AER: Insights 2023, 5(3): 377–392
https://doi.org/10.1257/aeri.20220180

The Educational Consequences of Remote and Hybrid Instruction during the Pandemic†

By Dan Goldhaber, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger*

Using testing data from over two million students in nearly 10,000 schools in 49 states (plus the District of Columbia), we investigate the role of remote and hybrid instruction in widening gaps in achievement by race and school poverty. We find that remote instruction was a primary driver of the widening gaps. Math gaps did not widen in areas that remained in person (although reading gaps did). We estimate that high-poverty districts that went remote in 2020–2021 will need to spend nearly all of their federal aid on helping students recover from pandemic-related academic achievement losses. (*JEL* H75, I12, I21, I24, I32, J15)

Since the pandemic started in March 2020, there have been multiple reports of large declines in students' math and reading achievement as well as widening gaps by race and school poverty.[1] If allowed to become permanent, such losses will have major impacts on future earnings and intergenerational mobility.[2]

We investigate the role of remote and hybrid instruction in contributing to the losses, using student-level data from over two million students in nearly 10,000 schools from 49 states (plus the District of Columbia). To do so, we compare students' achievement growth during the pandemic (fall 2019 to fall 2021) to a prepandemic period (fall 2017 to fall 2019). A prior study using district-level data in 11 states by Jack et al. (2023) documented declines in proficiency rates in districts that shifted to remote instruction, especially in districts serving larger shares

* Goldhaber: CALDER at the American Institutes for Research (AIR) and CEDR at the University of Washington (email: dgoldhaber@air.org); Kane: Center for Education Policy Research at Harvard University (email: tom_kane@gse.harvard.edu); McEachin: NWEA (email: andrew.mceachin@nwea.org); Morton: NWEA (email: emorton@air.org); Patterson: Center for Education Policy Research at Harvard University (email: tpatterson@uchicago.edu); Staiger: Department of Economics at Dartmouth College (email: douglas.o.staiger@dartmouth.edu). Peter Klenow was coeditor for this article. We thank anonymous referees for helpful suggestions. The research was supported by grants from the Carnegie Corporation of New York, the Walton Family Foundation, Kenneth C. Griffin, and an anonymous foundation.

† Go to https://doi.org/10.1257/aeri.20220180 to visit the article page for additional materials and author disclosure statement(s).

[1] For instance, see Curriculum Associates (2020, 2021a, b); Darling-Aduana et al. (2022); Dorn et al. (2020); Kogan and Lavertu (2021); Kuhfeld et al. (2021); Lewis and Kuhfeld (2021); Lewis et al. (2021).

[2] Using evidence on test scores and achievement from Neal and Johnson (1996) and Murnane, Willett, and Levy (1995), Goldhaber, Kane, and McEachin (2021) estimated that the losses would cost the United States $2 trillion in lifetime earnings. The World Bank estimated that the worldwide losses in lifetime earnings would be $17 trillion (World Bank, UNESCO, and UNICEF 2021).

of Black and lower-income students. Due to data limitations, however, the authors were unable to investigate whether gaps by race or school poverty changed within districts or schools when schools went remote.[3]

We make four primary contributions. First, we estimate a model of achievement growth in the prepandemic period (conditioning on student and school characteristics as well as prior achievement) and then compare students' actual and expected achievement growth during the pandemic. By doing so, we distinguish pandemic-related achievement losses from preexisting differences in achievement growth by student and school characteristics.

Second, we investigate differential impacts of remote instruction in high- and low-income schools. We find that the shift in instructional mode was a primary driver of widening achievement gaps by race/ethnicity and by school poverty status. High-poverty schools were more likely to remain in remote instruction, and their students were more negatively impacted when they did so. Within school districts that were remote for most of 2020–2021, high-poverty schools experienced 50 percent more loss in achievement growth than low-poverty schools (0.46 versus 0.30 standard deviations in math). In contrast, math achievement gaps by school poverty status and by race did not widen in areas where instruction remained in person (although there was some widening in reading gaps in those areas).

Third, we investigate within-school differences in the impacts of the pandemic on student subgroups. We find that most of the widening by race/ethnicity occurred because the students in schools attended by Black and Hispanic students were more negatively impacted rather than because they fell behind classmates attending the same school. Put another way: the widening racial gap happened because of negative shocks to schools attended by disadvantaged students, not because of differential impacts within schools.

Fourth, we provide a lower-bound estimate of the cost of academic recovery by district. To do so, we compare the share of a typical school year that students have lost to the share of their annual budget they have received in federal aid. The estimate is likely to be a lower bound since the marginal cost for supplemental catch-up efforts is likely to be higher than during the typical school year. The federal government, through the American Rescue Plan (ARP), has provided \$190 billion in aid to education agencies to help with academic recovery. We estimate that high-poverty districts that were remote for most of 2020–2021 will need to spend nearly all of their federal aid on academic recovery in order to eliminate achievement losses, far more than the 20 percent required under federal law.

I. Data Sources

For a national sample of student achievement, we rely on data from the Growth Research Database of NWEA (NWEA 2021), a nonprofit assessment provider. Roughly 3,000 school districts administer NWEA's Measures of Academic Progress

[3] Kilbride et al. (2021) also find larger declines in achievement in schools that went remote in the state of Michigan.

(MAP) Growth assessments. Unlike state-mandated tests, districts typically administer the MAP assessment three times per year, in the fall, winter, and spring. Though some remote testing occurred during the pandemic, nearly all MAP Growth tests were administered in person at the students' schools during the three fall terms included in the present study.

The MAP Growth assessment is a computer adaptive test, meaning that the difficulty of test questions increases or decreases in response to a student's prior responses. In contrast to tests with a standard test form for all students, the adaptive tests are designed to improve reliability at both the high end and low end of achievement (which is particularly useful when so many students are performing below grade level). We have standardized scores by using the means and standard deviations by grade and subject and control for testing date using NWEA's prepandemic norms[4] (Thum and Kuhfeld 2020). The NWEA data also include student-level demographic data on race/ethnicity and gender, as well as district and school identifiers.

We supplement the NWEA data with administrative data from the Common Core of Data (CCD): enrollment by school and grade in 2019–2020, the urbanicity of the school, expenditures on elementary and secondary education, and the percent of students in each school qualifying for free and reduced-price lunch (FRPL) through the National School Lunch Program (National Center for Education Statistics 2020). We categorized a school as "low poverty" if fewer than 25 percent of students were receiving FRPL and "high poverty" if more than 75 percent of students were receiving FRPL.[5]

In addition, we added information on the population density (population per square mile) within each school district and population by county using data from the US Census Bureau, the National Historical Geographic Information System, and the National Center for Education Statistics (US Census Bureau 2020; Manson et al. 2021; National Center for Education Statistics 2019), COVID infection rates by county from Johns Hopkins University (Center for Systems Science and Engineering 2020), and federal Elementary and Secondary School Emergency Relief (ESSER) funds by district (US Department of Education 2019, 2020; Office of Elementary and Secondary Education 2020, 2021).[6]

To measure schools' instructional mode during 2020–2021, we rely on the Return to Learn Tracker (American Enterprise Institute 2021) assembled by the American Enterprise Institute (AEI).[7] AEI collected data on districts with three or more

[4]The NWEA national norms have been weighted to reflect the national population of K–8 public schools in 2015–2016.

[5]To be eligible for federally subsidized lunch programs, a student must have income below 185 percent of the federal poverty threshold or attend a school where greater than 40 percent of students are categorically eligible by living in households receiving Temporary Assistance for Needy Families (TANF), Supplemental Nutrition Assistance Program (SNAP) benefits, Medicaid, or being homeless, a migrant student, or in foster care. Where FRPL values were unavailable, we used the percent of students eligible through direct certification. This included the entirety of two states (Delaware and Massachusetts) and the District of Columbia, as well as 2.6 percent of schools outside these areas.

[6]We estimated ESSER allocations by district using state ESSER totals and prior Title 1 allocations for each district.

[7]Given missing data in the early weeks, we start from September 7, 2020, the date for which over 95 percent of available districts have data.

schools and captured weekly data on mode of instruction from August 2020 through June 7, 2021, for 91 percent of K–12 enrollment in the United States.[8,9]

II. Representativeness of the Analysis Sample

Our analytic sample for math consists of 2.1 million students at 9,690 schools from 49 states (plus the District of Columbia).[10] The sample includes students who were in grades 3–8 in the follow-up year. We included schools that were covered in the AEI data and had valid test scores for at least ten students on the English language versions of the mathematics or reading assessments in fall 2017, fall 2019, and fall 2021 (all three years). In addition, individual students were required to have scores for both a baseline year (i.e., fall 2017 or fall 2019) and a follow-up test two years later (i.e., fall 2019 and fall 2021). Finally, students were excluded if their school tested less than 60 percent of their grade's enrollment based on data from the CCD.

In comparison to the universe of public schools with students in grades 3–8, our analytic sample for studying math achievement contains a smaller percentage of Hispanic students (20 percent versus 28 percent nationally), slightly less representation of high-poverty schools (22 percent versus 27 percent), and greater representation among suburban schools (44 percent versus 39 percent).[11] The analytic sample also had similar percentages of the year spent in remote and hybrid instruction (21 and 47 percent, respectively) as for all schools with both CCD and AEI data (24 and 46 percent). The mean total enrollment in grades 3–8 in our analysis sample was larger than in the CCD overall (3,300 versus 1,400).[12]

The requirement that students have a follow-up score led us to exclude roughly a quarter of students with valid baseline tests (25 percent in fall 2017 and 29 percent in fall 2019).[13,14] Given the change in attrition rates, we test the robustness of our findings by including the share of students tested in the school as a covariate.

[8] The AEI data cover 98 percent of students in districts with three or more schools, and such districts account for 92 percent of K–12 enrollment. We inferred the school attended during the academic year preceding the fall follow-up assessment. For the majority of students, we used the school code where the student was tested during 2018–2019 or 2020–2021. If students attended the same school in the baseline and follow-up year, we assume that they attended that school during the intervening year. For students who changed schools (and advanced two grade levels), we use grade-span data for their former and current school to infer which they attended. If both schools served the student's grade level in the intervening year, we treat the school as missing.

[9] The AEI data capture instructional mode at the district, not school, level. Although they cover a smaller share of schools, the COVID-19 School Data Hub (COVID-19 School Data Hub 2022) captures school-level data in 34 states. Even in those states, 89 percent of the variance in percent of year remote occurs at the district (rather than school) level.

[10] The NWEA analysis file only included students taking the English language version of the test.

[11] In online Appendix Table 3, we report descriptive statistics for our analysis samples as well as for the full CCD universe of public schools with students in grades 3–8.

[12] In unpublished results, we reweighted by $\frac{TotalEnrollmentinCCD_d}{TotalEnrollmentinAnalysisSample_d}$, where d corresponds to the decile of a student's district size in the CCD. The results are essentially unchanged. We also reestimated the specifications in Table 1 for the smallest quintile of districts in the analysis sample. The results were similar.

[13] In online Appendix Table 5, we report the degree to which each of the covariates is related to attrition in both the prepandemic and pandemic periods.

[14] We also excluded students whose schools tested less than 60 percent of their grade's enrollment, dropping an additional 0.3 and 2.2 percent of students in follow-up years fall 2019 and fall 2021, respectively.

III. Differing Incidence of Remote Instruction by School Poverty Level

As others have found (Parolin and Lee 2021; Camp and Zamarro 2021; Grossmann et al. 2021; Oster et al. 2021), we observe a higher incidence of remote schooling for Black and Hispanic students. We also find that high-poverty schools spent about 5.5 more weeks in remote instruction during 2020–2021 than low- and mid-poverty schools.[15] We observed large differences in remote instruction by state.[16] High-poverty schools were more likely to be remote in high- and low-closure states, but the gaps were largest in those states with higher rates of remote instruction overall. For example, in high-remote-instruction states (including populous states such as California, Illinois, New Jersey, Virginia, Washington, and the District of Columbia), high-poverty schools spent an additional nine weeks in remote instruction (more than two months) than low-poverty schools. In states with the lowest rates of remote instruction (including Florida and Texas), high-poverty schools were also more likely to be remote, but the differences were small: three weeks remote in high-poverty schools versus one week in low-poverty schools.[17]

IV. Inferring the Impacts of Remote and Hybrid Instruction

Figure 1 describes the mean change in unadjusted scores between fall 2017 and fall 2019 (prepandemic) and between fall 2019 and fall 2021, in districts that were fully in person and in districts that were remote for a majority of the 2020–2021 school year. In low-poverty schools that remained in person during 2020–2021, the average student improved their scores between 2019 and 2021, but by 0.08 standard deviations less than in the same districts before the pandemic. The average two-year test score gain also fell in mid- and high-poverty schools by 0.15 and 0.18 standard deviations, respectively. However, in districts that were remote for more than half a year, the decline in growth was larger and more unequal by school poverty—0.27, 0.42, and 0.48 standard deviations in low-, mid-, and high-poverty schools, respectively.

To establish prepandemic growth expectations and to adjust the estimates in Figure 1 for differences in student baseline scores, demographics, and school characteristics, we first estimate the following model of achievement growth (Todd and Wolpin 2003) during the prepandemic period:

$$S_{i0} = \beta_0 + \mathbf{Race}_i \beta_{Race} + \mathbf{Pov}_{j0} \beta_1 + \mathbf{Mode}_{j,2021} \beta_2 + \mathbf{Pov}_{j0} \mathbf{Mode}_{j,2021} \beta_3 + \mathbf{X}_{ij0} \beta_4 + \varepsilon_{i0}, \tag{1}$$

where i subscripts the student, j subscripts the school attended in 2018–2019 (the school year between the baseline year and follow-up), and the zero subscript refers

[15] After adjusting for population density, the urbanicity of the school, and county-level COVID infection rates, the gap in weeks of remote instruction between high- and low-poverty schools is only slightly smaller (roughly 4.6 weeks).

[16] See online Appendix Figure 1.

[17] States with low closure rates included Arkansas, Florida, Idaho, Louisiana, Maine, Montana, North Dakota, Nebraska, South Dakota, Texas, Utah, Vermont, and Wyoming.

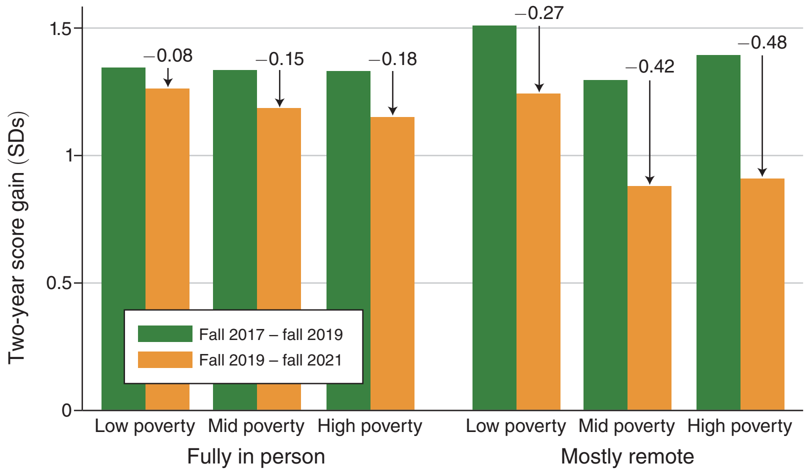

Figure 1. Changes in Mean Scores before and during Pandemic by 2020–2021 Closure Status and School Poverty

Notes: Test scores have not been adjusted for student demographics, grade level, school characteristics, and so forth. We report the mean gain in math scores in grades 3–8 in standard deviation units, using the standard deviation in scores in grade attended during 2020–2021.

to the prepandemic period. $\mathbf{Race}_i$ is a vector of dummies for students' race/ethnicity (Black, Hispanic, Asian, and Other with White as the reference group), $\mathbf{Pov}_{j0}$ is a vector of dummies for the poverty status of the school attended (mid and high poverty with low poverty as the reference group), $\mathbf{Mode}_{j,2021}$ is a vector with the percentage of the year that a school was hybrid and remote during the 2020–2021 school year, and $\mathbf{X}_{ij0}$ is a vector of student and school characteristics (including a cubic in baseline achievement fully interacted with grade level, gender, and the date of testing in the baseline and in the follow-up year included as linear terms).

Even before the pandemic, there were significant differences in achievement growth by race/ethnicity and school poverty status after controlling for baseline achievement.[18] For example, relative to White students with similar baseline scores and school poverty levels, Black students' math test scores were 0.12 standard deviations lower two years later and Hispanic students' scores were 0.02 standard deviations lower. The magnitude of widening for Black and Hispanic students was similar in reading. Conditioning on student race/ethnicity and baseline scores, students in high-poverty schools also fell behind by approximately 0.18 standard deviations in math and 0.14 standard deviations in reading during 2017–2019.

In the growth model above, we also included controls for the instructional mode used by their school during the 2020–2021 school year. Although there should be no causal relationship between remote/hybrid schooling in 2020–2021 and student growth between 2017 and 2019, we estimate such differences to identify any preexisting relationships between a school's subsequent use of remote/hybrid

[18] See online Appendix Table 4.

schooling and growth. The differences were small but, in some cases, statistically significant. As we will describe, we difference those out from 2019–2021 growth.

Equation (1) establishes a benchmark for how achievement, conditional on prior scores, varied by race, school poverty, and pandemic instructional mode before the pandemic.[19] We use those estimates to construct our primary outcome, which is the degree to which each student in 2019–2021 underperformed (or overperformed) growth expectations from the 2017–2019 period.[20] Specifically, we apply the 2017–2019 parameters to the 2019–2021 sample to estimate the difference between a student's actual and expected growth during the pandemic as follows:

$$(2) \qquad R_{i1} = S_{i1} - \Big(\hat{\beta}_0 + \mathbf{Race}_{i1}\hat{\beta}_{Race} + \mathbf{Pov}_{j1}\hat{\beta}_1 + \mathbf{Mode}_{j,2021}\hat{\beta}_2 + \mathbf{Pov}_{j1}\mathbf{Mode}_{j,2021}\hat{\beta}_3 + \mathbf{X}_{ij1}\hat{\beta}_4\Big).$$

Thus, when we refer to a "loss" or "decline" in achievement growth, we mean that actual achievement growth was less than expected given prepandemic relationships ($R_{i1} < 0$). Because the dependent variable, R_{i1}, includes estimated parameters, we calculated bootstrapped standard errors, sampling districts and reestimating both equations (1) and (2).

In Tables 1 and 2, we focus on math achievement impacts.[21] With one important exception (which we will highlight), the pattern of results in reading are similar (though the magnitudes are smaller). For brevity, we also pool results across grades 3–8. The magnitudes of differences are larger in grades 3–5 than in 6–8, but the patterns are similar.[22]

In Table 1, we describe how 2020–2021 growth diverged from expectations for different subgroups of students by regressing R_{i1} on different combinations of covariates. In column 1, we report that Black and Hispanic students lost more ground relative to White students with similar baseline achievement during the pandemic period than in the prepandemic period: Black students lost an additional 0.119 standard deviations, and Hispanic students lost an additional 0.092 standard deviations. (As reflected in the constant term, White students, the excluded subgroup, also lost 0.208 standard deviations relative to prepandemic growth.)

In column 2, we report differences in R_{i1} by students' baseline achievement. As reflected in the constant term, actual growth for students in the highest quartile on the baseline assessment (the excluded category) during the pandemic period was 0.194 standard deviations lower than expected growth. Students who were in the middle two quartiles of achievement in fall 2019 lost an additional 0.053 standard

[19] Given that attrition rates were 4 percent higher between 2019 and 2021, it is possible that the 2017–2019 growth expectations differed for those with nonmissing 2021 scores. In online Appendix Table 7, we reestimated equations (1) and (2) for those with nonmissing scores in all three years (limited to those who were in grades 3–6 in fall 2019 given that we lack scores for those above grade 8). The correlation in the 2019–2021 residuals, when limiting the sample to those with nonmissing 2017–2019 scores versus those with nonmissing scores all three years, was 0.9998.

[20] Our approach to measuring growth differs from that used by NWEA in its national reports. NWEA typically conditions on baseline scores, testing date, and grade—but not race/ethnicity or school poverty level.

[21] Online Appendix Tables 1 and 2 contain reading analogues to Tables 1 and 2, respectively. Online Appendix Figures 2 and 3 contain reading analogues to Figures 2 and 3, respectively.

[22] In math, the pattern of results by race, school poverty, and instructional mode are similar in elementary and middle school grades. In reading, the direct effect of school poverty status is larger in middle school grades.

Table 1—Pandemic Achievement Gains by Student and School Characteristics, Math

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Race (reference: White)							
Black	−0.119		−0.101	−0.036	−0.040	−0.057	−0.040
	(0.012)		(0.012)	(0.006)	(0.006)	(0.007)	(0.006)
Hispanic	−0.092		−0.077	−0.032	−0.014	−0.043	−0.014
	(0.015)		(0.014)	(0.005)	(0.007)	(0.006)	(0.007)
Asian	−0.013		−0.020	−0.029	0.005	−0.026	0.005
	(0.012)		(0.012)	(0.007)	(0.008)	(0.007)	(0.008)
Other	−0.041		−0.035	−0.019	−0.017	−0.025	−0.017
	(0.007)		(0.007)	(0.005)	(0.007)	(0.006)	(0.007)
Panel B. Baseline score (reference: top quartile)							
Middle quartiles		−0.053	−0.040	−0.012	−0.030	−0.016	−0.030
		(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Bottom quartile		−0.107	−0.078	−0.022	−0.053	−0.030	−0.053
		(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
School poverty (reference: low <25%)							
Middle (25%–75%)					−0.018	0.020	−0.017
					(0.011)	(0.015)	(0.011)
High (>75%)					−0.002	0.024	−0.001
					(0.021)	(0.022)	(0.021)
Panel C. Remote schooling							
Percent remote in 2020–2021					−0.201	N/A	−0.199
					(0.025)		(0.025)
Interactions:							
Middle poverty					−0.086	−0.103	−0.086
					(0.027)	(0.030)	(0.027)
High poverty					−0.158	−0.183	−0.159
					(0.039)	(0.040)	(0.039)
Panel D. Hybrid schooling							
Percent hybrid in 2020–2021					−0.033	N/A	−0.033
					(0.017)		(0.017)
Interactions:							
Middle poverty					−0.051	−0.023	−0.051
					(0.017)	(0.023)	(0.017)
High poverty					−0.118	−0.084	−0.118
					(0.031)	(0.035)	(0.031)
Percent tested in school							0.027
							(0.030)
Constant	−0.208	−0.194	−0.175	N/A	−0.098	N/A	−0.122
	(0.006)	(0.007)	(0.006)		(0.011)		(0.030)
Fixed effects?	No	No	No	School	No	District	No

Notes: Sample includes 2,102,010 students in grades 3–8 at the time of their follow-up test. The dependent variable is defined in equation (1) as the difference between a student's standardized 2021 fall NWEA MAP score and the expected score of students with similar baseline achievement and characteristics during the prepandemic period. Bootstrapped standard errors were estimated by resampling at the district level and reestimating both equations (1) and (2) 1,000 times.

deviations, while students in the bottom quartile in the baseline lost an additional 0.107 standard deviations.

In column 3, we report the gaps by race and by baseline score while conditioning on student characteristics. Because student race/ethnicity and baseline score are correlated, the magnitude of the loss for each is somewhat smaller when conditioning on both.

In column 4, we include school fixed effects. In this specification, the Black-White and Hispanic-White achievement gaps in math achievement are greatly diminished though both are still positive. Specifically, with the inclusion of school fixed effects, the race/ethnicity test achievement gaps are only about a third of the magnitude of the specification without school fixed effects (reported in column 1), falling to 0.036 for the Black-White differential and 0.032 for the Hispanic-White differential. The smaller magnitudes suggest that much of the gap in test scores reported in column 3 is a result of school-level shocks rather than differential effects of the pandemic on racial/ethnic subgroups within schools. Likewise, the gap in math achievement between students in the highest and lowest quartile of baseline achievement shrinks by 72 percent with the inclusion of school fixed effects $(0.022/0.078 = 0.28)$.

The results in column 4 have implications for academic recovery efforts: to reverse pandemic-related losses (as opposed to addressing inequities that existed before the pandemic) districts might focus on the hardest hit schools, rather than target subgroups within schools.

In column 5, we parameterize school effects with three factors: school poverty status, the percentage of the 2020–2021 school year that the school was in remote or hybrid instruction, and the interaction between the two.[23] The conditional difference by race/ethnicity remains small (0.04 standard deviations for White-Black and 0.01 standard deviations for White-Hispanic), implying that the simple parameterization captures much of the information in the school effects specification in column 4.[24,25]

Several other findings from Table 1 are noteworthy. In column 5, the main effects of school poverty status—which apply to those schools that were in person for all of 2020–2021—are small and no longer statistically significant. In other words, as long as schools were in person throughout 2020–2021, there was no significant widening of math achievement gaps between high-, middle-, and low-poverty schools.

The main effects of hybrid and remote instruction are negative, implying that even at low-poverty (high-income) schools, students progressed more slowly when their schools went remote or hybrid. Specifically, if their schools were remote throughout 2020–2021, students in low-poverty schools lost 0.201 standard deviations relative to expected growth. The losses associated with hybrid instruction were smaller, equal to 0.033 standard deviations for those that remained hybrid the whole year.

Perhaps the most striking finding in column 5 is that the consequences of hybrid and remote instruction were substantially larger in mid- and high-poverty schools than in low-poverty schools: the interaction between percent remote and high poverty was

[23] The variance in school effects increased by 81 percent between 2017–2019 and 2019–2021, as schools were differentially impacted during the pandemic. However, when we controlled for three variables (school poverty status, the percent of weeks remote/hybrid, and the interaction), the variance in school effects returned to 2017–2019 levels. Thus, the parameterization seemed to account for between 57 and 66 percent of the increase in variance (see online Appendix A).

[24] The differences by baseline score bounce back partially between columns 4 and 5 but remain far smaller than those in column 3. Apparently, the schools attended by low-baseline-score students are different in ways not captured by school poverty status or by percent remote/hybrid.

[25] In unpublished results, we added to column 5 interactions between student race/ethnicity and the share of the year schools were remote or hybrid. Conditional on the other covariates, the change in the race gaps in schools that were not remote or hybrid was similar to those reported in column 5.

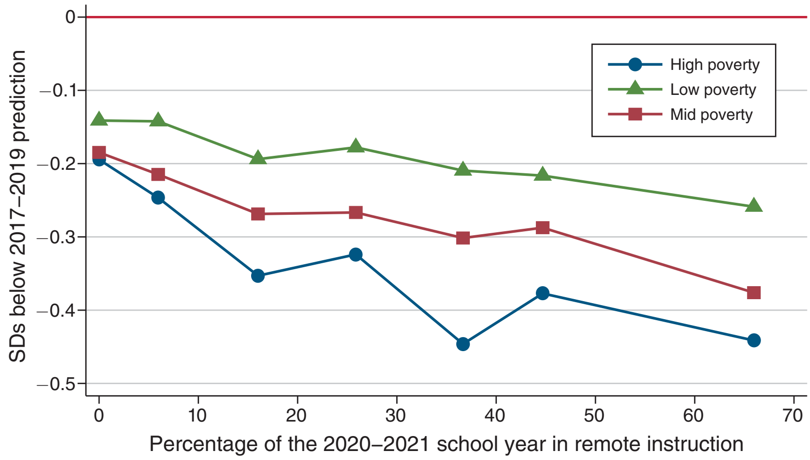

Figure 2. Pandemic Achievement Effects by Remote Schooling and School Poverty, Math

Notes: The vertical axis represents the difference between mean fall 2021 achievement and expected achievement based on prepandemic growth model estimates. The horizontal axis is the percentage of the 2020–2021 school year that a school was in remote instruction. Given the small number of districts that were remote all year, the top category of percent remote combines those that were remote between 50 and 100 percent of the year. Low-poverty schools had fewer than 25 percent of students receiving FRPL, while high-poverty schools had more than 75 percent of students receiving FRPL.

−0.158, which means that high-poverty schools that were remote all year lost 0.359 standard deviations (−0.158 − 0.201) more than high-poverty schools that were in person all year. High-poverty schools spending the year in hybrid instruction lost 0.151 standard deviations (−0.033 − 0.118) relative to high-poverty schools that remained in person. When we focus on within-district differences (by including district fixed effects in column 6), the losses associated with remote and hybrid instruction remained similar for mid- and high-poverty schools.

In column 7, we adjust for attrition by including the ratio of the number of tested students in the school to the number of students enrolled in the relevant grades in the school during the 2019–2020 school year. The substantive results are unchanged.

In Figure 2, we report the mean R_{i1} for schools by share of year in remote instruction and by school poverty. The vertical axis intercepts for the three lines are similar, implying that among those schools that were not remote during 2020–2021, the losses were similar for low-, medium-, and high-poverty schools—about 0.17 standard deviations on average. Presumably, such losses reflect some combination of the disruptions during spring 2020 (when all schools spent time in remote instruction) and the effect of pandemic-related stresses during 2020–2021. However, the gaps between high- and low-poverty schools are wider for schools that spent a larger share of the year in remote and hybrid instruction. For schools that spent more than 50 percent of the year in remote instruction, students in high-poverty schools lost roughly 0.44 standard deviations relative to prepandemic growth, while students in low-poverty schools lost 0.26 standard deviations.

In terms of standard deviation units, the losses were smaller in reading, but we see the same pattern of small racial/ethnic losses within schools and larger impacts of

TABLE 2—DECOMPOSING THE DIFFERENCE IN PANDEMIC ACHIEVEMENT GAINS BETWEEN HIGH- AND LOW-POVERTY SCHOOLS, MATH

	Amount	Percent of total
Total difference between high- and low-poverty schools	0.168	100%
Due to direct effects of:		
Race	0.014	8%
Baseline scores	0.016	9%
Conditional learning loss in high-poverty schools that were fully in person	0.002	1%
Due to differing incidence of remote and hybrid learning	0.051	30%
Due to differing effects of remote and hybrid learning	0.086	51%

Notes: Decomposition based on regression estimates from Table 1, column 5 and based on mean characteristics of high- and low-poverty schools in the analysis sample used in Table 1. The dependent variable is defined in equation (1) as the difference between a student's standardized 2021 fall NWEA MAP score and the expected score of students with similar baseline achievement and characteristics during the prepandemic period. See online Appendix B for details on the decomposition and online Appendix Table 6 for mean characteristics of high- and low-poverty schools.

remote and hybrid schooling on students attending mid- and low-poverty schools.[26] However, one substantive difference between math and reading is that gaps in reading achievement by school poverty and race did widen somewhat in districts that remained in person. While students learn math primarily in school, student learning in reading may depend more on parental engagement at home. Thus, the contrast between the math and reading findings for in-person districts may reflect differential family stresses outside of school.

Qualitatively, our findings are similar to Jack et al. (2023). Their finding that proficiency rates declined more in districts that had large shares of poor, Black, and Hispanic students is consistent with our finding that mean achievement declined more in high-poverty schools within districts that remained remote during 2020–2021. However, given that proficiency thresholds differ by state, it is not possible to infer impacts on mean achievement from their analysis (Ho 2008).

V. Disparate Incidence versus Disparate Impact of Remote and Hybrid Schooling

On average, high-poverty schools remained in remote instruction for longer than low-poverty schools. Moreover, in Table 1 and Figure 2, we presented evidence that students in high-poverty schools were more negatively impacted when schools went remote (i.e., disparate impacts of remote instruction). In Table 2, we now decompose the role played by the two factors—disparate incidence and disparate impacts—in widening the gap between low- and high-poverty schools. In the top row, we report the total difference in actual versus expected math achievement gains between high- and low-poverty schools, which is 0.168 standard deviations. As reported in the next two rows, a small share of this difference $(0.014 + 0.016)$ was due to the fact that Black and Hispanic students and students with low baseline achievement scores gained less and that those students were more likely to attend high-poverty schools. In the fourth row, we add in the differential loss in achievement gains between high- and

[26] See online Appendix Table 1.

low-poverty schools in areas that were in person throughout 2020–2021. As noted earlier, there was essentially no widening in math achievement gaps in districts that were fully in person (0.002 standard deviations). In the fifth row, we report the effect of greater incidence of remote/hybrid instruction in high-poverty schools, which was about one-third of the total difference (0.051/0.168). The remaining half of the gap (0.086/0.168) was due to the differing impact of hybrid/remote instruction on high-poverty schools.[27]

In reading, a larger share of the widening gap between high- and low-poverty schools occurred in areas that remained in person (26 percent).[28] Accordingly, the shares that were due to disparate incidence (19 percent versus 30 percent) and disparate impacts of remote/hybrid instruction (35 percent versus 51 percent) were lower in reading than in math.

VI. Paying for Academic Recovery

From the beginning of the pandemic through to the American Rescue Plan in spring 2021, the federal government provided state and local education agencies with $190 billion to pay for COVID-related expenses. States are required to allocate 90 percent of that funding to districts based on the Title I formula, which reflects child poverty rates and public assistance receipt in each district. Importantly, the funds were committed before the impact of the pandemic and instructional mode was clear. In this section, we provide a simple rule of thumb for judging whether the federal dollars are likely to be sufficient to pay for the catch-up in each district.

To put the achievement impacts and the federal aid on a comparable scale, we convert each into the share of each district's annual budget it represents. It is straightforward to convert the federal aid into an annual budget share, dividing each district's allocation by its spending on K–12 education in 2019–2020 (minus capital expenditures).

To convert recovery costs into an annual budget share, we use the share of a district's typical operating budget equivalent to the number of weeks of instruction lost. The NWEA data are especially well suited to this task. Unlike the official state tests, the NWEA's MAP assessment is administered at different points in the academic calendar for different schools. Thus, the test developers have observed how scores differ depending on the number of instructional weeks students received between test dates (which would yield unbiased estimates of gains per week of instruction as long as timing is exogenous; Thum and Kuhfeld 2020).[29] After using the parameters in column 5 of Table 1 to estimate the impacts on math scores, we divide by the average estimated gain per instructional week in grades 3–8 from NWEA to estimate the number of instructional weeks required for schools to get back to prepandemic growth expectations. We then divide the estimated weeks by 40 (the number of

[27] We describe the methodology for the decomposition in online Appendix B.

[28] See online Appendix Table 2.

[29] Because the tests are given in the spring and in the fall, the gains per instructional week during the school year do not include summer learning loss.

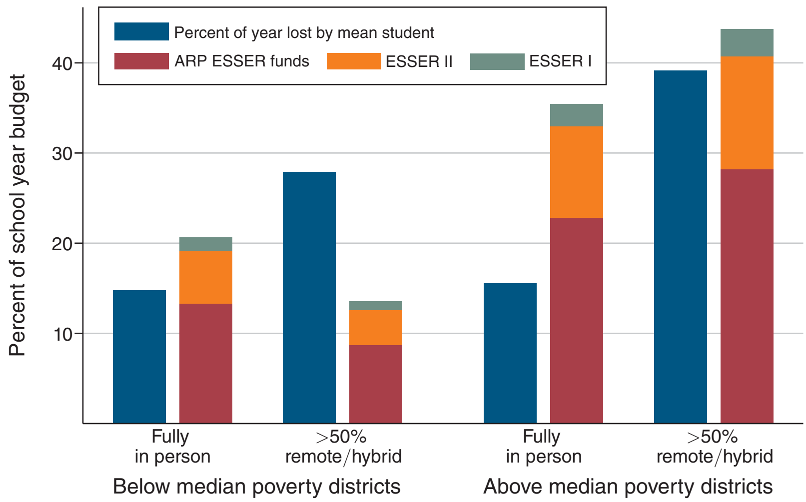

FIGURE 3. PANDEMIC ACHIEVEMENT LOSSES AND FEDERAL AID AS A SHARE OF ANNUAL SPENDING, MATH

Notes: Achievement effects were converted into weeks of instruction using NWEA growth norms and divided by a 40-week school year (to reflect the fact that salaries and operational expenses are paid by calendar weeks, not the number of instructional weeks in a school year, which is typically 36 weeks). Federal aid is reported relative to the district's annual budget for K–12 schooling, minus capital expenditures. High-poverty districts are the half of districts with the highest percent of students receiving FRPL (and low-poverty districts are the bottom half). Districts are considered "fully in person" if the AEI reports no remote or hybrid instruction in the district during the 2020–2021 school year.

calendar weeks in the typical school year) and aggregate to the district level (where ARP spending decisions will be made).[30]

Ours is likely to be a conservative estimate of cost. To make up 20 percent of a school year's worth of learning with supplemental interventions is likely to cost *more than* 20 percent of a district's typical annual spending. For instance, if a district were to extend the school year or lengthen the school day, it would have to pay teachers more than their normal wage rate (e.g., "time and a half"). Moreover, if students or teachers are tired at the end of the day or year, the marginal learning gain from additional time is likely to decline.

The correlation between the share of a year of unfinished learning and the share of an annual budget received in federal aid is positive (0.35) because both are positively related to poverty. But federal allocations across districts diverge from unfinished learning in key ways. We show this in Figure 3, which compares the shares of a school year required to eliminate the achievement loss and shares of annual budgets represented by federal aid. On the left are school districts that have below-median percentages of students receiving FRPL; on the right are the above-median (higher-poverty) districts. Within each, we report separately for districts that were fully in person during 2020–2021 and for those that spent the majority of the year remote or hybrid. (We excluded districts between the two extremes—those that were remote/hybrid for some weeks but less than half the year.)

[30]We assume that district operational expenditures are spread over 40 calendar weeks rather than the standard 36 instructional weeks (180 days).

Ironically, lower-poverty districts face the biggest gap between federal aid and remediation costs. Because the federal aid was based on the Title I formula, the lowest-poverty public school districts received less than 15 percent of their annual budgets in federal aid. The low-poverty districts that were remote or hybrid for most of the year lost 27 percent of a year's learning.

For high-poverty districts that remained in person, the losses were similar to those of low-poverty schools that remained in person (about 15 percent of a school year). However, because the federal dollars were based on poverty (and not achievement losses), they received considerably *more* funding (about a third of their annual budgets) than the 15 percent of a school year of unfinished learning their students experienced.

On the far right, we report the average losses for high-poverty districts that remained remote. This was the hardest hit group of districts, with achievement losses amounting to slightly under 40 percent of a year's worth of learning. That is roughly equivalent to the share of the annual budgets they received in federal aid.

The ARP only *requires* districts to spend 20 percent on academic recovery. According to an analysis of district plans by the nonprofit Future-Ed at Georgetown University, the average district is planning to spend not much more than the minimum on academic recovery (28 percent), with the remainder planned for facilities, technology, staffing, and mental and physical health.[31]

VII. Conclusion

Throughout the country, leaders made different choices about whether to hold classes in person or remotely during the COVID-19 pandemic. They had valid reasons to do so—including differing risks related to local demographics or population density as well as legitimate uncertainty about the public health consequences of in-person schooling. Regardless of possible public health benefits, the shifts to remote or hybrid instruction seem to have had profound consequences for student achievement. Where instruction was remote, achievement growth was lower for all subgroups but especially for students in high-poverty schools. Where instruction remained in person, there were still modest losses in achievement but there was no widening of gaps between high- and low-poverty schools in math (and less widening in reading).

It is possible that the relationships we have observed are not entirely causal. For instance, family stress may have contributed to the decline in achievement and forced school officials to keep school buildings closed for longer.[32] However, even if that were the case, our results have implications for the targeting of recovery efforts.

Although school leaders can readily measure the decline in their own students' performance, the bigger challenge is to formulate a commensurate plan of action.[33] We propose one relevant benchmark—start with the share of the typical school year's learning that was lost and calculate the equivalent share of the annual operating

[31] https://www.future-ed.org/financial-trends-in-local-schools-covid-aid-spending/

[32] We also included covariates for county COVID infection rates, district levels of Title 1 funding, and broadband access in the census tracts surrounding each school with little change in the results.

[33] In fact, proficiency rates could decline even if there were no change in mean achievement, if there were a narrowing in the variance in scores.

budget. It is likely a lower bound since the marginal cost per unit of growth from supplemental recovery efforts is likely to be higher than the average cost during a typical school year. Another approach is to convert the achievement loss into standard deviation units and compare against the effect sizes for relevant interventions. For instance, the average high-poverty school that remained in remote instruction for a majority of 2020–2021 lost roughly 0.44 standard deviations in achievement. For comparison, a literature review by Nickow, Oreopoulos, and Quan (2020) estimated the effect of high-dosage tutoring (defined as tutors working with fewer than four students three to five times per week for at least 30 minutes) to be 0.38 standard deviations in math. Thus, in high-poverty schools that remained remote, leaders could provide high-dosage tutoring to *every* student and still not make up for the loss.

Depending on whether they remained remote during 2020–2021, some school agencies have much more work to do than others. We plan to use the NWEA test results to measure the degree to which students affected by school closures catch up to prepandemic expectations in the coming school years.

REFERENCES

American Enterprise Institute. 2021. "Return to Learn Tracker Data, Version 45e." Return to Learn Tracker. https://www.returntolearntracker.net/instructional_status/ (accessed February 8, 2022).

Camp, Andrew M., and Gema Zamarro. 2021. "Determinants of Ethnic Differences in School Modality Choices During the COVID-19 Crisis." *Educational Researcher* 51 (1): 6–16.

Center for Systems Science and Engineering (CSSE). 2020. "COVID-19 Data Repository." Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19 (accessed December 16, 2021).

COVID-19 School Data Hub. 2022. "All School Learning Model Data." https://www.covidschooldatahub.com/data-resources (accessed June 8, 2022).

Curriculum Associates. 2020. *Understanding Student Needs: Early Results from Fall Assessments*. North Billerica, MA: Curriculum Associates.

Curriculum Associates. 2021a. *What We've Learned about Unfinished Learning: Insights from Midyear Diagnostic Assessments*. North Billerica, MA: Curriculum Associates.

Curriculum Associates. 2021b. *Academic Achievement at the End of the 2020–2021 School Year: Insights after More Than a Year of Disrupted Teaching and Learning*. North Billerica, MA: Curriculum Associates.

Darling-Aduana, Jennifer, Henry T. Woodyard, Tim R. Sass, and Sarah S. Barry. 2022. "Learning-Mode Choice, Student Engagement, and Achievement Growth During the COVID-19 Pandemic." CALDER Working Paper 260-0122.

Dorn, Emma, Bryan Hancock, Jimmy Sarakatsannis, and Ellen Viruleg. 2020. "COVID-19 and Learning Loss—Disparities Grow and Students Need Help." *McKinsey & Company*, December 8. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-and-learning-loss-disparities-grow-and-students-need-help.

Goldhaber, Dan, Thomas J. Kane, and Andrew McEachin. 2021. "Analysis: Pandemic Learning Loss Could Cost U.S. Students $2 Trillion in Lifetime Earnings. What States & Schools Can Do to Avert This Crisis." *The 74*, December 13. https://www.the74million.org/article/analysis-pandemic-learning-loss-could-cost-u-s-students-2-trillion-in-lifetime-earnings-what-states-schools-can-do-to-avert-this-crisis/.

Goldhaber, Dan, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger. 2023. "Replication data for: The Educational Consequences of Remote and Hybrid Instruction during the Pandemic." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E179641V1.

Grossmann, Matt, Sarah Reckhow, Katharine O. Strunk, and Meg Turner. 2021. "All States Close but Red Districts Reopen: The Politics of In-Person Schooling During the COVID-19 Pandemic." *Educational Researcher* 50 (9): 637–48.

Ho, Andrew Dean. 2008. "The Problem with 'Proficiency': Limitations of Statistics and Policy Under No Child Left Behind." *Educational Researcher* 37 (6): 351–60.

Jack, Rebecca, Clare Halloran, James Okun, and Emily Oster. 2023. "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts." *American Economic Review: Insights* 5 (2): 173–90.

Kilbride, Tara, Bryant Hopkins, Katharine O. Strunk, and Scott Imberman. 2021. *K–8 Student Achievement and Achievement Gaps on Michigan's 2020–21 Benchmark and Summative Assessments*. East Lansing, MI: Education Policy Innovation Collaborative.

Kogan, Vladimir, and Stéphane Lavertu. 2021. *The COVID-19 Pandemic and Student Achievement on Ohio's Third-Grade English Language Arts Assessment*. Columbus, OH: John Glenn College of Public Affairs, The Ohio State University.

Kuhfeld, Megan, Erik Ruzek, Karyn Lewis, Jim Soland, Angela Johnson, Beth Tarasawa, and Lindsay Dworkin. 2021. *Understanding the Initial Educational Impacts of COVID-19 on Communities of Color*. Portland, OR: NWEA.

Lewis, Karyn, and Megan Kuhfeld. 2021. *Learning during COVID-19: An Update on Student Achievement and Growth at the Start of the 2021–22 School Year*. Portland, OR: NWEA.

Lewis, Karyn, Megan Kuhfeld, Erik Ruzek, and Andrew McEachin. 2021. *Learning during COVID-19: Reading and Math Achievement in the 2020–21 School Year*. Portland, OR: NWEA.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. "IPUMS National Historical Geographic Information System, Version 16.0." IPUMS. http://doi.org/10.18128/D050.V16.0 (accessed March 23, 2022).

Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77 (2): 251–66.

National Center for Education Statistics. 2019. "Composite School District Boundaries." US Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries (accessed March 23, 2022).

National Center for Education Statistics. 2020. "Public Elementary/Secondary School Universe Survey 2019–20, Version 1a"; "School District Finance Survey (F-33), Version 1a." US Department of Education, Institute of Education Sciences. https://nces.ed.gov/ccd/elsi/tableGenerator.aspx (accessed July 22, 2022).

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869–95.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2020. "The Impressive Effects of Tutoring on PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." NBER Working Paper 27476.

NWEA. 2021. "Growth Research Database." Accessed December 16, 2021.

Office of Elementary and Secondary Education. 2020. *Elementary and Secondary School Emergency Relief Fund State Allocation Table*. Washington, DC: US Department of Education.

Office of Elementary and Secondary Education. 2021. *American Rescue Plan Elementary and Secondary School Emergency Relief Fund Revised State Allocations Table*. Washington, DC: US Department of Education.

Oster, Emily, Rebecca Jack, Clare Halloran, John Schoof, Diana McLeod, Haisheng Yang, Julie Roche, and Dennis Roche. 2021. "Disparities in Learning Mode Access Among K–12 Students During the COVID-19 Pandemic, by Race/Ethnicity, Geography, and Grade Level—United States, September 2020–April 2021." *Morbidity and Mortality Weekly Report* 70 (26): 953–58.

Parolin, Zachary, and Emma K. Lee. 2021. "Large Socio-economic, Geographic, and Demographic Disparities Exist in Exposure to School Closures." *Nature Human Behavior* 5 (4): 522–28.

Thum, Yeow Meng, and Megan Kuhfeld. 2020. *NWEA 2020 MAP Growth Achievement Status and Growth Norms for Students and Schools*. Portland, OR: NWEA.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113 (485): F3–33.

US Census Bureau. 2020. "CC-EST2019-AGESEX-[ST-FIPS]: Annual County and Puerto Rico Municipio Resident Population Estimates by Selected Age Groups and Sex: April 1, 2010 to July 1, 2019." https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/asrh/ (accessed December 16, 2021).

US Department of Education. 2019. "Estimated ESEA Title I LEA Allocations—FY 2019." https://www2.ed.gov/about/overview/budget/titlei/fy19/index.html (accessed March 29, 2022).

US Department of Education. 2020. "Estimated ESEA Title I LEA Allocations—FY 2020." https://www2.ed.gov/about/overview/budget/titlei/fy20/index.html (accessed February 7, 2022).

World Bank, UNESCO, and UNICEF. 2021. *The State of the Global Education Crisis: A Path to Recovery*. Washington, DC, Paris, and New York: World Bank, UNESCO, and UNICEF.