Pandemic Learning Loss by Student Baseline Achievement: Extent and Sources of Heterogeneity

Ian Callen

Dan Goldhaber

Thomas J. Kane

Anna McDonald

Andrew McEachin

Emily Morton

February 2024

WORKING PAPER No. 292-0224

CALDER
National Center for Analysis of
Longitudinal Data in Education Research

AIR®

Pandemic Learning Loss by Student Baseline Achievement: Extent and Sources of Heterogeneity

Ian Callen
American Institutes for Research / CALDER

Dan Goldhaber
American Institutes for Research / CALDER
University of Washington / CEDR

Thomas J. Kane
Harvard University

Anna McDonald
American Institutes for Research / CALDER

Andrew McEachin
NWEA

Emily Morton
American Institutes for Research / CALDER

Contents

Acknowledgments

This research was supported by the Carnegie Corporation of New York, the Walton Family Foundation, Kenneth C. Griffin, and an anonymous foundation. CALDER is funded by a consortium of foundations. We are grateful to James Cowan, Roddy Theobald, and Elise Dizon-Ross for assistance with various aspects of this work.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

Pandemic Learning Loss by Student Baseline Achievement: Extent and Sources of Heterogeneity

Ian Callen, Dan Goldhaber, Thomas J. Kane, Anna McDonald, Andrew McEachin, & Emily Morton

CALDER Working Paper No. 292-0224
February 2024

Abstract

It is now well established that the COVID-19 pandemic had a devastating and unequal impact on student achievement. Test score declines were disproportionately large for historically marginalized students, exacerbating preexisting achievement gaps and threatening educational and economic inequality. In this paper, we use longitudinal student-level NWEA MAP Growth test data to estimate differences in test score declines for students at different points on the pre-pandemic test distribution. We also test the extent to which students' schools and districts accounted for these differences in declines. We find significant differences in learning loss by baseline achievement, with lower-achieving student's scores dropping 0.100 SD more in math and 0.113 SD more in reading than higher-achieving students' scores. We additionally show that the school a student attended accounts for about three-quarters of this widening gap in math achievement and about one-third in reading. The findings suggest school and district-level policies may have mattered more for learning loss than individual students' experiences within schools and districts. Such nuanced information regarding the variation in the pandemic's impacts on students is critical for policymakers and practitioners designing targeted academic interventions and for tracking disparities in academic recovery.

1. Introduction

Since the outbreak of COVID-19 in 2020, school closures and other disruptions have had a devastating impact on student learning (Kuhfeld, Soland, and Lewis, 2022; Goldhaber et al., 2023; Kogan and Lavertu, 2022; U.S. Department of Education, 2023). Test score declines have been disproportionately large for students in high-poverty schools and for Black and Hispanic students (e.g., Goldhaber et al., 2022a, Kilbride et al., 2022; Lewis et al., 2022). The academic fallout from COVID-19 threatens to have long-run economic impacts on labor markets and economic productivity (Werner and Woessmann, 2023) and is likely to increase inequality in college and labor market outcomes into the future (e.g., Darling-Aduana, Schachner, and Edgerton 2020; Kane et al., 2022).

In response, school districts across the country have launched unprecedented efforts to recover academically, supported in part by federal funds from the American Rescue Plan's Elementary & Secondary School Emergency Relief (ESSER) program. The recovery efforts have included district-wide initiatives (e.g., curriculum updates, staff professional development, building improvements) as well as initiatives targeted at subsets of struggling students (e.g., high-dosage tutoring, school-based interventions, after-school programs). Unfortunately, the academic recovery has been sluggish so far. The 2021-22 and 2022-23 school years show that the pace of recovery appears to be slow, on average, and targeted interventions have served few students with little effectiveness to date (Carbonari et al., 2022; Callen et al., 2023; Halloran et al., 2023; Lewis and Kuhfeld, 2023). The unequal impacts of the pandemic remain two years after students returned to in-person learning in the fall of 2021 (Lewis and Kuhfeld, 2023; Kogan, 2023).

We build on prior research that has examined widening achievement gaps by race and economic disadvantage at the district-level by using a student-level panel to study losses by baseline academic achievement. We document the degree to which losses varied for students with different levels of pre-pandemic achievement. We also study the degree to which those losses were driven by school or district-level declines in achievement, as opposed to within-school and within-district widening. The organization level at which the widening occurred (whether it be state, district, school or within-school) offers clues regarding the mechanisms of learning losses. For example, if district-level closure decisions or the quality of schools' or districts' implementation of hybrid learning were the source of the loss, we would expect to see much of the decline happening at the schools and districts enrolling large shares of low-achieving students. However, if the source of the heterogeneity were differential access to household resources, such as broadband internet or the availability of a parent to oversee remote learning or quiet study space at home, we would expect to see more of the variation in loss within school and district. Such evidence would also be useful in planning recovery efforts, as it could inform where recovery efforts are focused: on district-wide improvement efforts, on school-wide efforts, or on subgroups of students within schools.

In this paper, we investigate heterogeneity in learning loss by baseline achievement using longitudinal student-level data from over 2.1 million students in over 10,000 schools across 49 states (and D.C.). By doing so, we extend earlier work examining national changes in students' test score gains (i.e., growth) during the pandemic at the extremes (10^{th} and 90^{th} percentiles) of the pre-pandemic achievement distribution (Peters et al., 2023) by estimating the degree to which test achievement and growth distributions in fall 2021 declined across the entire baseline test distribution. We also build on Fahle et al.'s (2023) study, which found that pandemic-related achievement declines were similar across race and economic disadvantage within districts. In contrast to that study, which used district-by-subgroup mean achievement, we use student-level data to compare "learning loss"[1] for students with high and low baseline achievement, within-schools and districts as well as between them.

We find significant differences in average learning loss by students' prior achievement level, with larger losses for students with lower pre-pandemic achievement. Moreover, using school and district fixed effects specifications and the decomposition methodology described in Gelbach (2016), we show that about three-quarters of the widening between top and bottom quintile students in math was associated with the school students attended, rather than widening gaps within schools. In other words, we found losses were larger in the schools (and, to a lesser extent, districts) attended by lower-scoring students, as opposed to finding differential losses for high and low-achieving students within schools and districts. In reading, where the average losses were smaller, about one-third of the widening by baseline achievement was associated with schools/districts. We speculate that the difference in the math and reading results may reflect the differential role of schools and families in math and reading growth.

Our results provide the first large-scale evidence suggesting that variation in learning loss between lower- and higher-performing students may be best explained by the school a student attended. In math (but not reading), the results suggest that school and district-level policies (and related characteristics) may have mattered *more* than individual students' experiences within their school for students' growth during the pandemic.

[1] Note that "learning loss" in this context does not necessarily imply that individual students' raw test scores decreased over time or that they necessarily "lost" content knowledge; rather, we use the term "learning loss" to refer to the differences between students' two-year test score gains and their expected two-year test score gains based on historical averages and typical patterns for students in the same grade with similar prior achievement.

2. Evidence on Disparities in Pandemic-Related Learning Loss

All U.S. public schools closed during the spring of the 2019-20 school year, and many remained closed (or operated hybrid[2] models) for much of the 2020-21 school year (Goldhaber et al., 2023b; Jack and Oster, 2023). Given the sudden shift to remote instruction and the broader economic and public health impacts of the COVID-19 pandemic, it is no surprise that numerous studies now show that students made substantially less academic progress during the pandemic relative to previous years.[3] Moreover, research shows that high-poverty and high-minority school districts remained closed for longer (Goldhaber et al., 2023b; Jack et al., 2023).

Test score patterns on state assessments (Fahle et al., 2023; Kogan & Lavertu, 2022), as well as benchmark assessments at the district, state, and national levels, document the negative impact of the pandemic on student learning (Darling-Aduana et al., 2022; Dorn, Hancock, and Sarakatsannis, 2021; Education Policy Innovation Collaborative [EPIC], 2021; Goldhaber et al. 2023a; Lewis et al., 2022; Lewis and Kuhfeld, 2021). There is evidence that these test score drops persist through the 2022-23 school year; the most recent National Assessment of Educational Progress (NAEP) long-term trend results show that the average math score of 13-year-old students in the fall of 2022 was the lowest since 1990 (USDE, 2023). Despite two school years of largely in-person school schedules, test scores in spring 2023 remained significantly lower than pre-pandemic scores across all grades in mathematics and reading, with the largest relative declines occurring in mathematics and earlier elementary grades (Lewis and Kuhfeld, 2023). Similar findings have been reported across the world; Betthaüser et al. (2023)

[2] Hybrid models typically consisted of shared in-person and online learning experiences.
[3] The effect of schooling interruptions (outside of the pandemic) on student achievement is well studied. Regularly scheduled breaks, such as summer (e.g., McEachin & Atteberry, 2017; Atteberry & McEachin, 2020; Gershenson, 2013) and transitional breaks (e.g., Slade et al., 2017) have been found to negatively impact student test scores. Even unscheduled events like closures brought upon by weather (e.g., Marcotte and Hemelt, 2008) and labor disputes (e.g., Baker, 2013; Wills, 2019) have been shown to negatively impact students.

reports a consistent pattern of widening achievement across 15 countries, with larger losses among more disadvantaged students and larger losses in low- and middle-income countries relative to wealthy nations.

The studies that estimate average impacts of the pandemic on student learning are important, but the averages potentially obscure important evidence on disparate impacts on subsets of students and the mechanisms behind them. A number of studies find that students attending high-poverty schools and Black and Hispanic students experienced the greatest losses (e.g., Camp & Zamarro, 2022; Kogan, 2023; Kuhfeld, Soland, and Lewis, 2022; Parolin & Lee, 2021; U.S. Department of Education, 2023).[4] The NAEP long-term trend cross-sectional analysis also shows larger declines in the 10th and 25th percentile test scores than the 75th and 90th percentiles, particularly in math (U.S. Department of Education, 2023). However, such data do not make it possible to parse the extent to which differences in pandemic test score declines by prior achievement over time are related to changing compositions of students in each cross-section (e.g., Dee and Murphy, 2023). More recent evidence that leverages longitudinal student-level NWEA MAP Growth test score data further suggests that students who were the lowest performing prior to the pandemic (who scored at or below the 10th percentile) experienced much larger declines in growth during the pandemic (relative to pre-pandemic norms) than their high-performing (90th percentile or above) peers (Peters et al., 2023). Nevertheless, little is known about how these declines in growth rates relative to pre-pandemic rates (i.e., "learning losses") varied across the full distribution of pre-pandemic achievement.

[4] For example, Kuhfeld et al. (2022) estimate that, relative to white students' NWEA MAP Growth test scores, Hispanic and Black students' scores respectively declined by an *additional* 0.08 and 0.11 standard deviations in math and 0.04 and 0.07 standard deviations in reading from fall 2019 to fall 2021.

Further examining variation in learning losses at the school- and district-level enhances our understanding of the variation in losses between students of the same district and schools. Several studies find that the proportion of the 2020-21 school year that a school district operated remotely (or offered hybrid instruction) was predictive of their students' achievement or growth declines (Fahle et al., 2023; Goldhaber et al., 2023b; Jack et al., 2023). These studies further show that a district's instructional modality accounted for much of the variation in achievement and growth declines by race and school-level poverty. For example, Goldhaber et al. (2023b) find that learning losses were generally greater for students attending higher poverty schools, but students attending higher and lower poverty schools lost about the same amount of ground when they returned to in-person instruction quickly (and had less remote or hybrid instruction) in 2020-21. Similarly, Jack et al. (2023) find that districts with higher proportions of Black students were less likely to have access to in-person schooling, amplifying the disparate incidence of pandemic learning loss. Taken together, these prior studies suggest that school- and district-level policies and characteristics may have played an important role in pandemic-related learning loss.

Fahle et al. (2023) provide the most comprehensive analysis to date of the variation in district-level learning loss by race and socioeconomic status within districts across 30 states. They find that disparities in achievement level declines between student subgroups within districts were small, implying that the mechanisms driving overall declines operated at the district or community level (i.e., broadband access, disruptions to social and economic activity, and trust in government institutions). Unfortunately, beyond the length of school closures, few of the other community-level predictors they tested provided much explanatory power.[5] While this study substantially advances our understanding of variation in pandemic-related declines, it is

[5] Losses were somewhat larger in areas with greater levels of disruption to families social and economics lives and in areas with higher COVID-19 death rates.

also limited by its use of aggregated district-level data. As a result, the authors are unable to examine heterogeneity in learning loss by baseline achievement (Werner and Woessmann, 2023).

As policymakers and practitioners grapple with the enduring challenge of academic recovery, a detailed understanding of how pandemic losses vary across students, schools, and districts will be critical for targeting academic recovery policies and programs at the students and communities most in need. To that end, the present study leverages longitudinal student-level data to contribute novel evidence about the heterogeneity in pandemic-related learning loss across the pre-pandemic achievement distribution, and the extent to which enrollment in different schools and districts accounts for learning losses.

3. Data Description and Sample Characteristics

The data used in this study come from the Growth Research Database at NWEA, the assessment division of Houghton Mifflin Harcourt (HMH). Roughly 3,000 school districts across 49 states partner with NWEA to administer its adaptive[6] Measures of Academic Progress (MAP) assessments three times (fall, winter, and spring) during the 2019-20 school year. Importantly, while some remote testing occurred during the pandemic, nearly all MAP Growth tests were administered in-person at the students' schools in our sample. Relative to fixed-form (i.e., nonadaptive) tests, adaptive assessments are designed to capture achievement more precisely at the high and low ends of the achievement distribution (Kingsbury, Nesterak and Freeman, 2014). MAP test scores, referred to as "RIT scores" are calculated using the Rasch item response theory (IRT) model and the tests are scaled vertically so that scores can be compared across grades. To allow for test distributions to be compared on a common scale, we standardize scores within each

[6] The MAP assessments are computer-adaptive, meaning the difficulty of exam questions changes in response to a student's performance.

grade-, subject-, and instructional week using the means and standard deviations from the pre-pandemic NWEA MAP Growth norms (Thum & Kuhfeld, 2020).[7] We refer to these scores as *"norms-standardized"* scores.

Our sample includes MAP Growth test scores for Grades 3 through 8 from three terms: fall 2017, fall 2019, and fall 2021.[8] We restrict our sample to the set of district-grades that test at least 60% of their enrolled students in all three terms to provide confidence that our estimates are not driven by the selection of students into testing.[9] We also incorporate sample restrictions for the purpose of measuring student growth over time. At the core of our analyses, we examine how students' test score growth changes over two years during the pre-pandemic period (i.e. fall 2017 to fall 2019) and the pandemic period (i.e. fall 2019 to fall 2021). To be included in the analysis, we require that: individual students have RIT tests for both a baseline year (i.e., fall 2017 or fall 2019) and a follow-up test two years later (i.e., fall 2019 or fall 2021); and that schools test 10 of the same students and districts test 100 of the same students over each two-year period. This additional restriction results in an analysis sample that includes more than 3.4 million students in 10,440 schools and 2,312 districts.[10] This sample comprises approximately 17 percent of public schools (and 20 percent of districts)[11] serving students in Grades 3–8 in the United States.[12]

[7] The means and standard deviations were estimated pooling data over three school years, 2015-16, 2016-17, and 2017-18.
[8] When examining students' 2-year growth (e.g., growth between fall 2017 and fall 2019), we include baseline data from Grades 1 and 2 to estimate growth for students in Grades 3 and 4.
[9] These restrictions omit approximately 0.5% of districts and 1.3% of schools from the NWEA universe of MAP tests.
[10] As we describe below, only students and districts who used MAP assessments pre-pandemic (fall 2017 and fall 2019) are included in the longitudinal sample during the pandemic. This is because we compare within-district pre-pandemic growth to within-district growth observed during the pandemic.
[11] The results that follow are consistent when we apply more restrictive sample criteria (i.e., requiring at least ten students per racial subgroup).
[12] The analysis sample is smaller than samples used in the aforementioned national cross-sectional achievement research (USDE, 2023), but consistent with Lewis and Kuhfeld's (2021) study.

We find that the overall 2-year attrition rate, defined as the percentage of students who took a test in a baseline year (i.e. fall 2017 or fall 2019) and did not take a test in a follow-up year (i.e. fall 2019 or fall 2021), is greater from fall 2019 to fall 2021 than from fall 2017 to fall 2019. Specifically, among the students who tested in the 2017 and 2019 baseline years, we observe respective 2-year attrition rates of 23 percent and 30 percent in mathematics and 25 percent and 37 percent in reading. Some attrition is expected because students who move districts (to a district that does not use MAP tests) would not test in the follow-up year. Unfortunately, the data do not enable us to distinguish whether students who did not test in a term are missing data because they were not enrolled in the school that year or because they were enrolled and did not take the test. However, attrition out of the samples did not vary significantly based on students' race or their achievement in the baseline year, as shown in Table 1, in which we report descriptive statistics for two groups of students.[13] Demographic characteristics are similar between these samples for both math and reading. The norms-standardized test scores of the students in the longitudinal samples are slightly higher in both math and reading.[14] This difference could be explained by the fact that mobile students are more likely to have lower test scores (Goldhaber et al., 2022b).

The descriptive test score patterns for the cross-sectional and longitudinal samples provide strong suggestive evidence that the pandemic impacted students' learning trajectories.

[13] That attrition rates did not vary based on prior achievement is consistent with Lewis and Kuhfeld's (2021) findings but contrasts with other previous research (e.g., Austin et al., 2021). One potential explanation for this finding could be that schools were more motivated than usual to test their low-achieving students in fall 2021, following the pandemic. Note also that the samples are also demographically very similar to the universe of students as defined by national data from the CCD, though our sample has a smaller proportion of Hispanic students and a larger proportion of students who identify as a race other than Asian, Black, Hispanic, or White. Although the NWEA sample of students is quite large and demographically resembles the population of U.S. public school students, it is important to note that all schools and districts in the sample choose to partner with NWEA, so they are distinct from schools not included in the sample in that way.

[14] For example, the mean Fall 2017 standardized math score for the full sample 0.08, compared to 0.11 for students in the full sample.

Comparing across years, we find that norms-standardized scores were noticeably lower on average during the pandemic (i.e., fall 2021) than in prior fall terms. For instance, these scores for the longitudinal sample are significantly lower in fall 2021 (-0.12 SD for math, -0.17 SD for reading) compared to fall 2019 (0.10 SD for math, 0.06 SD for reading).

In Table 2, we use transition matrices to describe differences in achievement based on prior achievement level. Each cell in the table shows the percentage of students in each longitudinal sample who score in the same norms-standardized quintile or move to a different quintile for each pair of baselines (i.e., 2017 and 2019) and follow-up years (i.e., 2019 and 2021). Panels A and B display these matrices for math scores, and Panels C and D display these matrices for reading scores.

While most students remain in the same quintile from year-to-year (e.g., students in the lowest quintile in fall 2017 remain in the lowest quintile in fall 2019), a meaningful portion of students transition to a different quintile. Furthermore, across pre-pandemic years, transition matrices are largely symmetric, meaning students were just as likely to transition up a quintile as they were to transition down a quintile.[15] But the situation is quite different when we focus on the transitions between 2019 and 2021 (Panels B and D). Across these years we observe that students were much more likely to move to a lower quintile than they were pre-pandemic. For instance, 38% of students scoring in the second quintile in math in 2019 scored in the lowest quintile two years later compared to 21% of students who scored in the second quintile in math in 2019. Interestingly, the increase in the percentage of students who move to a lower quintile during the pandemic (relative to pre-pandemic rates) is similar across different baseline quintiles.

[15] While total rows and columns of the matrices sum 100, individual rows and columns are not evenly distributed. This is because scores are standardized outside of the individual test distribution which causes quintile ranks to be uneven within individual years.

4. Empirical Strategy

We begin the analysis with simple descriptions of changes in norms-standardized test achievement across fall terms for different student subgroups in the longitudinal samples. As noted above, our standardization of achievement data using pre-pandemic norms allows for meaningful comparisons of student achievement between years. While it is clear that norms-standardized achievement declined during the pandemic, focusing on the average difference across years potentially misses important variation in these differences across students that could inform policymakers' and educators' academic recovery efforts. Therefore, we quantify the variation in learning loss along the prior achievement distribution and examine whether the school and/or district a student attended helps to explain variation in learning loss.

We begin by estimating students' fall 2019 norms-standardized achievement separately for each district and subject, accounting for students' fall 2017 norms-standardized achievement:

$$Y_{ij2019} = \alpha_{j2019} + \sum_{\tau=1}^{3} \beta_\tau Y_{(ij2017)}^{\tau} + \gamma X_{ij2017} + \varepsilon_i \quad (1)$$

where Y_{i2019} is the norms-standardized achievement score for student i in district j in fall 2019 and α_j captures the average fall 2019 norms-standardized achievement score for the omitted categorical groups (white 8th grade students) within each district j.[16] The term $\sum_{\tau=1}^{3} Y_{(ij2017)}^{\tau}$ includes the main effect, square, and cubic of students' lagged norms-standardized test score from the baseline testing period (fall 2017). Taken altogether, the estimates for β_τ represent average pre-pandemic growth patterns for students at various points along the prior test distribution.[17] $\boldsymbol{X}_{ij2017}$ is a vector of student-characteristics (i.e., students' race/ethnicity and

[16] Since equation (1) is estimated within districts, cell sizes for various factors (namely race) may be small.
[17] In Appendix B, we also estimate variants of equation (1) that use alternate specifications for prior achievement and obtain similar results.

grade level in 2017) that may be differentially related to growth patterns.[18] ε_i represents robust standard errors estimated at the student level.

Equation (1) estimates the relationship between norms-standardized test scores over a two-year period before the pandemic. We use the parameter estimates from (1) to predict what we would have expected students' norms-standardized achievement to be in fall 2021, had growth followed the same patterns observed prior to the pandemic: $\hat{Y}_{ij2021} = \hat{\alpha}_j + \sum_{\tau=1}^{3} \hat{\beta}_\tau Y_{(ij2019)}^{\tau} + \hat{\gamma}\boldsymbol{X}_{(ij2019)}$. In Equation (2), we then compare the predictions to students' observed norms-standardized achievement:[19]

$$G_{ij2021} = Y_{ij2021} - \left(\hat{\alpha}_{j2019} + \sum_{\tau=1}^{3} \hat{\beta}_\tau Y_{(ij2019)}^{\tau} + \hat{\gamma}\boldsymbol{X}_{(ij2019)} \right) \tag{2}$$

G_{ij2021} represents the difference between student *i's* observed 2021 norms-standardized achievement and their predicted 2021 norms-standardized achievement, as estimated from pre-pandemic growth patterns, conditional on observable characteristics. If $G_{ij2021} < 0$, we can infer students performed below pre-pandemic growth expectations, implying some degree of learning loss. We refer to estimates of G_{ij2021} as learning loss hereafter.

To examine variation in learning loss across different points in the prior achievement distribution, we estimate the following equation separately for math and reading:

$$G_{ij2021} = \alpha_s + \sum_{\substack{\tau=1 \\ \tau\neq 5}}^{\tau=4} \delta_\tau Q_{ij2019}^{\tau} + \theta\boldsymbol{X}_{ij2021} + \varepsilon_{ij} \tag{3}$$

[18] It is important to note that because equation (1) estimates pre-pandemic growth within a single two-year time period, time-varying student and school characteristics cannot be included in the model.
[19] Since we estimate pre-pandemic within districts, this means only students in districts that were included in the estimation of (1) are included in this analysis.

where i indexes students, s indexes schools, and j indexes districts. Q_{ij2019}^{τ} is the observed norms-standardized quintile rank of a student's score in fall 2019.[20,21] In the model above, the primary coefficients of interest are δ_τ. These coefficients measure the difference in learning loss between a given quintile and the omitted (highest) quintile. For example, a δ_τ equal to 0 implies that students who scored in quintile τ in 2019 experienced a similar amount of learning loss as students who scored in the highest quintile in 2019, conditional on observable covariates (such as race/ethnicity, grade-level, and MAP testing date).[22] Estimates of δ_τ that are significantly different than zero would alternatively suggest that learning loss is not uniform across prior achievement levels, with negative estimates implying that lower-scoring students experienced greater learning loss. We run this model three times: without fixed effects, with district fixed effects (α_j, not shown above), and with school fixed effects (α_s, shown above). We include these fixed effects to account for differences between districts or schools (e.g., the percent of students eligible for free or reduced-price lunch) that may influence students' learning loss. Standard errors are clustered at the district level (Bertrand, Duflo and Mullainathan 2004; Abadie et al. 2023).

To understand the potential sources of variation in student growth patterns during the pandemic, we use an order-invariant decomposition described in Gelbach (2016) to account for the fact that the covariates in the model may be correlated with each other. This approach

[20] We extend this specification in two ways in Appendix B. First, we interact a student's 2019 norms-standardized quintile rank with their 2019 norms-standardized achievement score to account for differences within quintiles. Second, we estimate this specification including the cubic of a students' 2019 norms-standardized achievement score. Results are consistent across specifications.
[21] We also extend this specification by interacting a student's 2019 norms-standardized quintile rank with race to examine the extent to which the relationship between prior achievement and learning loss varied by race. The coefficients on the interaction terms were not significantly different from zero. Results are available upon request.
[22] Testing dates are included in this specification to account for variation in testing dates that may influence the magnitude of learning loss. There is limited variation in test dates within districts. As a result, they are excluded from growth estimation models in equation (1).

considers each factor as an "omitted variable" when estimating the relationship between prior achievement and learning loss and measures the variation that each factor would explain in a model. The approach proposed by Gelbach (2016) is helpful in our context as it can decompose how much of learning loss is due to a set of covariates. For example, we can estimate how much of the variation in learning loss is driven by the school a student attended relative to a student's race.

5. Results

5.1 Changes in Test Achievement Over Time

We report simple kernel density plots documenting the distribution of norms-standardized student achievement across Grades 3-8 in mathematics and reading for our cross-sectional samples of students tested in fall 2017, 2019, and 2021 (Figure 1).[23] Differences across years illustrate the extent to which distributions in achievement have shifted over time. Panels A (math) and B (reading) of Figure 1 show the distributions in math and reading hardly differ in the years prior to the pandemic. However, there is a large shift to the left in fall 2021: the median score decreases by about 0.24 SD in mathematics and 0.13 SD in reading.[24] These declines are larger for elementary grade-level tests (see Appendix Figure A1) than middle school grade-level tests (see Appendix Figure A2).[25] Additionally, we show that changes in median achievement

[23] We show that the fall 2018 distribution is consistent with the fall 2017 and fall 2019 distributions in Goldhaber et al. (2022a).
[24] One way to put the magnitudes of these changes in context is to compare them to changes in student achievement associated with Hurricanes Katrina and Rita in Louisiana. As Sacerdote (2012) reports, students who were displaced by the hurricanes experienced declines in test scores between 0.07 and 0.20 standard deviations. In other words, the changes we observe from the NWEA national data are of the same order of magnitude (indeed larger) to what has been considered large negative impacts associated with hurricane-related disruptions.
[25] We provide formalized testing in Appendix Tables A1 and A2 that test for differences in mean norms-standardized achievement between fall 2017, fall 2019, and fall 2021. Though we find substantially larger differences between the pre-pandemic scores and the 2021 scores than between the 2017 and 2019 scores, all differences are statistically significant due to large sample sizes.

scores between fall 2019 and fall 2021 are largest for Black and Hispanic students (see Appendix Figure A3).[26]

We further show the changes in norms-standardized achievement for our cross-sectional samples across the fall terms in Appendix Table A3. In these tables, we report the 10th, 25th, 50th, 75th, and 90th percentile (normalized within year and subject) norms-standardized test scores for our sample in the fall of each year.[27] These tables help expand our understanding of the impact of the pandemic on students' achievement across the distribution. For example, a 10th percentile norms-standardized math test score in 3rd grade in 2021 is 0.44 SD lower than the 10th percentile score in 2019. However, the 90th percentile 2021 score is only 0.14 SD lower than the respective 2019 score. These findings are consistent with prior work using cross-sectional data (and different test score data or metrics) to examine declines in pandemic-related test scores by prior achievement (e.g., Dorn, Hancock, and Sarakatsannis, 2021; EPIC, 2021; Kogan and Lavertu, 2022, Lewis and Kuhfeld, 2021; U.S. Department of Education, 2023).

The preceding findings replicate prior evidence of declines in student achievement *levels*. However, as noted above, these findings may not be indicative of how the pandemic impacted students at different levels of prior achievement because they are based on cross-sections of students and could be related to changes in the composition of the tested populations of students (Wosemann and Werner, 2023). Recent evidence, for instance, shows that there are significantly more students being homeschooled and more students who are unaccounted for after the pandemic (e.g., Dee and Murphy, 2023). In the next section, we examine differences in students' growth during the pandemic compared to that of similar students who attended the same school

[26] The results presented in Appendix Figure A3 are similar when using fall 2017 test scores as the pre-pandemic comparison year.
[27] We present these results separately for each grade for math and reading respectively in Appendix Tables A4 and A5.

district during the pre-pandemic period to hold constant some of the possible factors related to both changes in schooling contexts and test scores.

5.2 Heterogeneity in Student Growth Along Prior Performance Distribution

As outlined in Equation (2), we compare observed growth during the pandemic to predicted growth, using the parameter estimates from Equation (1) to measure how pandemic growth deviated from pre-pandemic growth (which we refer to as learning loss). Table 3 (panel A for math, panel B for reading) displays the variation in learning loss across students in different pre-pandemic norms-standardized achievement quintiles. Column 1 shows unadjusted estimates of the relationship between students' fall 2019 quintile ranks and their learning loss.[28] The constant term in column 1 can be interpreted as average learning loss among students in the highest quintile (0.20 SD in math and 0.06 SD in reading), and the coefficients on each other 2019 achievement quintile can be interpreted as the additional learning loss experienced by students in that quintile relative to students in the highest quintile. The coefficient estimates for each prior achievement quintile in both math and reading models indicate that lower-achieving students experienced substantially more learning loss than higher-achieving students. For example, students who scored in the first (lowest) quintile in 2019 experienced an additional 0.10 SDs of learning loss in math (0.30 SDs total) and an additional 0.12 SDs of learning loss in reading (0.18 SDs total) relative to students who scored in the fifth (highest) quintile in 2019.

The next set of analyses adds student characteristics as covariates and district or school fixed effects to the regression (see Table 3 columns 2, 3, and 4 and Figure 2). These models

[28] We note that the R-squared value in Table 3 Column 1 indicates that, alone, a student's prior norms-standardized achievement quintile explains very little (0.4%) of the variation in students' learning loss. This finding is not surprising given the results in the present study that show that: (1) the school a student attended explains a substantial portion of the variation, (2) student characteristics explain very little of the variation, and (3) much of the variation remains unexplained even after controlling for these factors.

allow us to examine the influence of district- and school-level factors (e.g. resources, practices, policies, politics, etc.) on the relationship between prior achievement and pandemic learning loss. Were it the case that the school and district a student attended were largely unrelated to their learning loss, we would expect to see little difference in the coefficient estimates across these different specifications. As it turns out, we find that the fixed effects models that account for differences in learning loss between schools and districts significantly attenuate the differences in learning loss across prior achievement quintiles (and more so in the school fixed effects model than the district fixed effects model). While these differences in learning loss by prior achievement quintile within schools are much smaller than the same differences across schools, it is worth noting that previously lower-performing students still experienced more learning loss than higher-performing students within schools, and the remaining differences within schools are larger in reading than in math.

5.3 Decomposition of Student Growth

Our results suggest that differences in learning loss by students' prior norms-standardized achievement levels can be explained, in part, by the district or schools in which students are enrolled. With our next analysis, we endeavor to formally test the extent to which a student's school, race, grade, and test date each explain variation in learning loss using the approach described in Gelbach (2016). Conceptually, this approach treats each factor as an "omitted variable" in the relationship between learning loss and prior achievement levels and measures the bias that would result if the factor were excluded. This approach is of particular value when these factors are correlated; for instance, the different racial compositions of schools mean that the school a student attends is related to their propensity to be a given race, such that controlling for a student's school may also capture something about their race. To utilize a Gelbach decomposition, we specify a “base” specification that excludes measures of student characteristics (X_{is2021}) and school fixed effects (α_s), while the “full” model is specified as equation (3).

In Figure 3, we plot the proportion of additional learning loss experienced by students in each fall 2019 baseline norms-standardized achievement quintile (relative to students in the highest quintile) for math (panel A) and reading (panel B). The net height of the bar represents the estimated unconditional average additional learning loss (in absolute value) experienced by that quintile relative to the highest baseline achievement quintile. Estimates greater than zero (shown in green) indicate an increase in the difference relative to the unconditional average difference between quintiles.

The decomposition estimates (see Table 4) are consistent with prior studies that find significant heterogeneity in pandemic learning loss associated with the districts that students attended (Fahle et al., 2023; Goldhaber et al., 2023b). Across all quintile comparisons, schools explain a significant portion of the differences in learning loss between prior achievement quintiles in both subjects, but they play a particularly significant role for math. In math, they explain 67 to 81 percent of the variation in learning loss between quintiles, whereas in reading they explain only 24 to 37 percent of the variation. Consistent with Fahle et al. (2023), these findings suggest that, especially in math, students in the same schools experienced a relatively similar magnitude of learning loss, with significant variation between schools. Controlling for each of the following factors: students’ race, grade levels, and testing dates, generally does not explain differences in learning loss between quintiles or slightly amplifies the differences in loss between prior achievement quintiles (e.g., students in higher grades have slightly larger differences between students by prior quintile than students in lower grades) for either subject.

Finally, our findings also highlight a challenge for school systems in responding to students' needs in the aftermath of the pandemic. In both subjects, but especially reading, and for all racial groups, a substantial portion of the differences in learning loss across prior achievement quintiles is left unexplained. In other words, students' unique experiences within their school explain a significant portion of how far they are behind where we would expect them to be in absence of the pandemic. To achieve full academic recovery from the pandemic, school systems are not only tasked with implementing systematic programs and policies that reach a majority of their students, but they also have to address unique student needs that vary across students within schools.

6. Discussion

Prior research has documented large disparities in learning loss by poverty and race. We add to that growing body of research by showing that learning loss also varied by baseline achievement: the variance in academic achievement widened because students who started out with lower achievement lost the most ground. We also find that schools account for nearly three quarters of the widening gaps by baseline achievement in math, albeit less than one-third of the variation in reading. This finding is consistent with the hypothesis that achievement growth in math is more associated with school-based activities, while growth in reading reflects a combination of home and school-based activities (e.g., Burgess, Rawal, and Taylor, 2023; Riehl and Welch, 2022).

More broadly, these results indicate that academic recovery from the pandemic will require a mix of school- and district-level initiatives as well as interventions that respond to students' unique needs within these contexts. Targeted student-level interventions may be particularly important to support students' recovery in reading.

The findings in this paper support the notion that returning U.S. public schools just to pre-pandemic levels of inequality in achievement will not be easy. It will require significant school- and district-wide learning acceleration and/or supplementation of learning opportunities for the majority of schools that were negatively impacted in addition to supplemental supports for the lower-achieving Black and Hispanic students who were disproportionately impacted within these schools. If not remedied, the disproportionate impacts of the pandemic on disadvantaged students have the potential to exacerbate long-term social and economic inequality in the U.S. (Kane et al., 2022). In our pursuit of academic recovery for all students, and particularly for those whose disadvantage has been exacerbated by the pandemic, we must leverage our nuanced understanding of where and for whom the impacts were largest to efficiently and adequately target support.

References

Atteberry, A. C., & McEachin, A. J. (2020). Not Where You Start, but How Much You Grow: An Addendum to the Coleman Report. *Educational Researcher*, 49(9), 678-685.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. "When should you adjust standard errors for clustering?." *The Quarterly Journal of Economics* 138, no. 1 (2023): 1-35.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?." *The Quarterly journal of Economics* 119, no. 1 (2004): 249-275.

Betthaüser, B.A., Bach-Mortensen, A.M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*.

Burgess, Simon, Shenila Rawal, and Eric S. Taylor. "Teachers' use of class time and student achievement." *Economics of Education Review* 94 (2023): 102405.

Callen, I., Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., ... & Staiger, D. O. (2023). Summer School as a Learning Loss Recovery Strategy After COVID-19: Evidence From Summer. Working Paper No. 291-0823. *National Center for Analysis of Longitudinal Data in Education Research (CALDER).*

Camp, A. M., & Zamarro, G. (2022). Determinants of ethnic differences in school modality choices during the COVID-19 crisis. *Educational Researcher*, *51*(1), 6-16.

Clearinghouse, W. W. (2012). What works clearinghouse. *Internet site: http://ies. ed. gov/ncee/wwc*.

Carbonari, M. V., Davison, M., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., ... & Staiger, D. O. (2022). The Challenges of Implementing Academic COVID Recovery Interventions: Evidence from the Road to Recovery Project. Working Paper No. 275-1222. *National Center for Analysis of Longitudinal Data in Education Research (CALDER).*

Darling-Hammond, L., Schachner, A., & Edgerton, A. K. (2020). Restarting and Reinventing School: Learning in the Time of COVID and Beyond. *Learning Policy Institute.*

Darling-Aduana, Jennifer, Henry T. Woodyard, Tim R. Sass, and Sarah S. Barry. "Learning-mode choice, student engagement, and achievement growth during the COVID-19 pandemic." *AERA Open* 8 (2022): 23328584221128035.

Dee. T., & Murphy. M. (2021). Patterns in the Pandemic Decline of Public School Enrollment. *Educational Researcher*.

Dorn, E., Hancock, B., & Sarakatsannis, J. (2021). *COVID-19 and education: The lingering effects of unfinished learning*. McKinsey & Company.

Evans, W.N., Schwab, R.M. and Wagner, K.L., 2019. The Great Recession and public education. *Education Finance and Policy*, *14*(2), pp.298-326.

Education Policy Innovation Collaborative (EPIC). (2021). *K–8 Student achievement and achievement gaps on Michigan's 2020–21 benchmark and summative assessments*.

Fahle, E. M., Kane, T. J., Patterson, T., Reardon, S. F., Staiger, D. O., & Stuart, E. A. (2023). School district and community factors associated with learning loss during the COVID-19 pandemic. *Center for Education Policy Research at Harvard University: Cambridge, MA, USA*.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much?. *Journal of Labor Economics, 34(2)*, 509-543.

Goldhaber, D., Kane, T. J., McEachin, A., & Morton, E. (2022a). *A comprehensive picture of achievement across the COVID–19 pandemic years: Examining variation in test levels and growth across districts, schools, grades, and students*. (CALDER Working Paper No. 266–0522).

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2023b). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights*, 5(3), 377-392.

Goldhaber, D., Koedel, C., Özek, U., & Parsons, E. (2022c). Using longitudinal student mobility to identify at-risk students. AERA Open, 8, 233285842110710.

Halloran, C., Hug, C. E., Jack, R., & Oster, E. (2023). Post COVID-19 Test Score Recovery: Initial Evidence from State Testing Data (No. w31113). *National Bureau of Economic Research.*

Jack, R., Halloran, C., Okun, J., & Oster, E. (2023). Pandemic schooling mode and student test scores: evidence from US school districts. *American Economic Review: Insights*, *5*(2), 173-190.

Jack, Rebecca, and Emily Oster. "COVID-19, School Closures, and Outcomes." *Journal of Economic Perspectives* 37, no. 4 (2023): 51-70.

Jackson, C.K., Wigger, C. and Xiong, H., 2021. Do school spending cuts matter? Evidence from the Great Recession. *American Economic Journal: Economic Policy*, *13*(2), 304-35.

Kane, T., Doty, E., Patterson, T., Staiger, D. (2022). What do changes in state test scores imply for later life outcomes? *Cambridge, MA: Center for Education Policy Research.*

Kilbride, T., Hopkins, B., Strunk, K., & Yu, D. (2022). *Michigan's fall 2021 benchmark assessments*. Education Policy Innovation Collaborative.

Kingsbury, G. G., Nesterak, M., & Freeman, E. (2014). The potential of adaptive assessment. *Education Leadership*, 71(6).

Kogan, V. (2023). Student achievement and learning acceleration on spring 2023 Ohio state tests. *The Ohio State University.*

Kogan, V & Lavertu, S. (2022). How the COVID-19 pandemic affected student learning in Ohio: Analysis of spring 2021 Ohio state tests. *The Ohio State University.*

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test score patterns across three COVID-19-impacted school years. *Educational Researcher*, 51(7), 500-506.

Lewis, K., & Kuhfeld, M. (2023). Education's Long COVID: 2022-23 Achievement Data Reveal Stalled Progress toward Pandemic Recovery. *Center for School and Student Progress at NWEA.*

Lewis, K., Kuhfeld, M., Langi, M., Peters, S., & Fahle, E. (2022). The widening achievement divide during COVID-19.

Parolin, Z., & Lee, E. K. (2021). Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature human behaviour*, *5*(4), 522-528.

Peters, Scott J., Meredith Langi, Megan Kuhfeld, and Karyn Lewis. (2023). Unequal Learning Loss: How the COVID-19 Pandemic Influenced the Academic Growth of Learners at the Tails of the Achievement Distribution. (EdWorkingPaper: 23-787).

Riehl, Evan, and Meredith Welch. "Accountability, test prep incentives, and the design of math and English exams." *Journal of Policy Analysis and Management* 42, no. 1 (2023): 60-96.

Sacerdote, B. (2012). When the saints go marching out: Long-term outcomes for student evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, 4(1), 109-135.

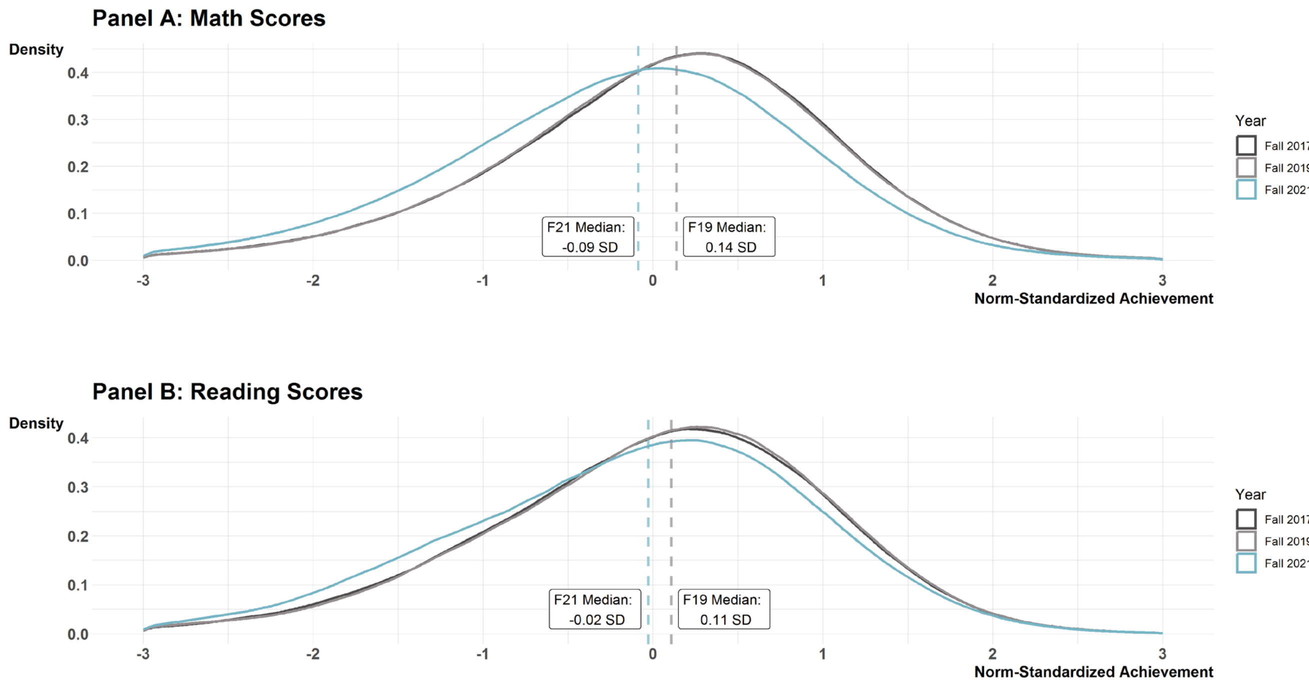
Thum, Yeow Meng, and Megan Kuhfeld. "NWEA 2020 MAP growth achievement status and growth norms for students and schools." *NWEA Research Report*. (2020).

U.S. Department of Education. (2023). *National Assessment of Educational Progress (NAEP) long-term trend assessment results: reading and mathematics*. Institute of Education Sciences, National Center for Education Statistics.

Werner, K., & Woessmann, L. (2023). The legacy of COVID-19 in education. *Economic Policy*..

Figures and Tables

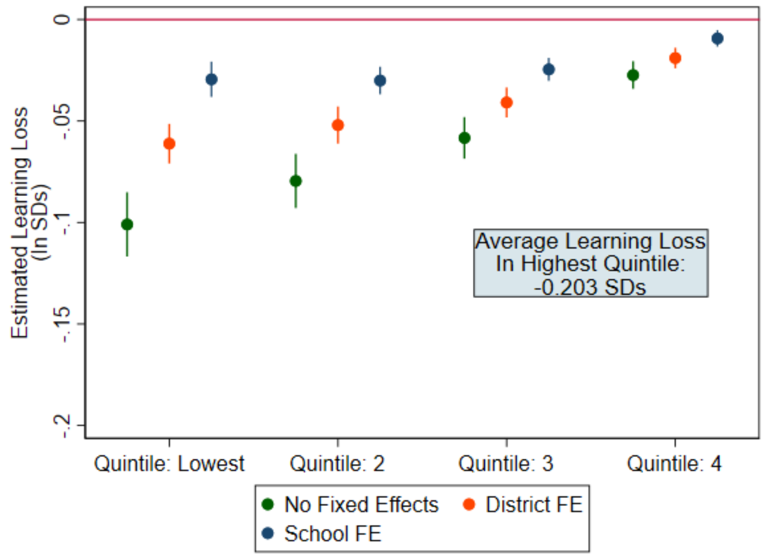
Figure 1. Distribution of Math and Reading Test Scores, Grades 3-8



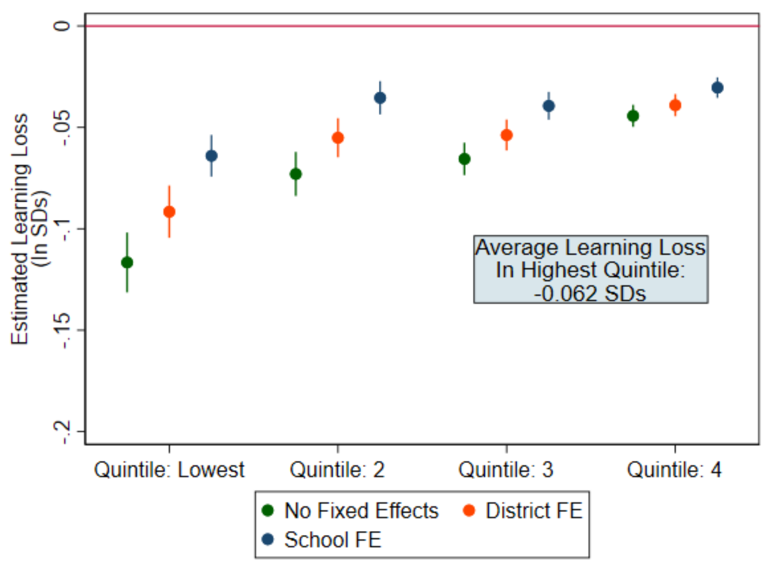
Note: These figures show kernel density plots for tests taken in fall 2017, 2019, and 2021 in math (panel A) and reading (panel B). Scores are normalized to MAP Growth norms reported in Thum and Kuhfeld, 2020.

Figure 2. Differences in Learning Loss Between Prior Achievement Quintiles

Panel A: Math



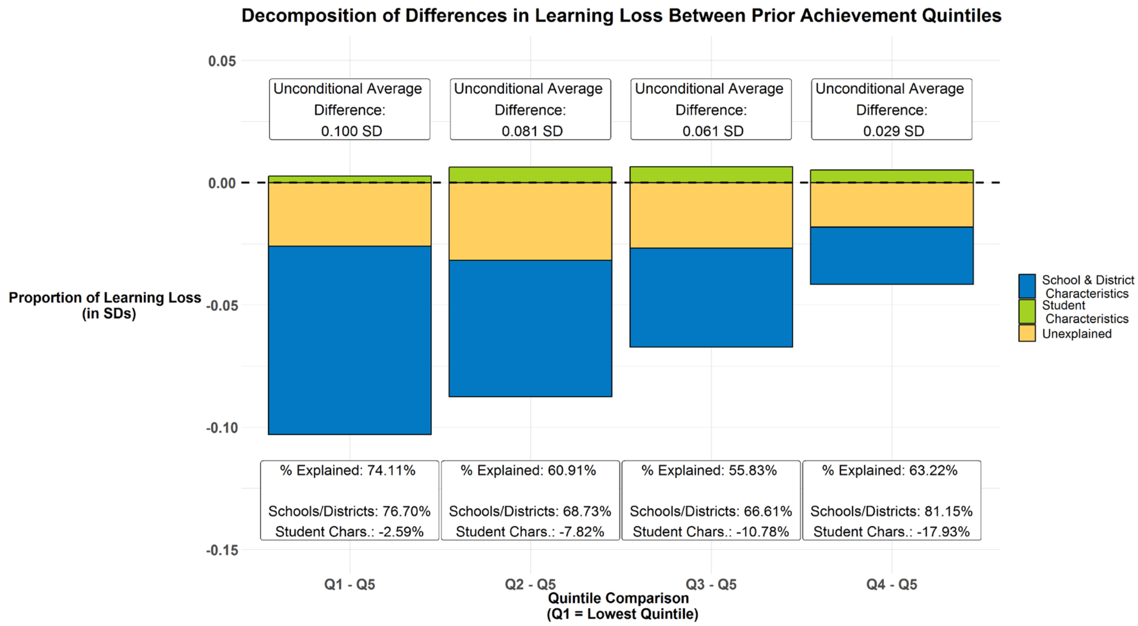
Panel B: Reading



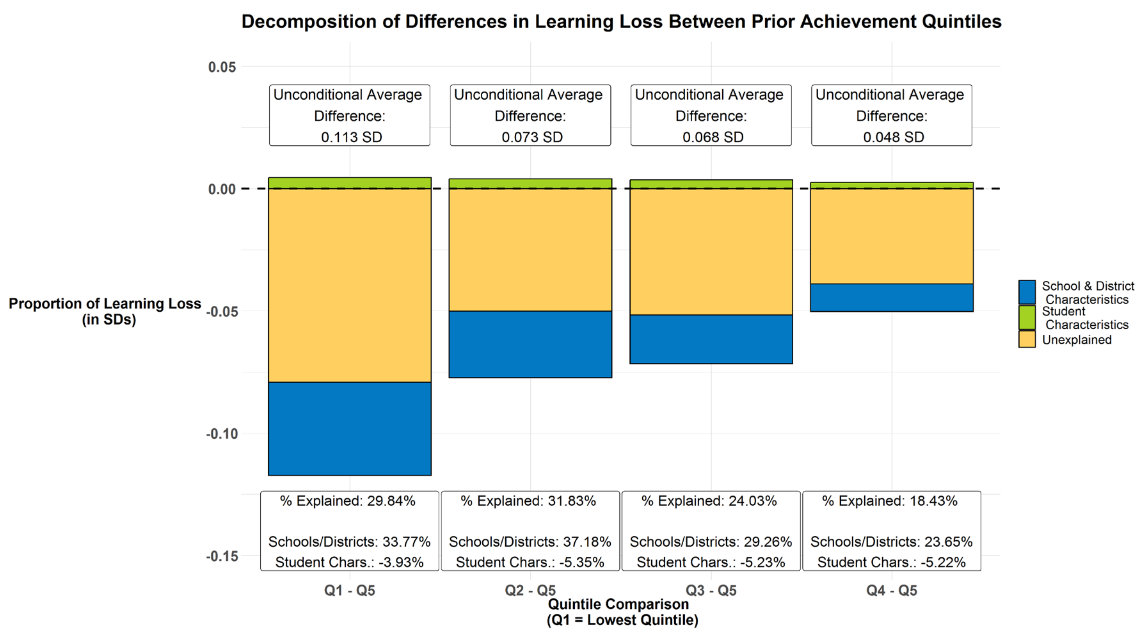
Note: The above figure compares point estimates from equation (3) with and without district and/or school fixed effects for differences in learning loss (G_{i2021}) by prior achievement quintiles relative to the highest quintile. Results for math are presented in panel A and results for reading are presented in panel B. Covariates in all models include student grade levels, race/ethnicity, and testing dates. 95% confidence intervals are calculated using robust standard errors clustered on the district level. If an estimate is statistically different than zero, we feel confident that learning loss for that prior achievement quintile was different than learning loss for the highest quintile. A negative and statistically significant estimate implies that the reference quintile experienced a higher incidence of learning loss as compared to the highest prior achievement quintile. Estimates correspond to estimates reported in Table 3, columns 2 through 4.

Figure 3. Decomposition of Gaps in Learning Loss Between Prior Achievement Quintiles Students

Panel A: Math



Panel B: Reading



Note: The above figures report estimates using the method described by Gelbach (2016) to plot the share of differences in learning loss (G_{i2021}) between prior achievement quintiles by each factor we consider: student characteristics (grade level, testing dates, and race) and school fixed effects. The net height of the bar represents the total unconditional estimate of learning loss differences between quintiles. All differences are relative to the fifth quintile (the highest). Positive values indicate that estimates of differences increase after accounting for these factors. The percent of difference explained is calculated relative to the unconditional average difference between quintiles.

Table 1. Descriptive Statistics by Year and Test Subject

Panel A: Descriptive Statistics for 2017-2019 Samples						
	Math			Reading		
	2017 Students	2019 Students	2017-2019 Longitudinal Sample	2017 Students	2019 Students	2017-2019 Longitudinal Sample
Norms-standardized score (F17)	0.09		0.09	-0.02		0.02
Norms-standardized score (F19)		0.13	0.13		0.01	0.03
% Asian	0.04	0.04	0.04	0.04	0.04	0.04
% Black	0.15	0.15	0.14	0.18	0.17	0.16
% Hispanic	0.18	0.19	0.19	0.19	0.20	0.19
% White	0.51	0.50	0.52	0.47	0.47	0.51
Unique districts	3,092	3,082	2,215	3,083	3,080	1,859
Unique schools	14,519	14,859	9,885	14,451	14,799	8,607
Unique students	4,941,889	5,156,259	2,341,440	4,253,731	4,710,384	1,631,558
Panel B: Descriptive Statistics for 2019-2021 Samples						
	Math			Reading		
	2019 Students	2021 Students	2019-2021 Longitudinal Sample	2019 Students	2021 Students	2019-2021 Longitudinal Sample
Norms-standardized score (F19)	0.05		0.09	0.01		0.01
Norms-standardized score (F21)		-0.18	-0.13		-0.12	-0.10
% Asian	0.04	0.04	0.04	0.04	0.04	0.04
% Black	0.15	0.15	0.14	0.17	0.16	0.16
% Hispanic	0.19	0.20	0.19	0.20	0.21	0.19
% White	0.50	0.50	0.52	0.47	0.47	0.51
Unique districts	3,082	3,079	2,156	3,080	3,075	1,793
Unique schools	14,859	14,354	9,747	14,799	14,102	8,451
Unique students	5,156,259	4,786,412	2,178,893	4,710,384	4,204,188	1,532,676

Note: This table includes only test observations that occurred at schools that tested at least 10 students, in districts that test at least 100 students in each of the three terms, and in schools where at least 60% of the within-grade population was tested in a given term. The math sample titled "2017 Students" includes all students who took a math MAP test in Fall 2017 who attended schools/districts that met minimum testing requirement. All test scores are standardized using the pre-pandemic NWEA MAP Growth norms.

Table 2. Quintile Transition Matrix for Fall 2017-Fall 2019 and Fall 2019-Fall 2021 MAP Growth Scores

Panel A: Quintile Transition Matrix for Math MAP Scores; 2017-2019

		Quintile Rank in 2019					
		Lowest	2nd	3rd	4th	Highest	Total
Quintile Rank in 2017	Lowest	**10.22**	3.99	1.45	0.46	0.08	16.21
	2nd	3.53	**6.79**	4.75	1.61	0.21	16.89
	3rd	1.02	4.68	**8.37**	5.82	1.01	20.91
	4th	0.22	1.37	5.78	**11.47**	5.57	24.41
	Highest	0.03	0.12	0.83	5.20	**15.39**	21.58
	Total	15.02	16.96	21.19	24.57	22.27	100.00

Panel B: Quintile Transition Matrix for Math MAP Scores; 2019-2021

		Quintile Rank in 2021					
		Lowest	2nd	3rd	4th	Highest	Total
Quintile Rank in 2019	Lowest	**12.96**	2.56	0.69	0.20	0.04	16.44
	2nd	6.50	**6.87**	2.91	0.76	0.09	17.14
	3rd	2.52	7.14	**7.45**	3.35	0.44	20.90
	4th	0.59	3.13	8.01	**9.55**	2.99	24.28
	Highest	0.06	0.34	1.81	6.71	**12.32**	21.23
	Total	22.63	20.04	20.86	20.57	15.89	100.00

Panel C: Quintile Transition Matrix for Reading MAP Scores; 2017-2019

		Quintile Rank in 2019					
		Lowest	2nd	3rd	4th	Highest	Total
Quintile Rank in 2017	Lowest	**11.42**	4.83	2.28	0.82	0.14	19.50
	2nd	4.09	**5.91**	5.02	2.25	0.36	17.63
	3rd	1.66	4.35	**7.10**	5.52	1.13	19.77
	4th	0.52	1.75	5.50	**9.93**	4.47	22.16
	Highest	0.09	0.25	1.22	5.77	**13.61**	20.94
	Total	17.78	17.10	21.12	24.28	19.72	100.00

Panel D: Quintile Transition Matrix for Reading MAP Scores; 2019-2021

		Quintile Rank in 2021					
		Lowest	2nd	3rd	4th	Highest	Total
Quintile Rank in 2019	Lowest	**13.36**	4.10	1.71	0.60	0.11	19.88
	2nd	5.49	**5.99**	4.24	1.72	0.28	17.72
	3rd	2.42	5.09	**6.83**	4.57	0.87	19.78
	4th	0.80	2.37	6.16	**9.28**	3.55	22.17
	Highest	0.12	0.36	1.55	6.17	**12.26**	20.46
	Total	22.20	17.92	20.48	22.33	17.07	100.00

Note: This table reports quintile transition matrices for student scores in Fall 2017/Fall 2019 and Fall 2019/Fall 2021 for math and reading respectively. The numbers reported in the matrices can be interpreted as the percentage of students in each longitudinal sample who score in the same quintile or move to a different quintile for each pair of baseline (i.e., 2017 and 2019) and follow-up years (i.e., 2019 and 2021). Quintile ranks are calculated using the NWEA MAP Growth Norms (Thum & Kuhfeld, 2020).

Table 3. Learning Loss Differences by Prior Achievement Quintile

Panel A: Math				
	(1)	(2)	(3)	(4)
Constant	-0.203***	-0.146***	-0.164***	-0.180***
	(0.007)	(0.007)	(0.007)	(0.006)
Fall 2019 Quintile Rank: 1 (Lowest)	-0.101***	-0.066***	-0.052***	-0.027***
	(0.008)	(0.007)	(0.005)	(0.004)
Fall 2019 Quintile Rank: 2	-0.079***	-0.056***	-0.047***	-0.030***
	(0.007)	(0.006)	(0.004)	(0.003)
Fall 2019 Quintile Rank: 3	-0.058***	-0.043***	-0.038***	-0.024***
	(0.005)	(0.004)	(0.003)	(0.003)
Fall 2019 Quintile Rank: 4	-0.027***	-0.020***	-0.018***	-0.009***
	(0.003)	(0.003)	(0.002)	(0.002)
F-Statistic	47.21***	35.00***	40.59***	26.17***
Student Characteristics	No	Yes	Yes	Yes
Fixed Effects	No	No	District	School
R-squared	0.004	0.013	0.068	0.103
Observations	2,178,893	2,178,893	2,178,893	2,178,893
Panel B: Reading				
	(1)	(2)	(3)	(4)
Constant	-0.062***	-0.043***	-0.057***	-0.070***
	(0.005)	(0.007)	(0.007)	(0.007)
Fall 2019 Quintile Rank: 1 (Lowest)	-0.117***	-0.096***	-0.086***	-0.063***
	(0.008)	(0.006)	(0.006)	(0.005)
Fall 2019 Quintile Rank: 2	-0.073***	-0.056***	-0.050***	-0.035***
	(0.006)	(0.005)	(0.005)	(0.004)
Fall 2019 Quintile Rank: 3	-0.065***	-0.054***	-0.050***	-0.039***
	(0.004)	(0.004)	(0.004)	(0.003)
Fall 2019 Quintile Rank: 4	-0.044***	-0.039***	-0.037***	-0.030***
	(0.003)	(0.003)	(0.003)	(0.003)
F-Statistic	83.47***	79.07***	62.89***	50.18***
Student Characteristics	No	Yes	Yes	Yes
Fixed Effects	No	No	District	School
R-squared	0.004	0.006	0.034	0.061
Observations	1,532,676	1,532,676	1,532,676	1,532,676

Note: Learning loss is defined as the difference between a student's norms-standardized 2021 fall NWEA MAP score and their expected score. The parameters for predicting expected scores were drawn from a pre-pandemic regression of fall 2019 norms-standardized scores on students' baseline characteristics from 2017. Point estimates can be interpreted as the magnitude of learning loss relative to that of students in the highest quintile in 2019. Robust standard errors (clustered at the district level) are in parentheses.

Table 4. Gelbach Decomposition of Differential Learning Loss Incidence for Various Prior Achievement Quintiles

Panel A: Differences in Learning Loss (Math)				
	(1) Q1 - Q5	(2) Q2 - Q5	(3) Q3 - Q5	(4) Q4 - Q5
Student Characteristics				
Grade Indicators	-0.0002* (-2.09)	0.0020*** (19.75)	0.0023*** (22.11)	0.0025*** (25.02)
Race Indicators	-0.0012** (-2.83)	-0.0008** (-2.59)	-0.0005* (-2.13)	-0.0002 (-1.20)
Testing Date	0.0039*** (19.68)	0.0051*** (25.74)	0.0047*** (25.18)	0.0028*** (16.06)
School Indicators	-0.0770*** (-117.19)	-0.0558*** (-97.90)	-0.0404*** (-81.61)	-0.0234*** (-54.88)
Total Explained	-0.0744*** (-114.83)	-0.0494*** (-87.39)	-0.0339*** (-69.13)	-0.0183*** (-43.67)
Unconditional Average Difference	-0.1004*** (-10.98)	-0.0812*** (-9.68)	-0.0607*** (-9.05)	-0.0288*** (-6.81)

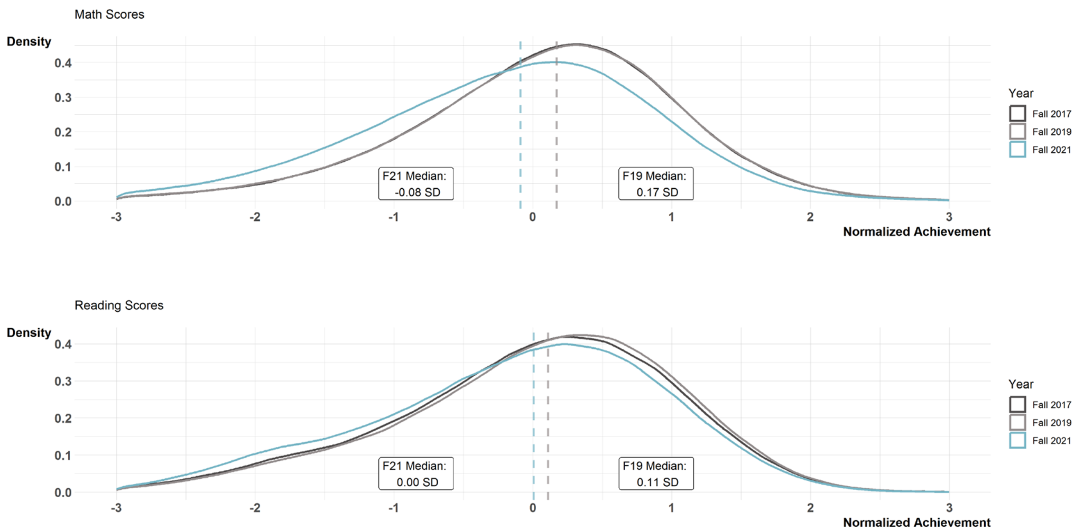
Panel B: Differences in Learning Loss (Reading)				
	(1) Q1 - Q5	(2) Q2 - Q5	(3) Q3 - Q5	(4) Q4 - Q5
Student Characteristics				
Grade Indicators	-0.0010*** (-8.35)	-0.0009*** (-9.66)	0.0003** (3.18)	0.0009*** (6.96)
Race Indicators	0.0014** (3.07)	0.0011** (3.11)	0.0008** (3.01)	0.0004** (2.65)
Testing Date	-0.0002 (-0.71)	0.0012*** (5.03)	0.0018*** (7.91)	0.0012*** (5.77)
School Indicators	-0.0513*** (-66.07)	-0.0374*** (-56.10)	-0.279*** (-48.16)	-0.0162*** (-32.01)
Total Explained	-0.0510*** (-69.25)	-0.0361*** (-56.78)	-0.0251*** (-45.61)	-0.0136*** (-29.30)
Unconditional Average Difference	-0.1128*** (-16.52)	-0.0734*** (-13.49)	-0.0680*** (-16.20)	-0.0478*** (-16.90)

Note: Table reports the estimated differences in learning between various prior achievement quintiles using the methodology proposed by Gelbach (2016). T-statistics are reported in parentheses.

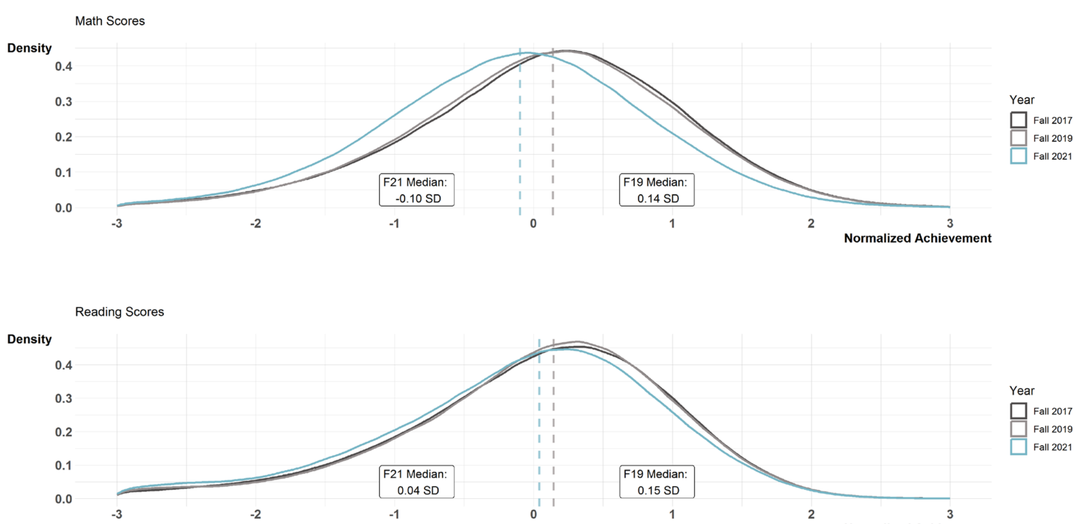
Appendix A: Figures and Tables

Appendix Figure A1. Distribution of Math and Reading Test Scores for Elementary Grades (3-5) and Middle Grades (6-8)

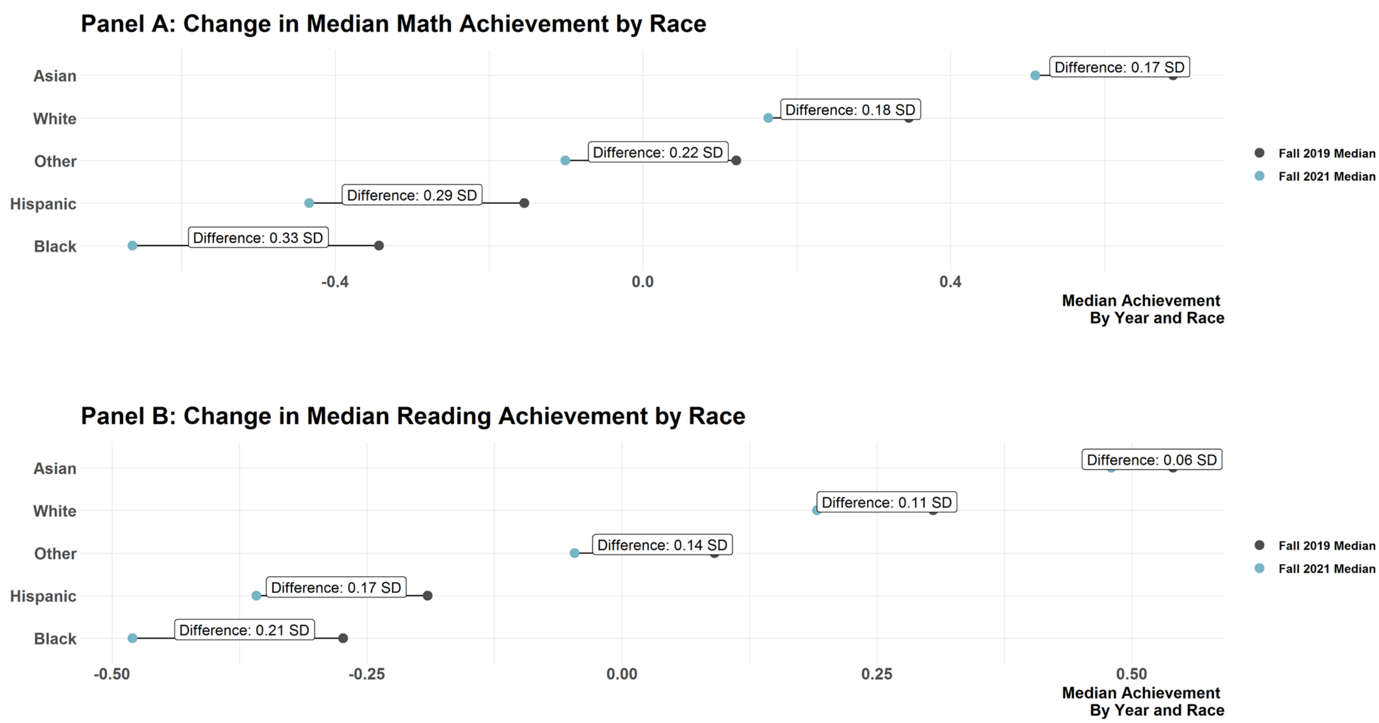
Panel A: Elementary Grades (3-5)



Panel B: Middle Grades (6-8)



Appendix Figure A2. Change in Median Achievement by Subject and Race



Appendix Table A1. Comparison of Means by Grade and Year (Math)

	Fall 2017	Fall 2019	Fall 2021			Fall 17 vs. 19	Fall 17 vs. 21	Fall 19 vs. 21
Grade 3								
Mean	0.046	0.043	-0.223		Difference	0.004	0.269	0.265
SD	1.000	1.018	1.110		T-Statistic	1.811	129.937	126.212
N	551,148	529,941	500,147		P-Value	0.070	0.000	0.000
Grade 4								
Mean	0.106	0.095	-0.174		Difference	0.012	0.281	0.269
SD	0.975	0.999	1.076		T-Statistic	6.061	139.029	132.046
N	536,258	538,941	502,755		P-Value	0.000	0.000	0.000
Grade 5								
Mean	0.103	0.077	-0.193		Difference	0.027	0.296	0.270
SD	1.018	1.028	1.088		T-Statistic	13.580	144.252	131.267
N	542,063	551,853	512,404		P-Value	0.000	0.000	0.000
Grade 6								
Mean	0.050	0.035	-0.182		Difference	0.015	0.232	0.217
SD	0.980	0.956	0.984		T-Statistic	7.971	117.260	111.811
N	519,936	486,824	465,522		P-Value	0.000	0.000	0.000
Grade 7								
Mean	0.124	0.106	-0.118		Difference	0.018	0.243	0.225
SD	1.005	0.992	0.983		T-Statistic	9.191	119.548	113.453
N	487,884	520,601	469,890		P-Value	0.000	0.000	0.000
Grade 8								
Mean	0.152	0.130	-0.099		Difference	0.022	0.251	0.230
SD	0.982	0.971	0.957		T-Statistic	10.939	123.285	114.208
N	464,391	481,613	439,499		P-Value	0.000	0.000	0.000

Appendix Table A2. Comparison of Means by Grade and Year (Reading)

		Fall 2017	Fall 2019	Fall 2021		Fall 17 vs. 19	Fall 17 vs. 21	Fall 19 vs. 21
Grade 3								
	Mean	-0.003	0.027	-0.152	Difference	-0.030	0.149	0.179
	SD	1.004	1.013	1.075	T-Statistic	-14.541	69.106	82.627
	N	485,238	485,789	451,153	P-Value	0.000	0.000	0.000
Grade 4								
	Mean	-0.003	0.054	-0.087	Difference	-0.057	0.084	0.141
	SD	1.003	0.997	1.034	T-Statistic	-26.843	37.640	65.084
	N	418,978	477,134	406,840	P-Value	0.000	0.000	0.000
Grade 5								
	Mean	0.009	0.039	-0.099	Difference	-0.030	0.108	0.138
	SD	1.033	1.011	1.040	T-Statistic	-13.808	46.793	62.698
	N	411,052	486,824	393,616	P-Value	0.000	0.000	0.000
Grade 6								
	Mean	-0.036	0.015	-0.089	Difference	-0.051	0.054	0.105
	SD	1.011	0.984	1.012	T-Statistic	-23.429	23.053	47.904
	N	378,678	468,619	376,128	P-Value	0.000	0.000	0.000
Grade 7								
	Mean	0.007	0.025	-0.078	Difference	-0.018	0.085	0.103
	SD	1.026	0.992	1.016	T-Statistic	-7.682	35.158	46.273
	N	337,542	451,469	370,130	P-Value	0.000	0.000	0.000
Grade 8								
	Mean	0.030	0.027	-0.072	Difference	0.003	0.102	0.100
	SD	1.000	0.973	0.998	T-Statistic	1.232	42.459	44.819
	N	319500	421,208	369,904	P-Value	0.218	0.000	0.000

Appendix Table A3. Norm-Standardized MAP Test Scores by Year and Percentile Rank

Panel A: Math						
	10th	25th	50th	75th	90th	90th-10th Difference
Fall 2017	-1.219	-0.515	0.154	0.751	1.274	2.493
Fall 2019	-1.227	-0.528	0.141	0.741	1.271	2.498
Fall 2021	-1.529	-0.811	-0.092	0.555	1.116	2.645
Panel B: Reading						
	10th	25th	50th	75th	90th	90th-10th Difference
Fall 2017	-1.334	-0.619	0.093	0.716	1.234	2.567
Fall 2019	-1.307	-0.592	0.114	0.727	1.240	2.546
Fall 2021	-1.526	-0.788	-0.019	0.633	1.175	2.701

Appendix Table A4. Distribution of Norm-Standardized MAP Math Test Scores by Year and Grade

		10th	25th	50th	75th	90th
Grade 3						
	Fall 2017	-1.26	-0.54	0.12	0.71	1.23
	Fall 2019	-1.28	-0.55	0.11	0.71	1.25
	Fall 2021	-1.70	-0.90	-0.13	0.53	1.09
Grade 4						
	Fall 2017	-1.15	-0.47	0.19	0.75	1.20
	Fall 2019	-1.17	-0.47	0.20	0.77	1.21
	Fall 2021	-1.56	-0.82	-0.06	0.57	1.06
Grade 5						
	Fall 2017	-1.22	-0.50	0.17	0.75	1.30
	Fall 2019	-1.24	-0.51	0.17	0.74	1.29
	Fall 2021	-1.60	-0.86	-0.10	0.55	1.10
Grade 6						
	Fall 2017	-1.32	-1.23	-0.54	0.11	0.68
	Fall 2019	-1.28	-1.20	-0.55	0.08	0.65
	Fall 2021	-1.54	-1.44	-0.79	-0.14	0.46
Grade 7						
	Fall 2017	-1.19	-0.50	0.18	0.80	1.33
	Fall 2019	-1.18	-0.52	0.15	0.77	1.32
	Fall 2021	-1.37	-0.75	-0.10	0.53	1.10
Grade 8						
	Fall 2017	-1.15	-0.47	0.18	0.81	1.33
	Fall 2019	-1.15	-0.49	0.15	0.79	1.32
	Fall 2021	-1.33	-0.73	-0.10	0.52	1.09

Appendix Table A5. Distribution of Norm-Standardized MAP Reading Scores by Year and Grade

		10th	25th	50th	75th	90th
Grade 3						
	Fall 2017	-1.45	-0.67	0.08	0.71	1.23
	Fall 2019	-1.42	-0.67	0.11	0.76	1.27
	Fall 2021	-1.70	-0.93	-0.06	0.65	1.19
Grade 4						
	Fall 2017	-1.42	-0.61	0.09	0.69	1.18
	Fall 2019	-1.35	-0.52	0.17	0.75	1.23
	Fall 2021	-1.57	-0.73	0.02	0.64	1.13
Grade 5						
	Fall 2017	-1.39	-0.60	0.11	0.73	1.21
	Fall 2019	-1.29	-0.52	0.16	0.74	1.20
	Fall 2021	-1.54	-0.71	0.01	0.62	1.10
Grade 6						
	Fall 2017	-1.38	-0.61	0.07	0.65	1.12
	Fall 2019	-1.29	-0.55	0.11	0.67	1.13
	Fall 2021	-1.47	-0.70	0.01	0.59	1.07
Grade 7						
	Fall 2017	-1.35	-0.57	0.13	0.71	1.18
	Fall 2019	-1.30	-0.54	0.13	0.70	1.17
	Fall 2021	-1.47	-0.68	0.03	0.61	1.10
Grade 8						
	Fall 2017	-1.28	-0.52	0.15	0.71	1.17
	Fall 2019	-1.24	-0.50	0.14	0.69	1.15
	Fall 2021	-1.41	-0.64	0.04	0.61	1.10

Appendix B: Alternative Measures of Differential Learning Loss by Prior Achievement

Robustness of Estimates

There are potential limitations to the empirical approach, namely in the representativeness of the counterfactual (i.e., pre-pandemic growth estimates). By using only three years of data, our approach is limited to two periods of growth. As a result, we do not know the degree to which changes in the norms-standardized achievement distribution might simply reflect changes in the distribution that happen over the course of any two years (in the absence of the pandemic). We examine the extent to which there is evidence that our results could represent statistical noise or changes in the underlying test score population. As previously mentioned, we compare fall test scores across fall 2017, fall 2019, and fall 2021, as well as growth between each temporal pair, in Appendix Table A2. These results show that changes in norms-standardized test score levels across pre-pandemic years are within the What Works Clearinghouse's threshold for baseline equivalence (i.e., <=.05SD; What Works Cleaninghouse, 2022). We show that there is little evidence of significant changes in test scores along the prior test distribution outside of the large changes that occur across the pandemic.

Alternative Specifications

The results provided in this study show how estimates of learning loss between norm-standardized quintiles differ on average. However, there may be significant different in learning losses within quintiles. To account for this possibility, we instead a model using norm-standardized prior achievement scores:

$$G_{i2021} = \alpha_s + \sum_{\tau=1}^{3} \beta_\tau Y_{i2019}^{\tau} + \theta \boldsymbol{X}_{i2021} + \varepsilon_{ij}$$

where $\sum_{\tau=1}^{3} Y_{(is2019)}^{\tau}$ is the set of lagged student normalized test score (and higher-order polynomials) from the prior testing period (fall 2019). Higher-order polynomials from the prior testing period account for nonlinear patterns that may have occurred in learning loss. Taken together, the estimates for β_{τ} describe how learning loss varied by prior achievement level.