	p < .001	-0.003	0.004	0.49
		Previous Assessment Score	1.036	0.004	p < .001	0.659	0.004	p < .001
	7	Intercept	-0.070	0.003	p < .001	-0.017	0.004	p < .001
		Female	0.035	0.003	0.01	-0.052	0.004	p < .001
		Previous Assessment Score	1.090	0.005	p < .001	0.698	0.004	p < .001
	8	Intercept	-0.065	0.003	p < .001	-0.012	0.004	p < .001
		Female	0.018	0.003	p < .001	-0.020	0.004	p < .001
		Previous Assessment Score	1.128	0.005	p < .001	0.677	0.004	p < .001
K-PREP Reading	5	Intercept	-0.059	0.003	p < .001	-0.008	0.004	0.06
		Female	0.030	0.003	p < .001	0.008	0.004	0.04
		Previous Assessment Score	0.964	0.005	p < .001	0.621	0.004	p < .001
	6	Intercept	-0.072	0.003	p < .001	-0.011	0.004	p < .001
		Female	0.056	0.003	p < .001	0.013	0.004	p < .001
		Previous Assessment Score	1.043	0.005	p < .001	0.651	0.004	p < .001
	7	Intercept	-0.081	0.003	p < .001	-0.005	0.004	0.17
		Female	0.064	0.003	p < .001	0.015	0.004	p < .001
		Previous Assessment Score	1.109	0.005	p < .001	0.646	0.004	p < .001
	8	Intercept	-0.090	0.003	p < .001	-0.002	0.004	0.55
		Female	0.100	0.003	p < .001	0.062	0.004	p < .001
		Previous Assessment Score	1.207	0.006	p < .001	0.698	0.004	p < .001

Table B2. Regression Coefficients for Race-Ethnicity Models

Test	Grade	Predictor	2017-2019			2019-2021		
			Estimate	SE	P	Estimate	SE	P
K-PREP Math	5	Intercept	-0.041	0.003	p < .001	-0.011	0.004	p < .001
		Asian	0.044	0.003	p < .001	0.033	0.004	p < .001
		African American	-0.073	0.003	p < .001	-0.073	0.004	p < .001
		Hispanic	-0.017	0.003	p < .001	-0.018	0.004	p < .001
		American Indian	-0.004	0.003	0.139	-0.004	0.004	0.32
		Multiracial	-0.010	0.003	0.001	-0.020	0.004	p < .001
		Native Hawaiian / Pacific Islander	-0.001	0.003	0.855	-0.005	0.004	0.18
		Previous Assessment Score	0.947	0.004	p < .001	0.632	0.004	p < .001
	6	Intercept	-0.056	0.003	p < .001	-0.018	0.004	p < .001
		Asian	0.037	0.003	p < .001	0.040	0.004	p < .001
		African American	-0.079	0.003	p < .001	-0.055	0.004	p < .001
		Hispanic	-0.026	0.003	p < .001	-0.012	0.004	p < .001
		American Indian	-0.001	0.003	0.718	-0.001	0.004	0.85
		Multiracial	-0.027	0.003	p < .001	-0.012	0.004	p < .001
		Native Hawaiian / Pacific Islander	-0.002	0.003	0.469	0.004	0.004	0.29
		Previous Assessment Score	1.007	0.005	p < .001	0.645	0.004	p < .001
	7	Intercept	-0.067	0.003	p < .001	-0.019	0.004	p < .001
		Asian	0.043	0.003	p < .001	0.045	0.004	p < .001
		African American	-0.082	0.003	p < .001	-0.047	0.004	p < .001
		Hispanic	-0.026	0.003	p < .001	-0.018	0.004	p < .001
		American Indian	0.000	0.003	0.911	-0.002	0.004	0.66
		Multiracial	-0.022	0.003	p < .001	-0.015	0.004	p < .001
		Native Hawaiian / Pacific Islander	-0.001	0.003	0.797	0.007	0.004	0.06
		Previous Assessment Score	1.059	0.005	p < .001	0.684	0.004	p < .001
	8	Intercept	-0.062	0.003	p < .001	-0.013	0.004	p < .001
		Asian	0.044	0.003	p < .001	0.042	0.004	p < .001
		African American	-0.053	0.003	p < .001	-0.030	0.004	p < .001
		Hispanic	-0.015	0.003	p < .001	0.004	0.004	0.34
		American Indian	0.000	0.003	0.903	-0.002	0.004	0.64
		Multiracial	-0.006	0.003	0.033	0.000	0.004	1.00
		Native Hawaiian / Pacific Islander	-0.002	0.003	0.457	0.004	0.004	0.29
		Previous Assessment Score	1.106	0.005	p < .001	0.669	0.004	p < .001

Test	Grade	Predictor	2017-2019			2019-2021		
			Estimate	SE	P	Estimate	SE	P
K-PREP Reading	5	Intercept	-0.057	0.003	p < .001	-0.010	0.004	0.01
		Asian	0.026	0.003	p < .001	0.031	0.004	p < .001
		African American	-0.067	0.003	p < .001	-0.050	0.004	p < .001
		Hispanic	-0.029	0.003	p < .001	-0.025	0.004	p < .001
		American Indian	-0.002	0.003	0.646	0.001	0.004	0.78
		Multiracial	-0.008	0.003	0.014	-0.010	0.004	p < .001
		Native Hawaiian / Pacific Islander	-0.002	0.003	0.484	-0.004	0.004	0.36
		Previous Assessment Score	0.943	0.005	p < .001	0.609	0.004	p < .001
	6	Intercept	-0.070	0.003	p < .001	-0.012	0.004	p < .001
		Asian	0.034	0.003	p < .001	0.057	0.004	p < .001
		African American	-0.061	0.003	p < .001	-0.019	0.004	p < .001
		Hispanic	-0.014	0.003	p < .001	-0.001	0.004	0.89
		American Indian	-0.001	0.003	0.667	0.000	0.004	0.94
		Multiracial	-0.016	0.003	p < .001	-0.006	0.004	0.15
		Native Hawaiian / Pacific Islander	-0.003	0.003	0.353	0.000	0.004	0.93
		Previous Assessment Score	1.025	0.005	p < .001	0.646	0.004	p < .001
	7	Intercept	-0.079	0.003	p < .001	-0.007	0.004	0.05
		Asian	0.032	0.003	p < .001	0.036	0.004	p < .001
		African American	-0.064	0.003	p < .001	-0.051	0.004	p < .001
		Hispanic	-0.014	0.003	p < .001	-0.019	0.004	p < .001
		American Indian	-0.002	0.003	0.526	-0.012	0.004	p < .001
		Multiracial	-0.010	0.003	0.002	-0.008	0.004	0.03
		Native Hawaiian / Pacific Islander	-0.002	0.003	0.465	0.000	0.004	0.94
		Previous Assessment Score	1.095	0.005	p < .001	0.636	0.004	p < .001
	8	Intercept	-0.088	0.003	p < .001	-0.004	0.004	0.29
		Asian	0.027	0.003	p < .001	0.028	0.004	p < .001
		African American	-0.068	0.003	p < .001	-0.039	0.004	p < .001
		Hispanic	-0.011	0.003	p < .001	-0.003	0.004	0.42
		American Indian	-0.003	0.003	0.313	0.005	0.004	0.19
		Multiracial	-0.007	0.003	0.028	-0.001	0.004	0.82
		Native Hawaiian / Pacific Islander	0.001	0.003	0.866	-0.002	0.004	0.53
		Previous Assessment Score	1.199	0.006	p < .001	0.695	0.004	p < .001

Table B3. Regression Coefficients for Lunch Status Models

Test	Grade	Predictor	2017-2019			2019-2021		
			Estimate	SE	P	Estimate	SE	p
K-PREP Math	5	Intercept	-0.041	0.003	p < .001	-0.012	0.004	p < .001
		Free Lunch	-0.118	0.003	p < .001	-0.146	0.004	p < .001
		Reduced Lunch	-0.023	0.003	p < .001	-0.029	0.004	p < .001
		Previous Assessment Score	0.926	0.004	p < .001	0.607	0.004	p < .001
	6	Intercept	-0.055	0.003	p < .001	-0.019	0.004	p < .001
		Free Lunch	-0.113	0.003	p < .001	-0.131	0.004	p < .001
		Reduced Lunch	-0.027	0.003	p < .001	-0.028	0.004	p < .001
		Previous Assessment Score	0.987	0.005	p < .001	0.621	0.004	p < .001
	7	Intercept	-0.066	0.003	p < .001	-0.020	0.003	p < .001
		Free Lunch	-0.127	0.003	p < .001	-0.125	0.004	p < .001
		Reduced Lunch	-0.029	0.003	p < .001	-0.029	0.004	p < .001
		Previous Assessment Score	1.033	0.005	p < .001	0.661	0.004	p < .001
	8	Intercept	-0.062	0.003	p < .001	-0.016	0.004	p < .001
		Free Lunch	-0.100	0.003	p < .001	-0.110	0.004	p < .001
		Reduced Lunch	-0.020	0.003	p < .001	-0.020	0.004	p < .001
		Previous Assessment Score	1.082	0.005	p < .001	0.644	0.004	p < .001
K-PREP Reading	5	Intercept	-0.057	0.003	0.01	-0.011	0.004	0.01
		Free Lunch	-0.125	0.004	p < .001	-0.121	0.004	0.04
		Reduced Lunch	-0.023	0.003	p < .001	-0.018	0.004	p < .001
		Previous Assessment Score	0.920	0.005	p < .001	0.590	0.004	p < .001
	6	Intercept	-0.069	0.003	p < .001	-0.015	0.004	p < .001
		Free	-0.102	0.003	p < .001	-0.112	0.004	p < .001
		Reduced Lunch	-0.022	0.003	p < .001	-0.023	0.004	p < .001
		Previous Assessment Score	1.004	0.005	p < .001	0.620	0.004	p < .001
	7	Intercept	-0.078	0.003	p < .001	-0.008	0.004	0.02
		Free	-0.115	0.003	p < .001	-0.108	0.004	p < .001
		Reduced Lunch	-0.021	0.003	p < .001	-0.015	0.004	p < .001
		Previous Assessment Score	1.067	0.005	p < .001	0.617	0.004	p < .001
	8	Intercept	-0.087	0.003	p < .001	-0.006	0.003	0.08
		Free Lunch	-0.108	0.003	p < .001	-0.093	0.004	p < .001
		Reduced Lunch	-0.017	0.003	p < .001	-0.015	0.004	p < .001
		Previous Assessment Score	1.172	0.006	p < .001	0.677	0.004	p < .001

Table B4. Regression Coefficients for IEP Status Models

Test	Grade	Predictor	2017-2019			2019-2021		
			Estimate	SE	P	Estimate	SE	p
K-PREP Math	5	Intercept	-0.043	0.003	p < .001	-0.010	0.004	.012
		IEP Students	-0.061	0.003	p < .001	-0.004	0.004	.308
		Previous Assessment Score	0.957	0.004	p < .001	0.651	0.004	p < .001
	6	Intercept	-0.054	0.003	p < .001	-0.016	0.004	p < .001
		IEP Students	-0.090	0.003	p < .001	-0.036	0.004	p < .001
		Previous Assessment Score	1.008	0.005	p < .001	0.653	0.004	p < .001
	7	Intercept	-0.068	0.003	p < .001	-0.017	0.004	p < .001
		IEP Students	-0.105	0.003	p < .001	-0.044	0.004	p < .001
		Previous Assessment Score	1.061	0.005	p < .001	0.689	0.004	p < .001
	8	Intercept	-0.064	0.003	p < .001	-0.014	0.004	p < .001
		IEP Students	-0.078	0.003	p < .001	-0.040	0.004	p < .001
		Previous Assessment Score	1.104	0.005	p < .001	0.669	0.004	p < .001
K-PREP Reading	5	Intercept	-0.060	0.003	p < .001	-0.012	0.004	0.04
		IEP Students	-0.109	0.004	p < .001	-0.046	0.004	p < .001
		Previous Assessment Score	0.945	0.005	p < .001	0.622	0.004	p < .001
	6	Intercept	-0.069	0.003	p < .001	-0.013	0.004	p < .001
		IEP Students	-0.119	0.003	p < .001	-0.059	0.004	p < .001
		Previous Assessment Score	1.014	0.005	p < .001	0.646	0.004	p < .001
	7	Intercept	-0.078	0.003	p < .001	-0.006	0.004	0.145
		IEP Students	-0.143	0.003	p < .001	-0.094	0.004	p < .001
		Previous Assessment Score	1.067	0.005	p < .001	0.627	0.004	p < .001
	8	Intercept	-0.086	0.003	p < .001	-0.005	0.004	0.139
		IEP Students	-0.139	0.003	p < .001	-0.101	0.004	p < .001
		Previous Assessment Score	1.165	0.006	p < .001	0.679	0.004	p < .001