Effects of COVID-19 on Student Achievement in Large Scale Assessments

Paper presented at the annual meeting of

the National Council on Measurement in Education,

San Diego

April 23, 2022

Aurore Yang Phenow, Dong-In Kim, Keith Boughton

Data Recognition Corporation

Effects of COVID-19 on Student Achievement in Large Scale Assessments

The COVID-19 pandemic disrupted educational learning both directly and indirectly. In March 2020 schools began to shift from traditional in-person learning to remote or hybrid instruction with some schools closed for an extended duration. At the peak of the pandemic, it is estimated that 55.1 million students and 124,000 schools in the United States were affected by school closures or switched to remote learning (Education Week, 2020). These impacts were not limited to education, as stay-at-home orders were implemented and non-essential companies and businesses were also ordered to close for extended periods (Wu et al., 2020). Combined with remote learning, this caused concerns about potential learning loss for students during the COVID-19 pandemic. However, the pandemic learning atmosphere differed amongst students as families faced differing challenges. Some parents worked from home while facilitating their student's learning, while others were essential workers and had to find alternate childcare. According to CCSSO (2020), although the disruptions and subsequent impact on learning were widespread, it may not be equal for all groups of students and may differ based on demographic factors. More specifically, the accessibility to broadband internet needed for remote learning became a highlighted issue with students from rural towns, lower social-economic status (SES), and Black and Latino or Hispanic households having less access when compared to white households (Fishbane et al., 2020). In terms of gender, there is concern that women may have been disproportionately affected in the workplace during the pandemic, which may have increased female students' role at home in terms of childcare or housework (Acosta and Evans, 2020; Bateman and Ross, 2020). There have also been concerns regarding the academic progress of students with disability and ethnic minorities. Studies from NWEA (2021) show that students with disabilities may not have progressed as far when compared to students without a disability when learning is halted. A study by Kuhfeld et al. (2020) identified that math test scores from Black, Indigenous, and people of color (BIPOC) were significantly lower in fall 2020 than in 2019 and students from high-poverty schools were the most at risk for not making academic gains. In this paper, the impact of COVID-19 on student learning across a wide range of demographic variables was studied by analyzing large scale state assessment data.

The main purpose of this study was to determine if there were differences in student performance as a result of COVID-19 across several demographic variables, using data from a large scale state assessment program. In order to study this type of learning loss, the scale score results from a 2019 administration of a large scale state program that was taken prior to the pandemic were utilized, as well as the 2021 assessment results for that same program which was administered after the pandemic mandates were put in place. Student data were analyzed by building four regression models to explore and explain the differences in scaled scores between the two years (i.e., before and after COVID-19 restriction were put in place). In education data, students are nested within schools, and students from the same school may be more related to each other than students from different schools. Failing to account for this nested structure may lead to biased estimates of the effect from the regression models (Raudenbush and Bryk, 2002; Osborne, 2000). Three of the models in this study utilized multilevel modeling to account for the hierarchical, nested, structure of the data. Scaled scores from this large-scale state assessment were regressed onto year of assessment and the student and school demographics of gender, ethnicity, social economic status (SES), disability, limited English proficiency (LEP), school location, and school type. The outcome variable is the scaled score for the assessment. The predictor variable is the year of assessment, depicted as cohort in the study. Student and school demographic variables served as covariates to help explain the differences due to year of assessment. The main research questions that were answered include the following:

1. Overall, is there a difference in scaled scores from 2019 to 2021?

The hypothesis was that there will be a significant difference between scaled scores in 2019 and 2021, with 2021 having lower scaled scores when compared to 2019, due to the pandemic related disruption to a state's educational system. This assumes that students who took the assessment in 2021 should have similar scores to students who took the assessment in 2019 if there was no pandemic impact.

2. When accounting for the nesting structure in the student data, will there be a significant difference in scaled scores between 2019 and 2021?

The hypothesis was that there will be a significant difference between scaled scores in 2019 and 2021 after accounting for the nesting structure with 2021 having lower scaled scores than 2019 due to the disruption from the pandemic.

3. At the student level, what covariates will contribute to the differences between scaled scores in 2019 and 2021 if student performance differences exist?

The hypothesis was that females, SES disadvantaged, LEP, disability, and minority ethnicity students will have a greater drop in scaled score between 2019 and 2021 than male, non-SES disadvantaged, non-LEP, non-disability students, and white students. Early 2020 testing data showed learning loss was greater for lower SES and BIPOC students (Kuhfeld et al., 2020; Lewis et al., 2021; NWEA, 2021). LEP students tend to also be of lower SES status and have additional barriers to learning, such as less access to internet and computers (Choi, 2021). Female students may hold more household responsibilities during the pandemic causing greater learning loss than male students (Acosta and Evans, 2020; Bateman and Ross, 2020).

4. At the school level, what covariates will contribute to the differences between scaled scores in 2019 and 2021 if student performance difference exists?

City, rural, and public schools were hypothesized to have a greater drop in scaled score between 2019 and 2021. City schools have students from lower SES families and these students have been shown to have increased learning loss during the early stages of the pandemic (Kuhfeld et al., 2020; Lewis et al., 2021). Rural towns were more likely to not have access to broadband internet, which may cause decreases in school attendance and lower learning gains (Fishbane et al., 2020). Mandates from governors were only applicable to public schools, so non-public schools had the option to continue in-person learning during the pandemic and may cause public school students to have decreased learning (Dickler, 2020).

Methods

Data

Large scale assessment data for English Language Arts (ELA) and math for grade 6 were used. The data consisted of 2019 and 2021 grade 6 academic and demographic student data. Prior academic performance has been shown to be highly correlated to current academic performance, so a prior academic indicator was included as a covariate. 2019 grade 6 students were matched to their 2017 grade 4 assessment scores, and 2021 grade 6 students were matched to their 2019 grade 4 assessment scores. The grade 6 scale is the same for 2019 and 2021 and the grade 4 scale is the same for 2017 and 2019. ELA had a scaled score range of 590 and math had a scaled score range of 430 for grade 6. ELA had a scaled score range of 590 for ELA and 395 for math for grade 4.

Because student participation in the assessment in 2021 was different from 2019, propensity score matching (PSM) was previously performed on the 2021 dataset to ensure representative sampling between the 2019 and 2021 grade 6 students for the calibration of the test items in 2021. PSM included matching on student demographics. This study includes only the students in 2021 that were chosen for the calibration sample to ensure similar student representation in 2021 as there was in 2019. Additionally, students were chosen for analysis if they had a 2021 or 2019 grade 6 scaled score, had a 2017 or 2019 prior indicator scaled score, earned a valid score on the assessment, and had complete demographic characteristics.

Model building

A simple linear regression and multilevel mixed effects models were built in R using the lme4 package (Bates et al., 2015) to analyze the differences between 2019 and 2021 scaled scores for ELA and math grade 6 students. Using a multilevel model accommodates the natural nesting structure of student data with students being nested within schools (Raudenbush and Bryk, 2002; Osborne, 2000). The main research question was, "is there a difference in scaled scores between the 2019 and 2021 cohorts?". To compare the difference in regression coefficients between 2019 and 2021, the models built combined both 2019 and 2021 data into one data set. The variable cohort was created and was dummy coded as 0 for 2019 and 1 for 2021. For the rest of the paper, the term cohort will be used interchangeably with year and cohort 0 pertains to the year 2019 while cohort 1 pertains to 2021. The dependent variable was the year 2 scaled score (Y2SS), which was the grade 6 2021 scaled score for students who took the assessment in 2021 and the grade 6 2019 scaled score for students who took the assessment in 2019. Because the scaled scores are restricted to a range, an intercept that is outside of the range, such as a zero, can be difficult to interpret. The Y2SS was centered among the student level so the predicted Y2SS seen in the model results can be interpreted as the amount of scaled score above or below the average student scaled score, which is set to zero. Therefore, a student who's predicted scale score is 15 can be interpreted as 15 scaled score points above the average Y2SS scaled score, and a Y2SS of -10 is 10 points below the average Y2SS scaled score. The primary predictor variable was cohort as the main purpose of the study is to compare the differences due to year. The effect of cohort 1 will show the increase or decrease in scaled score of students testing in 2021 instead of 2019. Student covariates included a previous academic indicator, gender, ethnicity, SES, disability status, and LEP status. Year 1 scaled score (Y1SS) was the variable used as the previous academic indicator and consisted of the 2019 scaled score for 2021 students and the 2017 scaled score for 2019 students. Similar to the Y2SS, the Y1SS was also centered around the average student scaled score. Gender was coded as male or female. Ethnicity included white, Hispanic, Black, Asian, and other. SES, disability, and LEP were binary variables of no or yes, with a value of no representing no SES, no disability, or no LEP. School covariates included school location and school type. School locations included city, suburb, town, rural and other. School type is a binary variable that is yes for public school and no for non-public school. The interaction between the cohort and the student and school variables were utilized to test the differences due to the cohort on the variables and shows the additional effect of students taking the assessment in 2021. The effects of students who took the assessment in 2019 are shown in the non-interaction variables. For all multilevel models, only the nesting of students within schools were treated as random and all other variables were treated as fixed effects.

Four models were produced for each subject area of ELA and math:

1.) Simple linear regression model to test for differences in scaled scores due to cohort year

$$Y2SS = B_0 + B_1 Cohort + e$$

where B_0 is the intercept, B_1 is the effect of taking the test in cohort 1 (2021), and e is the error term.

2.) Unconditional (null) multilevel mixed effects model to show the variability due to schools for each year 2019 and 2021.

$$Y2SS_{ij} = B_0 + \mu_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\mu_j \sim N(0, \sigma_\mu^2)$$

where B_0 is the intercept, μ_j is the error on the school level j, and e_{ij} is the error term for an individual i in school j.

3.) Multilevel mixed effects model that includes only the cohort predictor variable and the student covariates.

$$\begin{aligned} Y2SS_{ij} = B_0 &+ B_1 Y1SS + B_2 Female + B_3 Hispanic + B_4 Black + B_5 Asian + B_6 RaceOther \\ &+ B_7 SESDisadvantaged + B_8 Disability + B_9 LEP + B_{10} Cohort(2021) \\ &+ B_{11}(Cohort * Female) + B_{12}(Cohort * Hispanic) + B_{13}(Cohort * Black) \\ &+ B_{14}(Cohort * Asian) + B_{15}(Cohort * RaceOther) + B_{16}(Cohort \\ &* SESDisadvantaged) + B_{17}(Cohort * Disability) + B_{18}(Cohort * LEP) + \mu_j + e_{ij} \end{aligned}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\mu_j \sim N(0, \sigma_\mu^2)$$

where B_0 is the intercept, B_x is the effects for each student variable x, μ_j is the error on the school level j, and e_{ij} is the error term for an individual i in school j.

4.) Multilevel mixed effects model that includes both the cohort predictor variable and the student and school covariates.

$$\begin{aligned} Y2SS_{ij} = B_0 &+ B_1 Y1SS + B_2 Female + B_3 Hispanic + B_4 Black + B_5 Asian + B_6 RaceOther \\ &+ B_7 SESDisadvantaged + B_8 Disability + B_9 LEP + B_{10} Cohort(2021) \\ &+ B_{11}(Cohort * Female) + B_{12}(Cohort * Hispanic) + B_{13}(Cohort * Black) \\ &+ B_{14}(Cohort * Asian) + B_{15}(Cohort * RaceOther) + B_{16}(Cohort \\ &* SESDisadvantaged) + B_{17}(Cohort * Disability) + B_{18}(Cohort * LEP) \\ &+ B_{19} LocSuburb + B_{20} LocTown + B_{21} LocRural + B_{22} LocOther + B_{23} NonPublic \\ &+ B_{24}(Cohort * LocSuburb) + B_{25}(Cohort * LocTown) + B_{26}(Cohort * LocRural) \\ &+ B_{27}(Cohort * LocOther) + B_{28}(Cohort * NonPublic) + \mu_j + e_{ij} \end{aligned}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\mu_j \sim N(0, \sigma_\mu^2)$$

where B_0 is the intercept, B_x is the effects for each student variable x, μ_j is the error on the school level j, and e_{ij} is the error term for an individual i in school j.

In model 1, a simple linear regression tested the significance of the difference in means between cohorts 0 and 1. The intraclass correlation (ICC) of model 2 was calculated to determine the amount of variation due to the level 2 (school) level. This was conducted separately for 2019 and 2021, and the ICC in each of the models was compared. Model 3 and model 4 were used to account for the nesting structure in the student assessment data and to garner the effects of cohort and the student and school covariates on the predicted Y2SS. The equations in models 3 and 4 demonstrate how the effects due to cohorts were measured. When cohort was coded as 0 for 2019, then the effect of B_{10} to B_{18} and $\mathrm{B}_{24}$to B_{28} all went to zero and the effects shown were only from 2019. Thus, the effect for 2019 for model 3 is

$$\begin{aligned} Y2SS_{ij} = B_0 + B_1 Y1SS + B_2 Female + B_3 Hispanic + B_4 Black + B_5 Asian + B_6 RaceOther \\ + B_7 SESDisadvantaged + B_8 Disability + B_9 LEP + \mu_j + e_{ij} \end{aligned}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\mu_j \sim N(0, \sigma_\mu^2)$$

and for model 4 is

$$\begin{aligned} Y2SS_{ij} = B_0 + B_1 Y1SS + B_2 Female + B_3 Hispanic + B_4 Black + B_5 Asian + B_6 RaceOther \\ + B_7 SESDisadvantaged + B_8 Disability + B_9 LEP + B_{19} LocSuburb + B_{20} LocTown \\ + B_{21} LocRural + B_{22} LocOther + B_{23} NonPublic + \mu_j + e_{ij} \end{aligned}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\mu_j \sim N(0, \sigma_\mu^2)$$

When cohort was coded as 1 for 2021, then the effects of B_{10} to B_{18} and $\mathrm{B}_{24}$to B_{28} may be included and the effects from 2021 were added to the effects from 2019. The addition of these effects can be interpreted as the differences due to cohort. A significant cohort effect indicated differences in score from 2019 to 2021, and significant effects in the interaction variables showed which student and school covariates differed from 2019 to 2021.

Models 3 and 4 can be viewed as nested models, so these models were tested for comparative model fit. Akaike's Information Criteria (AIC) penalizes the use of additional variables to the model while the Bayesian Information Criteria (BIC) does not, and both were used to assess model fit. The R^2 value is the proportion of variation in the outcome variable that is explained by the independent variables in the model. A conditional R^2 value can be used in multilevel models to account for the fixed and random effects (Nakagawa and Schielzeth, 2013) and will also be used to assess model fit.

Results

Data

For ELA, there were 91,355 total students with 60,826 students in 2019, 30,509 students in 2021, 840 total schools in 2019, and 782 total schools in 2021. Math had a total 91,325 students with 60,966 students in 2019, 30,359 students in 2021, 841 schools in 2019, and 781 schools in 2021. Table 1 shows the distribution of the students separated by the model variables, subject, and years with the reference group in bold. Overall, the distributions between years were very similar. There were slightly more males than females (51% to 49%) and the ethnicity of white was the majority. The number of SES disadvantaged students is about 43%, showing that just under half the student population is considered lower economic status. About 12% of the students were disabled and about 6% had LEP. The city and suburb each comprise of about 28% of the student population and most students attend public schools. The student demographic distribution for math was similar to ELA, but this is expected as students should have taken both assessments. The minor differences seen are due to students meeting the inclusion criteria for the analysis only for one subject. The distributions also show that the 2019 and 2021 matched sample were highly similar in their distribution. Figure 1 was produced to show the distribution of SES disadvantaged students compared to the school location. City schools had more SES disadvantaged students than not and had more compared to the other locations of suburb, town, rural, and other.

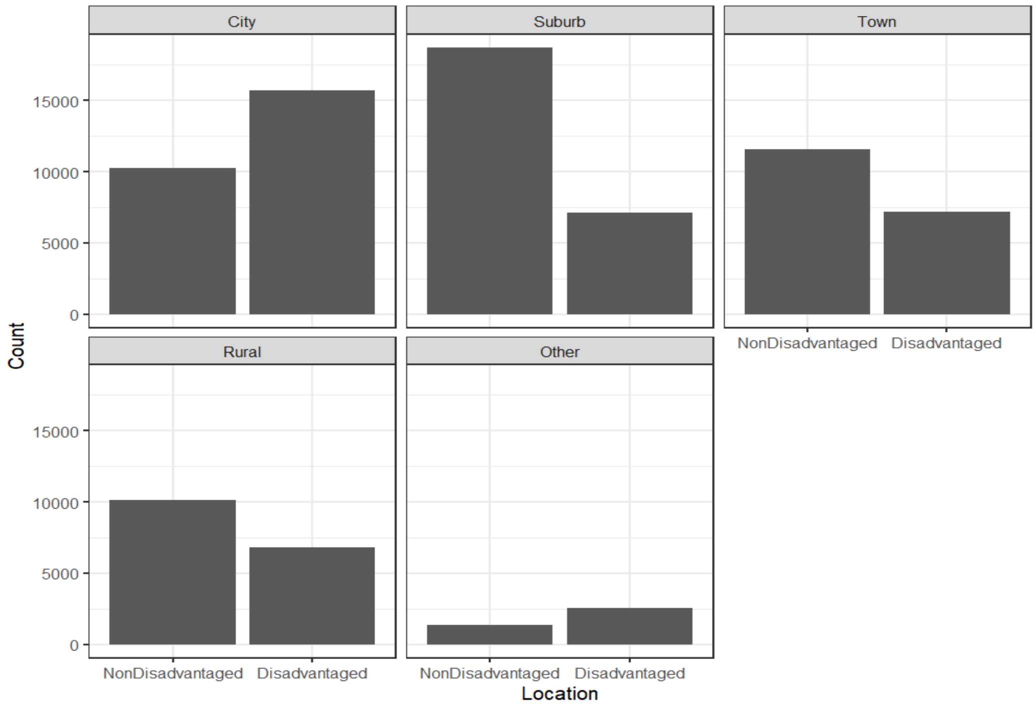
Table 1

Distribution of students based on subject, year, and model variables

	Variable	ELA		Math	
		2019	2021	2019	2021
		Pct (%)		Pct (%)	
Gender	**Male**	51.18	51.34	51.18	51.34
	Female	48.82	48.66	48.82	48.66
Ethnicity	**White**	67.28	67.73	67.16	68.00
	Hispanic	13.33	13.30	13.39	13.30
	Black	10.31	9.91	10.32	9.66
	Asian	3.86	3.99	3.90	3.99
	Other	5.22	5.06	5.23	5.05
SES	**NonDisadvantaged**	56.79	57.02	56.71	57.26
	Disadvantaged	43.21	42.98	43.29	42.74
Disability	**NonDisabled**	87.87	88.44	87.85	88.52
	Disabled	12.13	11.56	12.15	11.48
LEP	**NonLEP**	94.23	94.38	94.07	94.29
	LEP	5.77	5.62	5.93	5.71
Location	**City**	28.64	28.00	28.69	27.68
	Suburb	28.28	28.27	28.27	28.41
	Town	20.40	20.65	20.37	20.73
	Rural	18.46	18.73	18.43	18.81
	Other	4.22	4.36	4.23	4.37
School Type	**Public**	95.35	95.09	95.34	95.07
	NonPublic	4.65	4.91	4.66	4.93

Figure 1

Distribution of social economic status based on school location



Model Fit

ELA.

Result for model fit can be seen in Table 2.The R^2 for the simple linear regression model (model 1) was small ($R^2 = 0.002$) which signifies covariates need to be added to help explain the variance in the Y2SS. The conditional R^2 for the 2019 unconditional mixed effect model (model 2) was 0.230 and for the 2021 model was 0.204. Models 3 and 4 were considered nested models and were be compared using conditional R^2, AIC, and BIC. Both models had similar conditional R^2 values of 0.641 and 0.643. The AIC was lower for model 4, but the BIC was lower for model 3. Since AIC penalizes for additional variables but was lower for model 4, model 4 was considered to have better model fit.

Math.

Model 1 had a small R^2 of 0.008. The R^2 of both 2019 and 2021 in model 2 were similar with the 2019 model having an R^2 of 0.265 and 0.271 for the 2021 model. Models 3 and 4 also had similar conditional R^2 values of 0.636 and 0.642. The AIC and BIC was smaller for model 4, showing that model 4 had the better model fit.

Table 2

Fit statistics for regression models

Model	Model Number	R^2	R^2 Marginal	R^2 Conditional	AIC	BIC
			ELA			
Simple Linear Regression	1	0.002			972427	972455
2019 Unconditional Mixed Effects Model	2		0.000	0.230	636933	636960
2021 Unconditional Mixed Effects Model	2		0.000	0.204	320672	320697
Mixed Effects Model with Student Covariates	3		0.611	0.641	878875	879073
Mixed Effects Model with School Covariates	4		0.614	0.643	878830	879122
			Math			
Simple Linear Regression	1	0.008			1000462	1000490
2019 Unconditional Mixed Effects Model	2		0.000	0.265	655035	655062
2021 Unconditional Mixed Effects Model	2		0.000	0.271	326292	326317
Mixed Effects Model with Student Covariates	3		0.589	0.636	908297	908495
Mixed Effects Model with School Covariates	4		0.598	0.642	908154	908446

Models

Model 1: Simple linear regression model

Simple linear regression was conducted to answer research question 1. For ELA, cohort 1 (2021) was 4.86 scale scores points lower than cohort 0 (2021), and for math, cohort 1 scored 11.21 points lower than cohort 0. Both predictors are significant ($p < 0.001$) and showed that based on just the comparison of means, the 2019 cohort had higher scores on both the ELA and math assessments. Table 3 shows the results of the simple linear regression.

Table 3

Simple linear regression models for ELA and math

ELA Scaled Score

Predictors	Estimates	SE	*p*
(Intercept)	1.62	0.20	**<0.001**
Cohort [1]	-4.86	0.35	**<0.001**
Observations	91335		
R^2 / R^2 adjusted	0.002 / 0.002		

Math Scaled Score

Predictors	*Estimates*	*SE*	*p*
(Intercept)	3.73	0.23	**<0.001**
Cohort [1]	-11.21	0.41	**<0.001**
Observations	91325		
R^2 / R^2 adjusted	0.008 / 0.008		

Model 2: Unconditional multilevel mixed effects model

An unconditional (null) multilevel mixed effects model was built to calculate the ICC for determining the amount of variation due to schools and to provide support for using a multilevel model. The ICC showed that for ELA, 23% of the variation was due to schools in 2019 compared to 20% in 2021, indicating that the variations in scores between schools decreased in 2021. The ICC for math was about 27% for both 2019 and 2021, so the variation in scores between schools was about the same for both years (Table 4). The school level variation supports the use of a multilevel model.

Table 4

Unconditional mixed effects model for ELA and math

ELA	**2019**			**2021**		
Predictors	Estimates	SE	*p*	Estimates	SE	*p*
(Intercept)	603.97	0.90	**<0.001**	597.31	0.96	**<0.001**
Random Effects						
σ^2	1992.32			2050.89		
τ_{00}	595.11 SchoolNumber			526.00 SchoolNumber		
ICC	0.23			0.20		
N	840 SchoolNumber			782 SchoolNumber		
Observations	60826			30509		
Marginal R^2 / Conditional R^2	0.000 / 0.230			0.000 / 0.204		

Math	2019			2021		
Predictors	Estimates	SE	*p*	Estimates	SE	*p*
(Intercept)	-1.73	1.12	0.123	-15.44	1.26	**<0.001**
Random Effects						
σ^2	2611.49			2580.85		
τ_{00}	943.67 $_{\text{SchoolNumber}}$			959.38 $_{\text{SchoolNumber}}$		
ICC	0.27			0.27		
N	841 $_{\text{SchoolNumber}}$			781 $_{\text{SchoolNumber}}$		
Observations	60966			30359		
Marginal R^2 / Conditional R^2	0.000 / 0.265			0.000 / 0.271		

Model 3: Multilevel mixed effects model: cohort predictor and student level covariates

Models 3 and 4 were built to answer research questions 2 and 3. A multilevel mixed effects model including only cohort and the student level covariates was built to investigate the effect of cohort on the scaled scores. The effects of the variables from 2019 are the covariates without the cohort interactions. Scaled scores were centered so the interpretation was an increase in scaled scores above the mean for positive values and below the mean for negative values. The ELA Y2SS were centered around 606, and Y1SS were centered around 586. The math Y2SS were centered around 608, and Y1SS were centered around 577. For model 3, the reference group was male, white, non-SES disadvantaged, non-disability, non-LEP, cohort 0 (2019), and a Y1SS that was equal to the average Y1SS and centered to be zero. All result interpretations were compared to the reference group and to holding all other variables constant.

ELA.

Table 5 shows the results of model 3 for ELA. Cohort 1 was significant and decreases the scale score by 0.79 points, supporting the hypothesis that when accounting for the nesting structure of the data, scaled scores in cohort 0 will be lower than cohort 1. Covariates that had significant interaction with the cohort variable include female gender, Hispanic and Black ethnicities, SES disadvantaged, disability, and LEP students, indicating that for those covariates, the effects differed between 2019 and 2021. In 2019, being a female increased the scaled score 6.56 points compared to males, but the interaction effect with cohort lowered that by 4.02 points. Therefore, a female in 2021 would show a 2.54 scaled score point increase when compared to males in 2021. Figure 2 shows the moderating effect of gender by cohort with the slope lines representative of the reference group and allows only gender and cohort to change. Although the gap between males and females was large in 2019, the slope for females was more negative, showing that taking the assessment in 2021 had a greater negative effect on females than males. This can also be seen in the narrowing of the gap in Y2SS between females and males in cohort 1 compared to cohort 0. In model 3, all ethnicities except Asian and Hispanic students had lower scaled scores compared to the reference of white students, but the effect of Hispanic students was not significant. Asian students scored about 7.42 points higher than white students. The largest negative difference is between Black and white students with Black students scoring 6.54 points less. When cohort was included, only ethnicities of Hispanics and Blacks were significantly different than 2019 with both scoring lower in 2021. While in 2019 Blacks students scored 6.54 points lower than white students, the effect was increased in 2021 with Black students scoring 11.15 points lower than white students in 2021. This was consistent with the hypothesis in research question 3 that the ethnicity of Black would be the most negatively impacted. These results are also represented in Figure 4 that shows the steepest decrease in slope for Black ethnicity. Overall, the scaled scores in 2021 declined for all ethnicity groups. SES students scored 5.50 points lower than their non-SES counterparts in 2019 and in 2021, this is further increased to 7.38 points lower. Disabled students scored 16.71 points lower than non-disabled students in 2019, but unexpectedly, in 2021 that gap narrowed to 12.24. LEP students performed 6.20 points lower than non-LEP students in 2019 and their performance decreased further by 4.09 points in 2021. Except for students with disability, student scale scores in 2021 decreased when compared to 2019. All results except that from the disability covariate were congruent with the hypothesis in research question 3.

Math.

Cohort and all variables except Hispanic ethnicity for the 2019 covariates were significant. This parallels the hypothesis for research question 2. Students in 2021 scored about 11 points lower than those in 2019. Covariates that had significantly different effects in 2021 compared to 2019 are female gender; ethnicities of Hispanic, Black, and other; and SES disadvantaged, disability, and LEP students. Females in 2019 scored 4.28 points higher than males in 2019, but in 2021, the difference dropped to 0.37. Overall, scaled scores decreased for both males and females, but the decrease was greater for females than males. This can be seen in Figure 3 as the gap in performance narrows from cohort 0 to cohort 1. In 2019, an ethnicity of Black resulted in 9.54 scaled score points lower than the reference ethnicity of white and in 2021, that dropped even lower to 16.43. Asian students scored 7.97 points higher than white students in 2019, and the interaction of Asian students with cohort was not significant, showing that the effect of being an Asian student in 2021 was not different from 2019. Compared to white students, an ethnicity of other scored 2.61 points lower in 2019 and that increased in 2021 to 4.97 points lower. Overall, all ethnicities scored lower in 2021 than 2019, as shown in Figure 5. Students who were SES disadvantaged performed 7.02 scaled score points lower than non-SES disadvantaged students in 2019, and the effect of being in cohort 1 decreased the score by another 2.71 points. Students with disability in cohort 0 had scale scores that were 18.93 points lower than their non-disability counterparts, but similar to ELA, in 2021 that gap was reduced by 5.09 points, showing that the gap in scaled scores lessened in 2021 compared to 2019. Like ELA, except for disability, all results were as hypothesized in research question 3.

Comparison of ELA and Math Results.

Compared to ELA, the effect of cohort for math was greater. For the reference group, students who took the math assessment in 2021 compared to 2019 dropped 11.10 points while those who took the ELA assessment dropped 0.79 points. The effects can be visually represented by looking at the steepness of the slope. A visual representation of the gender and ethnicity covariates effects can be seen in Figures 2 to 5. For math, the slopes were more negative than for ELA, showing that the effect was greater in math. For all genders and ethnicities, the slope was negative, indicating that all genders and ethnicities performed better on both the ELA and math assessment in 2019 than in 2021. The figures also showed that overall, females performed better than males in 2019, but in 2021 the gap decreased.

Model 4: Multilevel mixed effects model: cohort predictor and student and school level covariates

Model 4 was built to address research questions 2 and 4. In addition to the student level covariates in model 3, school level variables were added as covariates to help explain the differences due to cohort. For model 4, the reference group for students was the same as model 3, and the reference group for the school variables was a location of city and a school type of public.

ELA.

Similar to the ELA results in model 3, all 2019 student variables except Hispanic were significant (Table 5). Results for the student predictors were similar to that seen in the results from model 3. With the addition of the school covariates, scaled scores in cohort 1 decreased by 1.46 points, a 0.67 scaled score point increase from model 3. Similar to model 3, adding school covariates to the model shows a decrease in scores from 2019 to 2021. Covariates that had significant interactions with 2021 cohort included female, Hispanic and Black ethnicities, SES disadvantaged, disability, LEP, and a location of rural, indicating that the effects for these covariates differed between 2019 and 2021. The only significant location variable in 2019 was suburb with suburban schools scoring 2.96 points higher than city schools. The interaction of cohort and suburb was not significant, showing that the effect of schools being in the suburb was not different between 2019 and 2021. Although the effect due to rural school location was not significant in 2019, it was significant in 2021. In 2019, rural schools scored 0.77 points lower than city schools, but in 2021, rural schools scored 3.29 points higher than city schools. Only the location of rural was significantly different in cohort 1 compared to cohort 0 and the effect was positive, contradicting the hypothesis to research question 4 that city and rural schools would have greater losses in scaled scores. School type was not significant in either 2019 or 2021, showing that public and private schools tended to score about the same in 2019 and 2021. This also opposed the hypothesis in research question 4 that public schools would be more negatively affected in 2021 than non-public schools.

Math.

When school level predictors were added, the differences due to cohort increased to 2021 students performing 15.80 scaled score points lower than 2019 students (Table 6). Similar to the model 3 results, adding the school covariates supported the hypothesis for research question 2 that 2021 students would score lower than 2019 students. Covariates that had significantly different effects in 2021 compared to 2019 were female gender; ethnicities of Hispanic and Black; SES disadvantaged; disability; LEP; and school locations of suburb, town, and rural. Like model 3, all student level covariates except an ethnicity of Hispanic were significant. The effects of the student level covariates in model 4 are similar to those in model 3 with only the effects from the ethnicity of Black showing a change. In model 3, Black students in 2021 scored 16.43 points lower than white students in 2021, but when school covariates are included in model 4, that decreases to 13.89. Two school covariates of location and school type were included in model 4. All locations except for a location of other had significant effects in 2019 with students in suburb, town, and rural scoring higher than city students. Suburban students scored the highest with an increase of 4.82 points in 2019 and increasing that gap by 4.62 points to 9.44 scaled scores higher in 2021 than city students. Students living in towns had similar results to suburb with a difference of 4.19 points higher in 2019 and 8.48 points in 2021. Rural students scored 3.27 points higher in 2019 than city students, but that increased an additional 9.34 points in 2021. The covariate of public school type was not significant in either cohort, showing that public and non-public schools had no statistically significant difference in scaled scores.

Table 5

ELA multilevel mixed effect model results

ELA	Student			Student+School		
Predictors	Estimates	SE	*p*	Estimates	SE	*p*
(Intercept)	3.61	0.39	**<0.001**	2.73	0.72	**<0.001**
Y1SS.center	0.65	0.00	**<0.001**	0.65	0.00	**<0.001**
Gender [Female]	6.56	0.24	**<0.001**	6.55	0.24	**<0.001**
Ethnicity [Hispanic]	0.49	0.44	0.267	0.26	0.45	0.557
Ethnicity [Black]	-6.54	0.52	**<0.001**	-6.80	0.54	**<0.001**
Ethnicity [Asian]	7.42	0.68	**<0.001**	7.17	0.68	**<0.001**
Ethnicity [Other]	-2.57	0.57	**<0.001**	-2.66	0.57	**<0.001**
SES [Disadvantaged]	-5.50	0.28	**<0.001**	-5.39	0.29	**<0.001**
Disability [Yes]	-16.71	0.39	**<0.001**	-16.66	0.39	**<0.001**
LEP [Yes]	-6.20	0.61	**<0.001**	-6.20	0.61	**<0.001**
Cohort [1]	-0.79	0.37	**0.034**	-1.46	0.57	**0.011**
Gender [Female] * Cohort[1]	-4.02	0.42	**<0.001**	-4.03	0.42	**<0.001**
Ethnicity [Hispanic] *Cohort [1]	-2.73	0.73	**<0.001**	-2.00	0.75	**0.007**
Ethnicity [Black] *Cohort [1]	-4.61	0.76	**<0.001**	-3.55	0.82	**<0.001**
Ethnicity [Asian] *Cohort [1]	-1.51	1.13	0.180	-0.72	1.14	0.528
Ethnicity [Other] *Cohort [1]	-0.22	0.97	0.816	0.16	0.97	0.870
SES [Disadvantaged] *Cohort [1]	-1.88	0.48	**<0.001**	-2.15	0.48	**<0.001**
Disability [Yes] * Cohort[1]	4.47	0.66	**<0.001**	4.40	0.66	**<0.001**
LEP [Yes] * Cohort [1]	-4.09	1.04	**<0.001**	-4.00	1.04	**<0.001**
Location [Suburb]				2.96	1.08	**0.006**
Location [Town]				1.62	1.05	0.125
Location [Rural]				-0.77	0.92	0.403
Location [Other]				-1.92	3.61	0.595
Public [No]				3.76	3.43	0.274
Location [Suburb] *Cohort [1]				-0.64	0.62	0.302
Location [Town] * Cohort[1]				0.08	0.66	0.905
Location [Rural] * Cohort[1]				4.06	0.70	**<0.001**
Location [Other] * Cohort[1]				0.33	3.15	0.915
Public [No] * Cohort [1]				-1.46	2.96	0.621
Random Effects						
σ^2	868.38			867.82		
τ_{00}	71.30 $_{\text{SchoolNumber}}$			70.45 $_{\text{SchoolNumber}}$		
ICC	0.08			0.08		
N	912 $_{\text{SchoolNumber}}$			912 $_{\text{SchoolNumber}}$		
Observations	91335			91335		
Marginal R^2 / Conditional R^2	0.611 / 0.641			0.614 / 0.643		

Table 6

Math multilevel mixed effect model results

Math	**Student**			**Student+School**		
Predictors	Estimates	SE	*p*	Estimates	SE	*p*
(Intercept)	9.87	0.52	**<0.001**	6.87	0.98	**<0.001**
Y1SS.center	0.72	0.00	**<0.001**	0.72	0.00	**<0.001**
Gender [Female]	4.28	0.29	**<0.001**	4.27	0.28	**<0.001**
Ethnicity [Hispanic]	0.08	0.52	0.874	-0.24	0.53	0.644
Ethnicity [Black]	-9.54	0.62	**<0.001**	-10.05	0.64	**<0.001**
Ethnicity [Asian]	7.97	0.80	**<0.001**	7.47	0.80	**<0.001**
Ethnicity [Other]	-2.61	0.66	**<0.001**	-2.78	0.66	**<0.001**
SES [Disadvantaged]	-7.02	0.33	**<0.001**	-7.04	0.33	**<0.001**
Disability [Yes]	-18.93	0.46	**<0.001**	-18.92	0.46	**<0.001**
LEP [Yes]	-5.18	0.71	**<0.001**	-5.35	0.71	**<0.001**
Cohort [1]	-11.10	0.44	**<0.001**	-15.80	0.67	**<0.001**
Gender [Female] * Cohort[1]	-3.91	0.49	**<0.001**	-3.90	0.49	**<0.001**
Ethnicity [Hispanic] *Cohort [1]	-6.78	0.85	**<0.001**	-5.03	0.88	**<0.001**
Ethnicity [Black] *Cohort [1]	-6.89	0.90	**<0.001**	-3.84	0.98	**<0.001**
Ethnicity [Asian] *Cohort [1]	-2.40	1.33	0.070	-0.44	1.34	0.740
Ethnicity [Other] *Cohort [1]	-2.36	1.14	**0.038**	-1.57	1.14	0.169
SES [Disadvantaged] *Cohort [1]	-2.71	0.56	**<0.001**	-2.60	0.57	**<0.001**
Disability [Yes] * Cohort[1]	5.09	0.77	**<0.001**	5.10	0.78	**<0.001**
LEP [Yes] * Cohort [1]	-4.83	1.21	**<0.001**	-4.20	1.21	**0.001**
Location [Suburb]				4.82	1.51	**0.001**
Location [Town]				4.19	1.46	**0.004**
Location [Rural]				3.27	1.27	**0.010**
Location [Other]				7.31	4.81	0.128
Public [No]				-3.08	4.56	0.500
Location [Suburb] * Cohort [1]				4.62	0.73	**<0.001**
Location [Town] * Cohort [1]				4.29	0.78	**<0.001**
Location [Rural] * Cohort[1]				9.34	0.83	**<0.001**
Location [Other] * Cohort[1]				-2.55	3.70	0.491
Public [No] * Cohort [1]				4.37	3.47	0.209
Random Effects						
σ^2	1195.76			1193.97		
τ_{00}	151.89 SchoolNumber			146.18 SchoolNumber		
ICC	0.11			0.11		
N	913 SchoolNumber			913 SchoolNumber		
Observations	91325			91325		
Marginal R^2 / Conditional R^2	0.589 / 0.636			0.598 / 0.642		

Figure 2

ELA moderator effect of gender on year 2 scaled score based on cohort

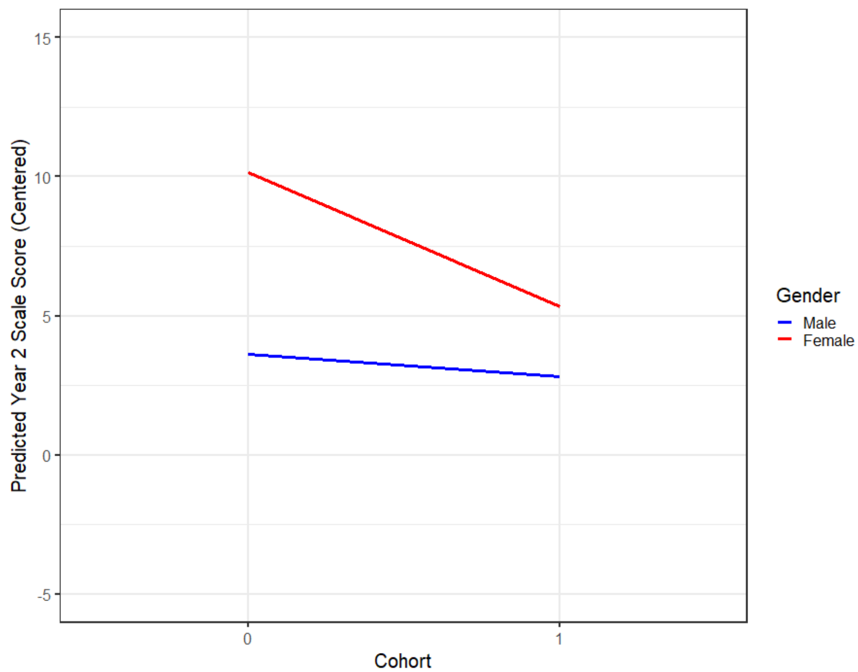


Figure 3

Math moderator effect of gender on year 2 scaled score based on cohort

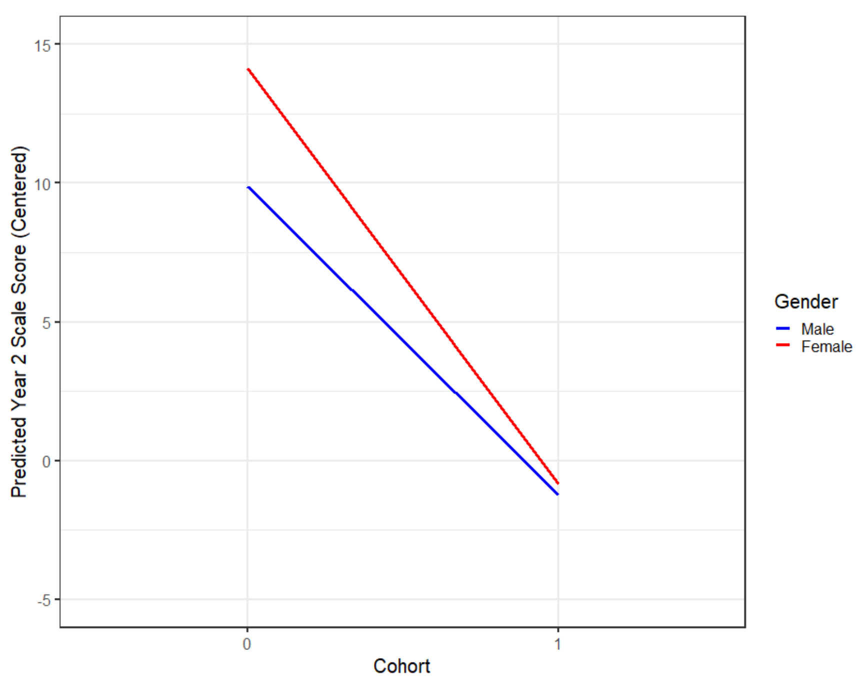


Figure 4

ELA moderator effect of ethnicity on year 2 scaled score based on cohort

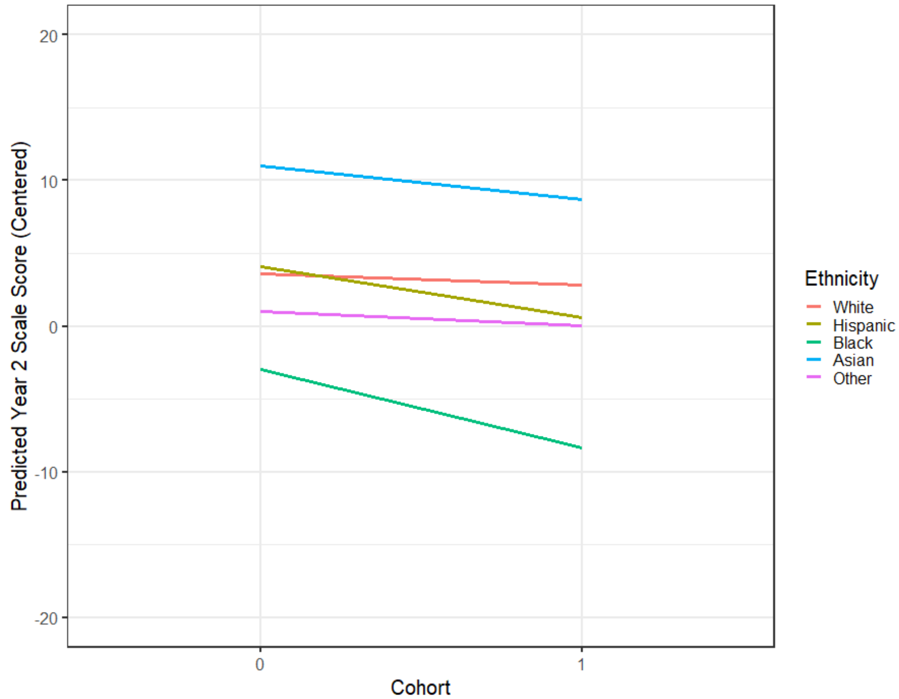
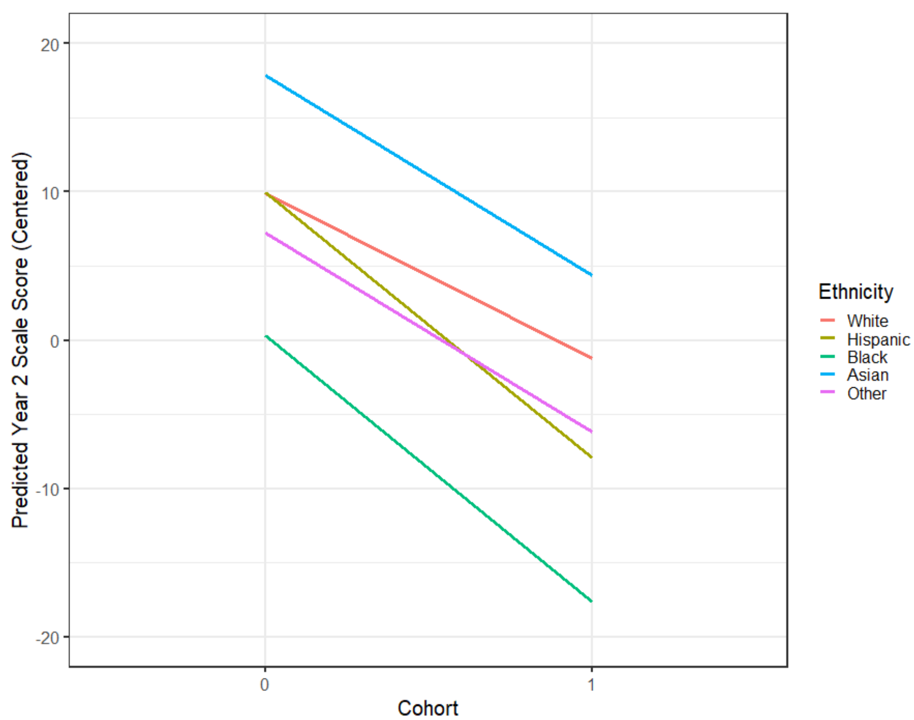


Figure 5

Math moderator effect of ethnicity on year 2 scaled score based on cohort



Discussion

The COVID-19 pandemic caused disruptions in education that were detrimental to student learning. Comparisons of scores from large scale state assessments can provide insight into the extent of the learning loss that occurred, and by using demographic covariates, can help identify how learning loss affected differing groups. In this study, scaled scores from a pre-pandemic year, 2019, was compared to scaled scores from a pandemic year, 2021, using multilevel regression models. The main predictor variable was cohort, which represented the year the assessment was taken. Student and school covariates were also included to help explain the differences seen between cohorts.

The first model was built to answer research question 1. When comparing the scaled scores using a simple linear regression, the average score dropped by 4.86 points for ELA and 11.21 points for math between 2019 and 2021. This difference in means was expected for both ELA and math as learning loss was expected due to the pandemic. The scaled score range for math was smaller than ELA (395 vs. 590), so the larger drop in math signifies even a greater learning loss.

Model 3 was built to address research questions 2 and 3. Model 3 results show that in ELA, females scored 6.56 points higher than males in 2019 but decreased to 2.54 points in 2021. The same trend occurs for math with the gap between female and male performance narrowing in 2021. Figure 2 shows that for ELA, the effect for male students in 2021 was only slightly lower than for 2019, while the effect for females is larger. Along with education, the pandemic caused disruptions to schedules and lifestyles, and Acosta and Evans (2020) and Bateman and Ross (2020) purport that the trend seen may be due to females being expected to take on more household responsibilities. Although female students are still performing better than males, their learning gains may have declined, as depicted in Figures 2 and 3. The trendline in the figures also only compares two cohorts, and if the effect continued at the same rate, scaled scores of female students would decline to become lower than males. Black and Hispanic students had the greatest drop in scaled scores for both ELA and math while scores also declined for SES disadvantaged students. The studies from Kuhfeld et al. (2020), Lewis et al. (2021), and NWEA (2021) looked at learning loss from a formative assessment and their results are in line with this study that the effects of the pandemic on education was more pronounced for lower SES and minority ethnic groups, especially Black and Hispanic students. These results are more concerning when the effects are compounded by being both a minority and an SES disadvantaged student. Although, this study showed that learning loss may be more affected by ethnic minority status than by SES status. LEP students also performed lower in 2021 than 2019 and their effect was larger than the effect from SES, but LEP students tend to also be from a lower SES so the effects can be compounded. The effects from model 3 for math for all covariates were more negative than for ELA. The effect of cohort 1 was -0.79 for ELA but was -11.10 for math. Although the score range on the assessments were different, math had a narrower score range, so the effects were even more pronounced. The larger drops in math scores between the cohorts showed that the learning losses in math were more substantial than for ELA. The results from this study parallel the findings from Lewis et al. (2021) that saw double the learning loss in math compared to reading.

Model 4 included the school covariates of school location and school type to answer research questions 2 and 4. The effect of public schools was not significant in either ELA or math, contradicting the hypothesis that public schools would perform lower than non-public schools due to the latter being allowed to continue in-person learning (Dickler, 2020). One possible explanation is that in March 2020, the Coronavirus Aid Relief and Economic Security (CARES) Act provided funding to public and charter schools, and this additional funding may have mitigated the differences between school type during the pandemic (Office of Elementary and Secondary Education, 2020). For school location, rural schools had significantly different scores in 2021 in both ELA and math, and rural schools performed significantly higher than city schools in 2021. This was unexpected as rural schools were assumed to have less internet connectivity (Fishbane et al., 2020). One possible explanation for these results is that rural schools have lower populations than city schools, so they may have been able to continue or return to in-person learning during the pandemic sooner than city schools. Second, city schools have a higher rate of SES, LEP, and ethnic minorities (e.g. Figure 1), so this may have caused lower observed scaled scores in the city schools. For math, suburb, town, and rural schools all performed better than city schools.

Results of the model fit statistics show the model that included both the student and school predictors is a better fit. Model 4 for ELA only had one significant school covariate (interaction of rural and cohort), so the more parsimonious model 3 may be more beneficial to use. Public school type was not significant for either subject, so this variable was omitted, and an ANOVA was run to test the fit between model 4 with and without school type. The results of the ANOVA showed that the full model with school type had better fit, so it was included as a covariate. The effects of the student level covariates did not differ much when the school covariates were included, so either model 3 or 4 could be used to assess the effect of student covariates.

Although this study was able to detect differences in performance on a large scale state assessment, there are limitations to this study. First, this study is performed on only one large scale assessment program, so the results may be difficult to generalize to other student populations as school closures and impact may differ by region (Parks et al., 2021). Second, no data on remote learning or school closures was gathered, so there is no direct study of school closures as the reason for lower scaled scores. This was also not a study on causal relationships, so even though the results showed differences in effects, the pathway for how learning loss occurred for each covariate is unknown. In addition, this study was done on two separate groups of grade 6 cohorts, so it was not possible to measure the learning loss within students over years. It is possible that scaled scores decreased when comparing the 2019 students to the 2021 students, but the students themselves could still have made learning gains as in Lewis et al. (2021). Although changes in learning due to COVID-19 were prominent between 2019 and 2021, there may be other reasons unrelated to the pandemic that may have caused decrease in learning. This analysis focused on student and school level nesting, but decisions may come from the district level. Adding the district as a third level into the multilevel model may yield different results. This analysis also focused on the differences due to cohort, but in future studies, the interaction of the covariate variables may provide more information, such as whether there is a further effect of being a minority and LEP.

Overall, this study showed a decrease in scaled scores in both ELA and math for students who took the grade 6 assessment in 2021 compared to 2019. Scores from females, minority, SES-disadvantaged, and LEP students saw more negative decreases while students with disability saw scores increase from 2019. Results on school location differed, but rural schools did not have the scaled score decrease as expected in 2021. There were no differences in scores in either 2019 or 2021 for public and non-public schools.

Works Cited

Bateman, N., & Ross, M. (2020, October 2). Why has COVID-19 been especially harmful for working women? Brookings. https://www.brookings.edu/essay/why-has-covid-19-been-especially harmful-for-working-women/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Choi, T. H., & Chiu, M. M. (2021). Toward equitable education in the context of a pandemic: supporting linguistic minority students during remote learning. *International Journal of Comparative Education and Development*.

Council of Chief State School (CCSSO). (2020, July 14). *RESTART & RECOVERY: Considerations for Teaching and Learning Overview*. https://ccsso.org/coronavirus.

Dickler, J. (2020, November 8). *Families jump to private schools as coronavirus drags on.* CNBC. https://www.cnbc.com/2020/11/08/coronavirus-why-families-are-jumping-to-private-schools.html

Education Week. 2020. *Map: Coronavirus and School Closures*. https://www.edweek.org/leadership/map-coronavirus-and-school-closuresin-2019-2020/2020/03

Fishbane, L., & Tomer, A. (2020, March 20). *As classes move online during COVID-19, what are disconnected students to do?*. Brookings. https://www.brookings.edu/blog/the-avenue/2020/03/20/as-classes-move-online-during-covid-19-whatare-disconnected-students-to-do/

Mendez Acosta, A., & Evans, D. (2020, October 2). *COVID-19 and Girls' Education: What We Know So Far and What We Expect*. Center for Global Development. https://www.cgdev.org/blog/covid-19-and-girls-education-what-we-know-so-far-and-what-we-expect-happen

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142. doi:10.1111/j.2041-210x.2012.00261.x

NWEA. (2021). *Academic Growth for Students with Disabilities: Lessons from school-year learning gains and summer learning loss*. https://www.nwea.org/content/uploads/2021/06/Academic-Growth-for-Students-with-Disabilities-Lessons-from-school-year-learning-gains-1.pdf

Kuhfeld, M., Ruzek, E., Lewis, K., Soland, J., Johnson, A., Tarasawa, B., & Dworkin, L. (2020). *Understanding the initial educational impacts of COVID-19 on communities of color*. NWEA. https://www.nwea.org/content/uploads/2021/06/Understanding-the-initial-educational-impacts-of-COVID19-on-communities-of-color.Resear-hBrief.pdf

Lewis, K., Kuhfeld, M., Ruzek, E., & McEachin, A. (2021). Learning during COVID-19: *Reading and math achievement in the 2020-21 school year*. NWEA. https://www.nwea.org/content/uploads/2021/07/Learning-during-COVID-19-Reading-and-math-achievement-in-the-2020-2021-school-year.research-brief-1.pdf

Office of Elementary and Secondary Education. (2020). *U.S. Elementary and Secondary School Emergency Relief Fund*. https://oese.ed.gov/offices/education-stabilization-fund/elementary-secondary-school-emergency-relief-fund/

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation*, 7(1), 1-3.

Parks, S., Zviedrite, N., Budzyn, S.E., Panaggio, M.J., Raible, E., Papazian, M., Magid, J., Ahmed, F., Uzicanin, A., & Barrios, L.C. (2021). *COVID-19–Related School Closures and Learning Modality Changes —United States, August 1–September 17, 2021*. CDC. https://www.cdc.gov/mmwr/volumes/70/wr/pdfs/mm7039e2-H.pdf

Raudenbush, S. W. Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, second edition*. Newbury Park, CA: Sage.

U.S. Department of Education. (2022). *U.S. Department of Education Announces Distribution of All American Rescue Plan ESSER Funds and Approval of All 52 State Education Agency Plans*. https://www.ed.gov/news/press-releases/us-department-education-announces-distribution-all-american-rescue-plan-esser-funds-and-approval-all-52-state-education-agency-plans

Wu, J., Smith, S., Khurana, M., Siemaszko, C., DeJesus-Banos, B. (2020, April 29). *Stay-at-home orders across the country.* NBC News. https://www.nbcnews.com/health/health-news/here-are-stay-home-orders-across-country-n1168736