JUNE 2022

DOES STUDENT INSTRUCTIONAL MODALITY PREDICT STUDENT ACHIEVEMENT?

LEARNING DURING THE PANDEMIC IN ILLINOIS SERIES
PART 3

WRITTEN BY:

Sarah Cashdollar, Mariana Barragan Torres, Yi Wang & Meg Bates

Illinois Workforce and Education Research Collaborative (IWERC)
Discovery Partners Institute
University of Illinois System

Illinois State Board of Education

IWERC

DPI

PART OF THE UNIVERSITY OF ILLINOIS SYSTEM

Suggested Citation:

Cashdollar, S., Barragan Torres, M., Wang, Y., & Bates, M. (2022). Does Student Instructional Modality Predict Student Achievement? *Learning During the Pandemic in Illinois Series*. Chicago, IL: Illinois Workforce and Education Research Collaborative (IWERC), Discovery Partners Institute, University of Illinois. https://dpi.uillinois.edu/applied-research/iwerc/current-projects/learning-modalities/

External Review:

This report was reviewed by scholars to ensure that its contents are rigorous, accurate, and useful to educators and policymakers with varying levels of background knowledge. The reviewers of this report included:

- Paul Bruno, Ph.D. (University of Illinois Urbana-Champaign)
- John Q. Easton, Ph.D. (UChicago Consortium on School Research)
- Elliot Regenstein (Foresight Law + Policy)
- Jose Eos Trinidad (University of Chicago)

We deeply thank these reviewers for their time and expertise.

Acknowledgments

We gratefully acknowledge the support of the Research Department and the Data Strategies and Analytics Department at ISBE, who provided support with accessing data, understanding data elements, and interpreting findings. This study was improved by feedback from Guanglei Hong of University of Chicago, who provided guidance on statistical methods.

Last revised: 5/23/2022

EXECUTIVE SUMMARY

For over two years, schools, teachers, students, and families across Illinois have worked to adapt instruction and learning under the ever-changing conditions wrought by COVID-19. This report, the third and final in the *Learning During the Pandemic in Illinois* series, investigates how much time students spent learning in-person and/or remotely during the 2020-21 school year (SY21), and how it related to students' achievement outcomes for grades 5-8. These findings are relevant for school leaders and policymakers as they implement learning recovery efforts.

As detailed in the first report of this series, *Trends in Instructional Modality During the 2020-21 School Year*, remote instruction was not experienced evenly across schools with different student populations in Illinois. In schools serving higher proportions of White students, students spent more time learning in-person during the SY21, on average, while schools serving higher proportions of Black, Latinx, low-income, and English learner (EL) students spent more time instructing remotely.

In the second report in this series, *Does School Instructional Modality Predict Average School Achievement?*, we showed that while all schools in Illinois declined in average test scores, schools serving grades 3-5 performed worse on standardized tests when students spent more time in remote learning. For schools serving grades 6-8 and/or 11, the relationships between learning modality and achievement were smaller and mixed.

The current report builds upon the previous two. For students in grades 5-8, we dig deeper into understanding instructional modality at the level of the individual student. We ask:

1. How did the proportion of time learning in-person versus remotely vary for students in different demographic groups in Illinois?
2. Was student instructional modality in SY21 related to average student achievement, controlling for student and school characteristics?
3. Did the relationship between modality and achievement vary depending on the school that a student attended?

We show that, across schools, the proportion of time students spent learning in-person was positively related to their test scores in English Language Arts (ELA) and math.

Key Findings

The amount of time students attended in-person versus remotely varied dramatically across student subgroups.

- On average, a typical Illinois student attended 33% of SY21 in-person, attended 61% remotely, and was absent 7% of days.
- Statewide, younger students (grades 5 and 6) spent slightly more time in-person than older students (grades 7 and 8).
- On average, White students spent 50% of the year in-person, while Latinx and English learner (EL) students spent 17% of the year in-person, and Black students spent just 14% of the year in-person.
- Students eligible for Free/Reduced-Price Lunch (FRPL) spent 24% of the year in-person, on average.
- Students with Individualized Education Programs (IEPs) attended in-person at about the same rate (34%) as the state average.

After controlling for student and school characteristics, there was a statistically significant and positive relationship between in-person instruction and Illinois Assessment of Readiness (IAR) test scores.

- Attending in-person versus remotely all year was associated with higher scores by 5.4 points (0.15 SDs) in ELA and 7.5 points (0.22 SDs) in math.
- Variation in instructional modality across student subgroups likely exacerbated racial/ethnic and socioeconomic achievement disparities.

Learning During the Pandemic in Illinois

Part 3: Does Student Instructional Modality Predict Student Achievement?

Background: Instructional Modality in Illinois

In July of 2020, the Illinois State Board of Education (ISBE) set forth guidance on learning modality for SY21. The Board "strongly recommend[ed] in-person learning," especially for students with IEPs, ELs, and students under age 13 (ISBE, 2020, p. 5). However, the Board also acknowledged that decision-making about whether in-person learning was safe and appropriate would vary across students, schools, and districts in differing communities. In turn, ISBE provided recommendations for instructional programs that were in-person, remote, and a combination of in-person and remote. Recommendations focused on communication with caregivers, social and emotional support, student engagement, professional development, accessibility, and assessment, among other topics.

In *Report 1* of this series, we showed that SY21 instructional modality varied dramatically across Illinois schools and across time (from the beginning to the end of the school year). This variation reflected cross-community differences both in district/school decision-making and in student and family choices given the options offered by their schools. Within districts that offered both in-person and remote modalities, modality patterns were driven largely by whether or not students and their families opted to return to in-person learning. Just over half of schools began the year remotely. Of these schools, the majority (6 out of 7) transitioned to hybrid or in-person instructional modalities over the course of the year. In about one-third of schools, students attended in-person for the majority of the year.

We also showed that schools' instructional modalities were strongly related to their demographic characteristics. Schools serving higher proportions of White students spent more of the year in-person, on average. Schools with higher proportions of Black, Latinx, low-income, and EL students experienced more remote instruction. Elementary and middle schools were more likely than high schools to start the year remotely and then transition into hybrid or in-person modalities, while high schools were more likely to offer both remote and in-person instruction throughout the year.

Research suggests several reasons to expect that students who returned to in-person instruction early in SY21 experienced fewer learning disruptions than those who were instructed remotely. Teachers and administrators reported challenges with instructional time and student engagement in remote learning, while parents reported decreased educational quality, and students experienced lower social and emotional wellbeing (Bartlett, 2022; Duckworth et al., 2021; Hanno et al., 2022; Kaufman & Diliberti, 2021; Rapaport et al., 2020; Trinidad, 2020). Research has found that pandemic-related challenges to learning were particularly deleterious for certain student subgroups, including students who were Black, Latinx, low-income, ELs, students with disabilities, and students experiencing homelessness (see Appendix A for a discussion of this literature). Overall, remote instruction during the pandemic has been estimated to reduce student performance on standardized tests nationwide (Darling-Aduana et al., 2022; Domina et al., 2022; EPIC, 2021; Goldhaber et al., 2022; Halloran et al., 2021; Kogan & Lavertu, 2021).

Consistent with these findings, in *Report 2* of this series we showed that in-person instruction predicted substantially smaller test score declines compared to remote instruction for Illinois schools serving grades 3-5. We estimated that being in a school that was in-person rather than remote for most of the year improved average test scores for schools serving grades 3-5 by 14 scale score points (0.40 SDs) in math and 8 points (0.19 SDs) in ELA. However, for schools serving grades 6-8 and high schools with students tested in grade 11, relationships were small and varied.

In the present study, we delve into the relationship between instructional modality and achievement at the student, rather than the school, level. We first explore differences in the amount of time students learned in-person across racial/ethnic groups, FRPL-eligible students, students with IEPs, ELs, and homeless students. This report is the first that we know of that uses student-level data to detail how student subgroups across Illinois differed in the amount of in-person instruction they experienced in SY21.

We then use multilevel models, controlling for student and school characteristics, to understand how the amount of time students spent learning in-person was related to ELA and math outcomes on the Illinois Assessment of Readiness (IAR). Because we analyze only cohorts that were also grades tested in SY19, we cannot include as many grade levels in this analysis as we were able to include in our *Report 2* school-level analysis, which analyzed data for grades 3-8 and 11. However, the student-level data for grades 5-8 allows us to model relationships using the full range of scores and characteristics for these students statewide (whereas the use of school-level data limited us to using school averages). Therefore, our findings allow us to verify and expand upon a subset of the findings from our previous report.

Method

Data

We examined the relationship between instructional modality and SY21 test scores, controlling for student and school characteristics as well as SY19 test scores. We drew on student-level achievement, attendance, and demographic data shared with our team by ISBE. We also controlled for school characteristics using publicly available data from the Illinois Report Card in SY19.

Illinois administers federally mandated annual state testing for grades 3-8 and 11, although testing was cancelled in SY20 due to the onset of the pandemic. We limited our analyses to students in grades that were tested in SY21 who were also in grades tested in SY19 so that we could control for their SY19 test scores. This limited our sample to students in grades 5-8, who were in grades 3-6 in SY19 (n=579,323). This sample includes students in major metropolitan areas like Chicago. We further limited the sample to students who were not missing data on test scores or modality of attendance, for a final n=346,596. We further discuss missing data below.

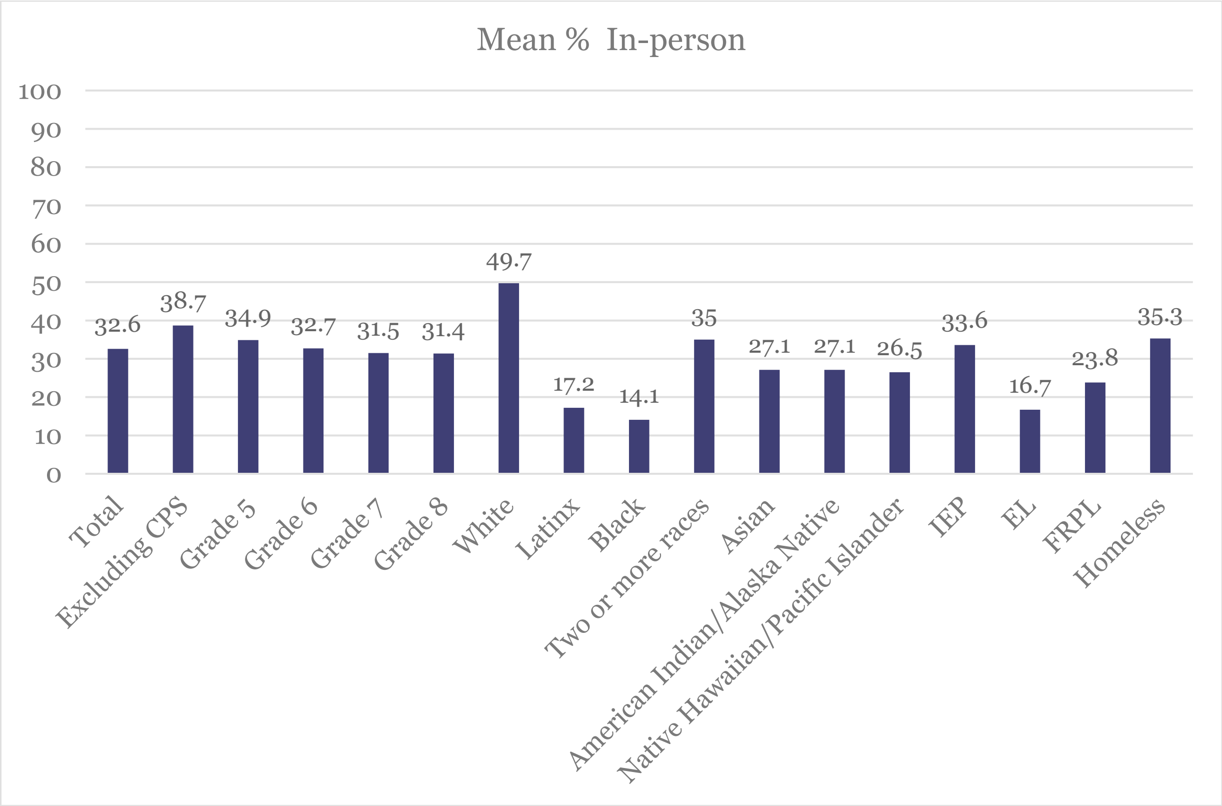
Student Achievement. We measured achievement using student scale scores on the SY21 IAR exams in ELA and math. In order to compare our findings to findings from other educational intervention studies, we also presented findings of our estimates in terms of SY19 standard deviations (SDs). On average, students in grades 5-8 scored lower in SY21 than students who were in grades 5-8 in SY19 – by 6.5 points (0.18 SDs) in ELA and 6.1 points (0.18 SDs) in math.

Proportion of Time Learning In-Person. Proportion of time learning in-person was measured using student-level data on daily mode of attendance in SY21. We calculated the number of days a student attended in-person (as opposed to remotely) as a percentage of the total number of school days (i.e., sum of days absent and present) in SY21. Data on proportion in-person was missing for just 0.03% of all students in grades 5-8. Our models estimate the association between student achievement and a one percentage point increase in the proportion of time a student spent learning in-person.

As shown in Figure 1, students statewide attended about 33% of school days in-person, on average. We also show a slightly higher mean proportion of days in-person of 39% when excluding students in CPS (because CPS students made up 13%[a] of the grades 5-8 student population in this sample and averaged just 5% of days in-person, their inclusion influences the mean for students outside of CPS). Students in grades 5 and 6 spent slightly more time in-person than students in grades 7 and 8.

There was wide variation in instructional modality across ethnic/racial groups. White students attended in-person for half of the year, on average, while Black students were in-person for just 14% of the year, and Latinx students for just 17%. ELs and FRPL-eligible students also spent less time than the state average attending in-person.

Figure 1
Average Percentage of SY21 Proportion In-Person Across Demographic Groups



[a] Students in CPS made up 18.4% of all Illinois students in SY21. Their proportion is lower in our sample due to relatively low SY21 test participation.

Controls. We controlled for a number of student and school characteristics that could hypothetically confound the relationship between in-person instruction and achievement.[b] At the student level (level 1), we controlled for race/ethnicity (White, Latinx, Black, Asian, Two or more races, American Indian/Alaska Native, Native Hawaiian/Pacific Islander), gender, IEP participation, EL status, FRPL eligibility, and homelessness. We assigned students to schools based on where they spent the majority of the year. School controls (level 2) included SY21 fall enrollment, proportions of students in each racial/ethnic group, students eligible for FRPL, students with IEPs, students classified as ELs, and students experiencing homelessness.

Using Illinois Report Card data, we also indicated whether a school belonged to a district that was classified as Evidence-Based Funding (EBF) Tier 4. The state categorizes all districts into an EBF Tier, ranging from Tier 1 to Tier 4, which reflects the extent to which local financial resources meet students' educational needs (ISBE, n.d.). Districts in Tier 4 are determined to have local funding exceeding that required to meet educational needs. This excess funding could, hypothetically, affect both test scores and the modality options schools provide to students. For example, schools with more funding nationally have been reported to implement mitigation strategies more quickly, making in-person instruction viable sooner (Godoy, 2022). Accordingly, we accounted for EBF Tier 4 status, which was the same in SY19 and SY21 for all schools.

Additional school-level controls were based on schools' characteristics in SY19. These controls included SY19 average attendance rate and school score on the Involved Families indicator of the *5Essentials* survey of students and teachers, which is administered at all Illinois schools annually (ISBE, 2022). These variables were correlated with the proportion of SY21 that students spent in-person as well as their ELA and math scores, potentially confounding the relationship between instructional modality and achievement. We did not control for these school characteristics in SY21, unlike our controls for demographic characteristics, because both were likely affected by instructional modality over the course of the year. If remote instruction caused lower attendance rates and/or changed families'

[b] Confounders refer to variables that influence the dependent and independent variables at the same time (Greenland et al., 1999). Statistically controlling for them allows us to better understand the relationship of interest – in this case, the relationship between in-person instruction and average student achievement.

level of involvement in school, then controlling for SY21 attendance or Involved Families would reduce our ability to understand variation accounted for by modality.[c][d]

Missing Data

Student test participation was much lower in SY21 compared to SY19. As detailed in *Report 2* of this series, the decline in test participation could be attributed to remote instruction itself. Students were required to attend school in-person in order to participate in testing, even if they were otherwise instructed remotely. This appeared to reduce participation in schools that spent more of the year in remote learning. These schools were disproportionately Black, Latinx, and FRPL-eligible. These student subgroups were especially likely to have missing data for SY21, as shown in Table 1. Overall, just 60% (346,596 out of 579,323 students) of our sample has test score data from both SY19 and SY21.

The amount of missing data and the way it is distributed presents a problem for understanding the relationship between instructional modality and achievement. As shown by comparing Figure 1 and Table 1, students who spent the least amount of time learning in-person were the most likely to have missing test scores in SY21, on average. Table 2 displays descriptive characteristics of students with any data ("All students") and students with no missing data on test scores ("Complete Cases"). As discussed, the sample of complete cases has higher rates of in-person instruction and lower proportions of Latinx and Black students as well as FRPL-eligible students and English learners than the sample of students with any data statewide. The sample of complete cases also has a larger proportion of White students.

[c] While it is possible that the demographic composition of schools also could have been affected by modality via student mobility in response to modality options, we believe (especially after comparing SY19 and SY21 demographic characteristics by school) that the impact of these demographic changes due to modality on test scores were minor. The impact of peer demographic composition on test scores, on the other hand, may be substantial (Sacerdote, 2014). We therefore considered it important to control for SY21 demographic composition.

[d] We considered several additional school-level controls, including scores on the other *5Essentials* Survey measures and hospital beds in use in the school's IDPH region, by month, during SY21. However, these controls did not improve goodness-of-fit according to Likelihood Ratio Tests (LRTs), so we did not include them in our final models.

Table 1
Missing Test Score Percentages Across Demographic Groups

	2019		2021	
	ELA % Missing	Math % Missing	ELA % Missing	Math % Missing
All Students	7.65	7.77	38.75	39.61
White	6.14	6.20	21.52	22.19
Hispanic/Latinx	6.49	6.61	48.18	49.08
Black	9.18	9.41	49.94	51.34
Asian	10.77	10.81	38.67	39.49
Two or more races	8.46	8.55	34.01	34.81
American Indian/Alaska Native	9.90	9.96	51.90	51.11
Native Hawaiian/Pacific Islander	11.66	11.99	40.56	41.87
IEP	11.62	11.91	40.13	41.11
FRPL	7.35	7.49	43.47	44.48
Homelessness indicator	16.78	17.1	40.52	41.69
English Learners	12.24	12.39	49.90	50.92

In turn, calculations based on complete cases do not precisely represent the relationship of interest at the lowest levels of in-person instruction. Estimates also cannot be considered representative of demographic subgroups least likely to participate in SY21 testing. However, we believe it is likely that estimates for students who are not represented in our dataset are at least as large as the estimates we find using the complete case sample. As discussed, research has found that Black, Latinx, low-income, and EL students faced difficult learning challenges in remote instruction. This research is summarized in Appendix A. For these subgroups that are underrepresented in our data, it is probable that the relationship between modality (remote or in-person learning) and achievement was even stronger than for the average student statewide. Therefore, we likely *underestimated* these relationships.

Analytic Strategy

We used two-level Hierarchical Linear Models (HLM) to estimate the relationship between a student's proportion of days learning in-person during SY21 and IAR test score outcomes. HLM is a statistical technique for data that is organized hierarchically–where smaller units (in our case, students) are nested within larger units (in our case, schools). Students who attend the same school are subject to a range of similar school-level influences, such as school leadership, discipline policy, or extracurricular options. As such,

Table 2
Descriptive Student Characteristics, Grades 5-8 (SY21)

Variable type	Variable	Mean All Students (n=579,323)	Mean Complete Cases (n=346,596)
Instructional Modality (%)	In-Person Instruction	32.6	43.6
	Remote Instruction	60.8	51.1
	Absent	6.6	5.3
Outcomes	ELA 2021	731.9	732.9
(SY21 IAR scale scores: Grades 5-8)	Math 2021	725.9	726.5
Prior Achievement	ELA 2019	736.7	740.1
(SY19 IAR scale scores: Grades 3-6)	Math 2019	734.1	737.5
Student Covariates (%)	Race/Ethnicity		
	White	45.8	56.6
	Hispanic/Latinx	27.7	21.9
	Black	16.9	12.4
	Asian	5.3	4.8
	Two or more races	3.9	3.9
	American Indian/Alaska Native	0.3	0.2
	Native Hawaiian/Pacific Islander	0.1	0.1
	Other Demographics		
	Female	49.0	48.6
	FRPL	47.8	41.1
	IEP	14.3	13.1
	English learner	11.7	8.3
	Homeless	1.1	0.9

Note: Prior achievement refers to students' IAR scores from SY19, when they were in grades 3-6.

we expect that students within the same school are more similar in a number of ways–both observed and unobserved–than students in different schools. Accounting for this hierarchical data structure is important in order to obtain accurate, unbiased estimates (Raudenbush & Bryk, 2001). HLM provides these estimates by parsing variation across schools and students. It also allows us to understand how the relationship between modality and achievement varies based upon the school that a student attends.

We nested students (level 1) within schools (level 2), running models separately for ELA and math. As a first step, we ran an intercept-only model (unconditioned), which allows us to explore how much of the variance in student scores occurs between schools. The intercept model looks as follows:

Level-1 Model.

$Y_{ij} = \pi_{0j} + e_{ij}$

Level-2 Model.

$\pi_{0j} = \beta_{01} + r_{0j}$

Where, at level 1 (student-level),

Y_{ij} is the ELA or math score for student i in school j;

π_{0j} is the intercept and represents mean of the SY21 test scores;

and at level 2 (school-level), β_{01} is the school-level intercept, which represents average test scores across schools.

The second step in our process was to estimate fully conditional models, or models that include our primary independent variable (% in-person) as well as student and school controls. We allowed for random effects when estimating the slope of the intercept π_0, percent of the days in-person π_1, and 2019 test score π_{2i}, which means that we allowed these estimates to vary across schools. For ease of interpretation, we used fixed effects to estimate the slopes for student demographic characteristics $\boldsymbol{X}$ – we did not allow these estimates to vary across schools. Now, our models are as follows:

Level-1 Model.

$Y_{ij} = \pi_{0j} + \pi_{1j}(perInp_{ij}) + \pi_{2j}(2019score_{ij}) + \boldsymbol{\Omega X}_{ij} + e_{ij}$

Level-2 Model.

$\pi_{0j} = \beta_{01} + \boldsymbol{\Omega D}_j + \beta_{02}(AttendRate_j) + \beta_{03}(Enrollment_j) + \beta_{04}(EBFTier4_j)$

$+ \beta_{05}(5InvolvedFams_j) + r_{0j}$

$\pi_{1j} = \beta_{11} + r_{1j}$

$\pi_{2j} = \beta_{21} + r_{2j}$

$\pi_{kj} = \beta_{k1}$

Where, at level 1 (student-level),

Y_{ij} is the SY21 ELA or math score for student i in school j;

π_{0j} is the intercept;

π_{1j} is the coefficient on proportion of SY21 spent in-person for student *i* in school *j*;

π_{2j} is the coefficient on ELA/math scale score in SY19 for student *i* in school *j*;

$\boldsymbol{X}_{ij}$ is a vector of demographic characteristics for student *i* in school *j,* including race/ethnicity (White, Latinx, Black, Asian, Two or more races, American Indian/Alaska Native, Native Hawaiian/Pacific Islander), gender, FRPL, IEP, EL, and homelessness.

At level 2 (school-level),

β_{01} represents the school-level intercept, which corresponds to the average test score across schools;

$\boldsymbol{D}_j$ is a vector for SY21 demographic characteristics for school *j*, including proportions of students by race/ethnicity and proportions of low-income students, students with IEPs, EL students, and students experiencing homelessness;

β_{02} is the coefficient for school *j*'s SY19 attendance rate,

β_{03} is the coefficient for school *j*'s SY21 fall enrollment;

β_{04} is the coefficient for a dummy for whether a school *j* was classified as EBF Tier 4; and

β_{05} is the coefficient for school *j*'s SY19 score on the *5Essentials* measure for involved families.

Complementary Analyses

We ran additional analyses to see how our estimates changed when excluding two subgroups of students: students who tested in the fall rather than the spring of 2021, and students in CPS. Due to disruptions caused by the pandemic, ISBE gave districts the option of delaying spring 2021 testing to the fall of 2021. Students in these schools experienced additional months of learning and a summer break, both of which could have influenced their test score outcomes (Papay, 2011). If these students also spent more or less time in remote instruction than other students, then including them in our analyses may bias our estimates of how modality related to achievement. We developed an indicator based on schools that reported to ISBE that they had tested in the spring – we assumed all other schools to be "fall testers." Overall, we estimated that 1.6% of students in this sample attended schools that tested in the fall, and on average they spent 35% of the year learning in-person (compared to 44% of the complete-case sample). This indicator is not precise given that some schools that tested in the spring may have failed to report the proper testing window. However, it allows us to develop a rough estimate of how sensitive our findings are to the exclusion of students in schools that mostly tested in the fall (resulting in n=341,371).

We also considered it important to analyze how much our findings were influenced by students in CPS, who made up 12.9% of our sample and spent 10% of the year learning in-person, on average (among our sample of complete cases). Because CPS makes up a relatively large proportion of students in the state, any unobserved district-level confounders from this single district have the potential to disproportionately impact the size of our estimates. In turn, we analyzed how our estimates changed when excluding CPS students (resulting n = 301,842).[e]

Finally, we reran our analyses with the addition of district fixed effects at level 2. Fixed effects control for district characteristics that are difficult to measure (e.g., the role of teacher unions and varied district administration) and may confound the associations between modality and student achievement. A caveat of using district fixed effects is that these models do not account for how differences in modality policy across districts changes the relationship between in-person learning and achievement. Further discussion of fixed effects models, and the results of our models with district fixed effects, are presented in Appendix C.

Results

Results from the intercept-model are shown in Table 3. Estimates of the intra-class correlation (ICC) show that approximately 22% of the variation in average SY21 student test scores occurs between schools. As such, nesting students within schools is necessary in order to account for school-level influences on the relationship between modality and achievement. Table 3 also displays residual variance, or variation that is not accounted for by the model. We see that, by not including any predictors in the model, we have large residual variances.

[e] Descriptive characteristics of our samples excluding fall testers and CPS are shown in Appendix Table B1.

Table 3
Intercept-Only Model Results

	ELA	Math
Intercept	728.61***	722.58***
	(0.34)	(0.35)
ICC	0.22	0.25
	(0.01)	(0.01)
Random-effect parameters (Variances)		
Schools	274.30	297.66
	(8.24)	(8.82)
Residual	980.89	876.03
	(2.37)	(2.11)

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10.

Estimates from our fully conditional models are shown in Table 4. The first row represents the estimated change in IAR ELA and math scores for each additional percentage point of days attended in-person during SY21. The second row shows the effect sizes (Cohen's *d*, calculated using SY19 student-level standard deviations). We see that for each additional percentage point of in-person instruction in SY21, which is equivalent to about 1.7 days, IAR test scores in ELA and math were higher by 0.05 and 0.08 points, respectively. In other words, students who attended in-person scored higher than those who attended remotely, on average, when controlling for student and school characteristics.

In pragmatic terms, the size of this relationship means that test scores in ELA are predicted to be about 5.4 points higher after a year in-person compared to fully remote. In math, a year of learning in-person predicts scores that are 7.5 points higher compared to a year of learning remotely. The typical Illinois student attended 33% of the year in-person. If that student instead had attended 53% of the year in-person, we would expect them to score 1.08 points higher in ELA and 1.5 points higher in math.

The random effect parameter of .008 on in-person instruction suggests that, after accounting for student and school covariates, very little variation remains between schools in the relationship between modality and achievement. In other words, the relationship is constant across students and schools that are similar on the observed covariates. We also see that residual variances were reduced by half, meaning that our models were able to explain a large proportion of the variance in test scores across students.

Table 4
Estimated associations between percentage of SY21 in-person and IAR scale scores, grades 5-8

	ELA	Math
SY21 Scale Score	0.054***	0.075**
	(0.004)	(0.003)
Effect size	.0015	.0022
	(.0001)	(.0001)
Constant	220.39	223.82
	(11.39)	(10.47)
ICC	0.00	0.00
	(0.00)	(0.00)
Random Effect Parameters (Variance)		
In-Person Instruction	0.008	0.007
	(0.001)	(0.001)
SY19 Scores	0.000	0.000
	(0.00)	(0.00)
School (constant)	0.000	0.000
	(0.00)	(0.00)
Residual	469.17	394.22
	(1.13)	(0.95)
N Schools	2,600	
N Students	346,596	

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05
Note: Effect sizes are Cohen's *d*. Controls included the following student characteristics: ELA/math IAR score in 2019, race/ethnicity, gender, eligibility for Free/Reduced-Price Lunch (FRPL), English learner (EL) status, Individualized Education Program (IEP) participation, Homelessness; and 2019 school characteristics: % White, % Black, % Hispanic/Latinx, % Two or more races, % Low-income, % Homeless, % ELs, % IEP, mean attendance rate, indicator for Evidence-based Funding Tier 4, and scores on the *5Essentials* measure for Involved Families.

Results from our complementary analyses are displayed in Tables B2, B3, and C1 in the Appendix. As shown in Table B2, excluding presumed fall testers made no change to our estimates. Excluding students in CPS increased our estimates slightly, by 0.008 points in both ELA and math (see Appendix Table B3). This finding suggests that for CPS students, on average, there was a smaller relationship between in-person instruction and achievement than for students in other districts statewide, and this slightly decreased the statewide estimates. Including district fixed effects also increased our estimates slightly (see Appendix Table C1). This finding suggests that, when limiting our comparisons to students in districts with similar modality options, students who spent more time learning in-person scored even higher relative to students who stayed remote. The overall similarity between our models with and without district fixed effects suggests that unobserved district characteristics are not creating significant bias in our estimates. We further discuss findings from models with district fixed effects in Appendix C.

How Much Stock Should We Put in These Findings?

Our study controls for several potentially confounding variables between instructional modality and student outcomes, including student test scores from SY19. However, it is possible that other, unobserved variables influenced both a student's propensity to learn in-person and their SY21 test score outcomes (Bacher-Hicks & Goodman, 2021). In turn, our estimates should not be interpreted causally. Instead, our findings indicate "promising evidence" for the benefits of in-person instruction over remote, within the contexts studied, according to the Every Student Succeeds Act (ESSA) of 2015 evaluation framework.

The large amount of missing data means our estimates should be interpreted cautiously. Students who did not participate in SY21 testing had relatively low rates of in-person instruction. Because they also disproportionately represented demographic groups that faced additional challenges in remote learning, we believe that our findings underestimate the size of the positive relationship between in-person learning and achievement.

While we aggregate grades 5-8 in our analyses, estimates are likely to differ across grade levels, a possibility for future work. Findings from grades 5-8 also may not generalize to other grade levels. In *Report 2* of this series, *Does School Instructional Modality Predict Average School Achievement?*, we showed that instructional modality was more strongly related to test score outcomes in schools serving grades 3-5 than those serving grades 6-8, and we found no significant relationships for grade 11. Other studies have also found greater losses for students in lower grade levels (Lewis et al., 2021; The World Bank, UNESCO, & UNICEF, 2021). Thus, estimates in this report, *Report 3*, confirm and expand our general understanding of the relationship between modality and student achievement.

Findings may not be generalizable outside of pandemic contexts. As we showed in *Report 2* of this series, students statewide performed lower in SY21 compared to SY19, even in schools that spent most of the year instructing in-person. This suggests that other factors related to the pandemic contributed to test score declines. The relationship between in-person learning and student outcomes may differ outside of this context.

Discussion and Implications

While the COVID-19 pandemic impacted schooling for all students in Illinois, in the year following the pandemic onset (SY21) there was dramatic variation across students in the amount of remote versus in-person instruction they received. In this report, we showed that this variation followed distinct patterns. Students in lower grade levels (grades 5 and 6) spent slightly more time in-person than students in higher grade levels (grades 7 and 8). This finding accords with ISBE's recommendation to districts and schools to prioritize in-person learning for younger students. Much more predictive of modality than grade level, however, was a student's race/ethnicity. White students on average spent about half of the school year in-person, while Black and Latinx students spent just 14% and 17% of the year in-person, respectively. English learners also had very low rates of in-person instruction – on par with the Latinx student population–despite ISBE's recommendation that in-person learning should be prioritized for this group. FRPL-eligible students spent 24% of the year in-person, while students with IEPs learned in-person at about the same rate as the state average (34% versus 33%, respectively). These findings echo the results of *Report 1* in this series, which showed that schools serving higher proportions of White students spent more of the year in-person, while schools serving higher proportions of Black, Latinx, EL, and low-income students spent more of the year remote.

We then analyzed the relationship between the amount of time students spent in-person and one indicator of learning–IAR test score outcomes. We showed that attending more of the year in-person was associated with higher scores in ELA and math, controlling for student and school characteristics. We estimated that after a year of learning, a student who attended in-person would score 5.4 points (0.15 SDs) higher in ELA and 7.5 points (0.22 SDs) higher in math than a student who attended the year remotely. These findings align with other studies that have shown a negative relationship between remote learning and achievement (Darling-Aduana et al., 2022; Domina et al., 2022; EPIC, 2021; Goldhaber et al., 2022; Halloran et al., 2021; Kogan & Lavertu, 2021). Our estimates confirm and expand upon our general findings from *Report 2* of the series, which explored the relationship between modality and achievement at the school level.

Given the nature of this study's data and analyses, the magnitude of these findings can be considered small to moderate (Kraft, 2020). For a student starting at the 50th percentile in ELA, after a year of remote learning they would be expected to perform at the

43rd percentile. In math, a student starting at the 50th percentile would be expected to perform at the 42nd percentile. These effect sizes are similar to those found for attending a small class instead of a large one (about 0.20 SDs) (Shin & Young Chung, 2009) or for attending a class with exclusively whole-group instruction rather than with small group differentiated instruction (0.12-0.17 SDs) (Lou et al., 1996).

Moving forward, the findings from this study are potentially informative for policymakers and practitioners implementing learning recovery efforts. Average test scores declined statewide, including in schools that instructed in-person for all or part of SY21 (as shown in *Report 2* of this series). However, students who attended a large proportion of the year remotely experienced greater learning declines, on average. These students were disproportionately students of color, FRPL-eligible students, and English learners. These findings help explain the widened racial/ethnic and socioeconomic achievement disparities observed statewide. Addressing these disparities will likely require targeted learning recovery resources and supports.

Works Cited

Agostinelli, F., Doepke, M., Sorrenti, G., & Zilibotti, F. (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of Public Economics*, *206*, 104574. https://doi.org/10.1016/J.JPUBECO.2021.104574

Averett, K. H. (2021). Remote learning, COVID-19, and children with disabilities. *AERA Open*, *7*, 23328584211058470. https://doi.org/10.1177/23328584211058471

Bacher-Hicks, A., & Goodman, J. (2021, July 13). The COVID-19 pandemic is a lousy natural experiment for studying the effects of online learning. *Education Next*. https://www.educationnext.org/covid-19-pandemic-lousy-natural-experiment-for-studying-the-effects-online-learning/

Bartlett, L. (2022). Specifying hybrid models of teachers' work during COVID-19. *Educational Researcher*, *51*(2). https://doi.org/10.3102/0013189X211069399

Becker, S. P., Breaux, R., Cusick, C. N., Dvorsky, M. R., Marsh, N. P., Sciberras, E., & Langberg, J. M. (2020). Remote learning during Covid-19: Examining school practices, service continuation, and difficulties for adolescents with and without Attention-Deficit/Hyperactivity Disorder. *Journal of Adolescent Health*, *67*(6), 769–777. https://doi.org/10.1016/j.jadohealth.2020.09.002

Darling-Aduana, J., Woodyard, H. T., Sass, T. R., & Barry, S. S. (2022). *Learning-mode choice, student engagement, and achievement growth during the COVID-19 pandemic*. National Center for Analysis of Longitudinal Data in Education Research (CALDER). https://caldercenter.org/sites/default/files/CALDER%20WP%20260-0122.pdf

Domina, T., Hashim, A., Kearney, C., Pham, L., & Smith, C. (2022). *COVID-19 and the system resilience of public education: A view from North Carolina*. Urban Institute.

https://www.urban.org/research/publication/covid-19-and-system-resilience-public-education-view-north-carolina

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2021). *Mind the gap: COVID-19 is widening racial disparities in learning, so students need help and a chance to catch up*. McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/covid-19-and-learning-loss-disparities-grow-and-students-need-help

Duckworth, A. L., Kautz, T., Defnet, A., Satlof-Bedrick, E., Talamas, S., Lira, B., & Steinberg, L. (2021). Students attending school remotely suffer socially, emotionally, and academically. *Educational Researcher, 50*(7), 479–482. https://doi.org/10.3102/0013189X211031551

Enders, O. G., & Kostewicz, D. (2022). Secondary teachers' remote instructional practices in mathematics for students with disabilities. *Journal of Special Education Technology*, 01626434211059486. https://doi.org/10.1177/01626434211059486

EPIC. (2021). *K-8 student achievement and achievement gaps on Michigan's 2020-21 benchmark and summative assessments*. Education Policy Innovation Collaborative. https://epicedpolicy.org/k-8-student-achievement-and-achievement-gaps-on-michigans-2020-21-benchmark-and-summative-assessments/

Godoy, M. (2022, March 14). Better air in classrooms matters beyond COVID. Here's why schools aren't there yet. *NPR*. https://www.npr.org/sections/health-shots/2022/03/14/1086125626/school-air-quality

Goldhaber, D., Kane, T., Mceachin, A., Morton, E., Patterson, T., & Staiger, D. (2022). *The consequences of remote and hybrid instruction during the pandemic*. National Center for Analysis of Longitudinal Data in Education Research.

https://caldercenter.org/publications/consequences-remote-and-hybrid-instruction-during-pandemic

Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour*, *5*(10), 1273–1281. https://doi.org/10.1038/s41562-021-01212-7

Greenland, S., Pearl, J., & Robins, J. M. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*(1), 29–46. https://doi.org/10.1214/ss/1009211805

Haderlein, S. K., Saavedra, A. R., Polikoff, M. S., Silver, D., Rapaport, A., & Garland, M. (2021). Disparities in educational access in the time of COVID: Evidence from a nationally representative panel of American families. *AERA Open*, *7*. https://doi.org/10.1177/23328584211041350

Halloran, C., Jack, R., Okun, J. C., Oster, E., & Mit, J. O. (2021). *Pandemic schooling mode and student test scores: Evidence from US states*. NBER Working Papers. https://www.nber.org/papers/w29497

Hamilton, L. S., Kaufman, J. H., & Diliberti, M. K. (2020). *Teaching and leading through a pandemic: Key findings from the American Educator Panels Spring 2020 Covid-19 surveys*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA168-2.html

Hanno, E. C., Fritz, L. S., Jones, S. M., & Lesaux, N. K. (2022). School learning format and children's behavioral health during the COVID-19 pandemic. *JAMA Pediatrics*, *176*(4), 410–411. https://doi.org/10.1001/jamapediatrics.2021.5698

Hoffman, J. A., & Miller, E. A. (2020). Addressing the consequences of school closure due to COVID-19 on children's physical and mental well-being. *World Medical & Health Policy*, *12*(3), 300–310. https://doi.org/10.1002/wmh3.365

ISBE. (n.d.). *Evidence-Based Funding*. Illinois State Board of Education. Retrieved March 18, 2022, from https://www.isbe.net

ISBE. (2020). *Fall 2020 learning recommendations*. https://www.isbe.net/Documents/Fall-2020-Learning-Rec.pdf

ISBE. (2022). *5Essentials Survey*. Illinois State Board of Education. https://www.isbe.net

Jack, R., Halloran, C., Okun, J., & Oster, E. (2022). *Pandemic schooling mode and student test scores: Evidence from U.S. school districts*. NBER Working Papers. https://emilyoster.net/wp-content/uploads/MS_Updated_Revised.pdf

Kamenetz, A. (2020a, May 27). Survey shows big remote learning gaps for low-income and special needs children. *NPR*. https://www.npr.org/sections/coronavirus-live-updates/2020/05/27/862705225/survey-shows-big-remote-learning-gaps-for-low-income-and-special-needs-children

Kamenetz, A. (2020b, July 23). Families of children with special needs are suing in several states. Here's why. *NPR*. https://www.npr.org/2020/07/23/893450709/families-of-children-with-special-needs-are-suing-in-several-states-heres-why

Kaufman, J. H., & Diliberti, M. K. (2021). *Divergent and inequitable teaching and learning pathways during (and perhaps beyond) the pandemic: Key findings from the American Educator Panels. Spring 2021 COVID-19 surveys*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA168-6.html

Kogan, V., & Lavertu, S. (2021). *How the COVID-19 pandemic affected student learning in Ohio: Analysis of Spring 2021 Ohio state tests*. John Glenn College of Public Affairs, The Ohio State University. https://glenn.osu.edu/how-covid-19-pandemic-affected-student-learning-ohio

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lambert, R., & Schuck, R. (2021). “The wall now between us”: Teaching math to students with disabilities during the COVID spring of 2020. *The Asia-Pacific Education Researcher*, *30*(3), 289–298. https://doi.org/10.1007/s40299-021-00568-8

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

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d’Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423–458. https://doi.org/10.3102/00346543066004423

Marshall, D. T., Shannon, D. M., & Love, S. M. (2020). How teachers experienced the COVID-19 transition to remote instruction. *Phi Delta Kappan*, *102*(3), 46–50. https://doi.org/10.1177/0031721720970702

Morita-Mullaney, T., Greene, M., Cushing-Leubner, J., Benegas, M., & Stolpestad, A. (2021). COVID-19 and pandemic teaching: Emergency remote teaching and learning and English learner teachers. *INTESOL Journal*, *18*(1), 24–50. https://doi.org/10.18060/25273

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, *48*(1), 163–193. https://doi.org/10.3102/0002831210362589

Peterson, L., Pham, S., & Sole, M. (2021). Supporting English Learners during Virtual Learning. *Communique*, *49*(6), 16–17.

Rapaport, A., Saavedra, A., Silver, D., & Polikoff, M. (2020). *Surveys show things are better for students than they were in the spring—Or do they?* Brookings Institution. https://www.brookings.edu/blog/brown-center-chalkboard/2020/11/18/surveys-show-things-are-better-for-students-than-they-were-in-the-spring-or-do-they/

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd edition). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/hierarchical-linear-models/book9230

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, *6*(1), 253–272. https://doi.org/10.1146/annurev-economics-071813-104217

Shin, I.-S., & Young Chung, J. (2009). Class size and student achievement in the United States: A meta-analysis. *KEDI Journal of Educational Policy*, *6*(2), 3–19.

Steed, E. A., Phan, N., Leech, N., & Charlifue-Smith, R. (2021). Remote delivery of services for young children with disabilities during the early stages of the Covid-19 pandemic in the United States. *Journal of Early Intervention*, 10538151211037672. https://doi.org/10.1177/10538151211037673

Stelitano, L., Mulhern, C., Feistel, K., & Gomez-Bendaña, H. (2021). *How are teachers educating students with disabilities during the pandemic?* RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1121-1.html

Sugarman, J., & Lazarin, M. (2020). *Educating English learners during the COVID-19 pandemic: Policy ideas for states and school districts*. Migration Policy Institute.

https://www.migrationpolicy.org/sites/default/files/publications/mpi-english-learners-covid-19-final.pdf

The World Bank, UNESCO, & UNICEF. (2021). *The state of the global education crisis: A path to recovery*. The World Bank, UNESCO, & UNICEF. https://www.unicef.org/reports/state-global-education-crisis

Trinidad, J. E. (2020). Equity, engagement, and health: School organisational issues and priorities during COVID-19. *Interdisciplinary Studies Faculty Publications*. https://archium.ateneo.edu/is-faculty-pubs/29

Valicenti-McDermott, M., O'Neil, M., Morales-Lara, A., Seijo, R., Fried, T., & Shulman, L. (2022). Remote learning experience for children with developmental disabilities during COVID-19 pandemic in an ethnically diverse community. *Journal of Child Neurology*, *37*(1), 50–55. https://doi.org/10.1177/08830738211040296

Werner, S. M. (2022). *Trauma-informed care for supporting the mental and physical health of youth experiencing homelessness (YEH)*. Illinois Workforce and Education Research Collaborative (IWERC), Discovery Partners Institute, University of Illinois. https://2hih5c1nlnt71of6g8yfvvnl-wpengine.netdna-ssl.com/wp-content/uploads/2022/03/IWERC_TIC_Youth-Experiencing-Homelessness_Werner2022.pdf

Appendix A

Experiences of Remote Learning Across Student Subgroups

While most research has focused on average student outcomes in response to instructional modality, there is reason to think that remote learning may have been particularly deleterious for some student subgroups. Students of lower socioeconomic status and Black and Latinx students experienced lower access to computers and Internet (Goudeau et al., 2021; Haderlein et al., 2021; Peterson et al., 2021; Sugarman & Lazarin, 2020; Trinidad, 2020), and they were less likely to meet with their teachers (Haderlein et al., 2021) when instructed remotely at the start of the pandemic. Low-income students in remote learning experienced lower positive peer influence than those who learned in-person, and they were less likely than higher-income students to have a parent who was able to work from home and provide instructional support during the school day (Agostinelli et al., 2022). Low-income parents of students receiving remote instruction also have expressed lower confidence in supporting their children with their schoolwork (Goudeau et al., 2021; Haderlein et al., 2021). It is therefore unsurprising that low-income students have been documented as engaging less in school than high-income students during remote instruction during the pandemic (Dorn et al., 2021). Nationally, studies have found that the negative impact of remote instruction on learning was twice as large for students in high-poverty schools than those in low-poverty schools, and for schools with larger shares of Black and Latinx students (Darling-Aduana et al., 2022; Goldhaber et al., 2022; Halloran et al., 2021; Jack et al., 2022).

Additional concerns exist about the impacts of remote learning on students with disabilities (SWD) and English learners (ELs). Many SWD struggled to adapt to the lack of structure and in-person feedback when learning remotely (Averett, 2021; Becker et al., 2020; Lambert & Schuck, 2021). For ELs, whose language development is shaped through daily peer interactions and school activities, the transition to remote learning drastically reduced opportunities for communicating in English (Peterson et al., 2021). Teachers experienced difficulties providing special education services and support for ELs remotely (Enders & Kostewicz, 2022; Marshall et al., 2020; Stelitano et al., 2021), and many teachers did not feel that they received sufficient recommendations from their schools or districts for these implementing challenging tasks (Hamilton et al., 2020). Many students received reduced special education services or no services at all (Averett, 2021; Becker et al., 2020; Kamenetz, 2020a, 2020b; Steed et al., 2021; Valicenti-McDermott et al., 2022). Likewise, many ELs did not receive language support programming (Morita-Mullaney et al., 2021). In turn, SWD participated in school and completed assignments less frequently in remote learning (Enders & Kostewicz, 2022; Valicenti-McDermott et al., 2022), while ELs have potentially experienced both academic and language delays (Sugarman & Lazarin, 2020). No studies to date have examined differential impacts of remote learning on these subgroups.

Students experiencing homelessness were among those most vulnerable to the cessation of in-person instruction. Schools not only provide students experiencing homelessness with myriad physical and mental health resources (including guaranteed meals, counseling, and physical safety), but also connect them with a range of social services (Hoffman & Miller, 2020; Werner, 2022). The transition to remote instruction disrupted access to these resources critical for learning and wellbeing. We also note that students experiencing homelessness may have been undercounted in Illinois during the SY21, making it difficult to collect data on their academic and other outcomes (Werner, 2022).

Appendix B

Table B1

Descriptive Student Characteristics (2021), Samples Excluding CPS and Presumed Fall Testers

		Mean	
Variable type	Variable	Complete Cases Excluding CPS (n = 301,842)	Complete Cases Excluding Presumed Fall Testers (n=341,371)
Instructional Modality (%)	In-Person Instruction	48.7	43.8
	Remote Instruction	46.0	50.9
	Absent	5.3	5.3
Outcomes	ELA 2021	741.2	740.2
(SY21 IAR scale scores)	Math 2021	727.9	726.6
Prior Achievement	ELA 2019	734.2	733.0
(SY19 IAR scale scores)	Math 2019	738.5	737.7
Student Covariates (%)	Race/Ethnicity		
	White	62.9	56.9
	Hispanic/Latinx	18.5	22.0
	Black	9.2	12.1
	Asian	4.9	4.8
	Two or more races	4.3	3.9
	American Indian/Alaska Native	0.2	0.2
	Native Hawaiian/Pacific Islander	0.1	0.1
	Other Demographics		
	Female	48.6	48.6
	FRPL	36.3	40.8
	IEP	12.9	13.1
	English learner	7.0	8.4
	Homeless	1.0	0.9

Note: We analyzed cases excluding CPS to check how much estimates from this single large district influence the size of our overall estimates. “Presumed fall testers” refers to schools that we assume tested in fall rather than spring of 2021 based on their reporting to ISBE. We analyze cases excluding presumed fall testers to check how sensitive our findings are to the influence of students who experienced additional months of learning.

Table B2

Estimated associations between percentage of SY21 in-person and IAR scale scores, grades 5-8, sample excluding presumed fall testers

	Scale Score	Effect Size
	.054***	.0015
IAR ELA	(.004)	(.0001)
	.075***	.0022
IAR Math	(.003)	(.0001)
N Students	341,371	
N Schools	2,479	

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05
Note: Effect sizes are Cohen's *d*. Controls included the following student characteristics: ELA/math IAR score in 2019, race/ethnicity, gender, eligibility for Free/Reduced-Price Lunch (FRPL), English learner (EL) status, Individualized Education Program (IEP) participation, Homelessness; and 2019 school characteristics: % White, % Black, % Hispanic/Latinx, % Two or more races, % Low-income, % Homeless, % ELs, % IEP, mean attendance rate, indicator for Evidence-based Funding Tier 4, and scores on the *5Essentials* measure for Involved Families.

Table B3

Estimated associations between percentage of SY21 in-person and IAR scale scores, grades 5-8, sample excluding students in Chicago Public Schools

	Scale Score	Effect Size
	.062***	.0017
IAR ELA	(.004)	(.0000)
	.083***	.0024
IAR Math	(.004)	(.0001)
N Students	301,842	
N Schools	2,118	

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05
Note: Effect sizes are Cohen's *d*. Controls included the following student characteristics: ELA/math IAR score in 2019, race/ethnicity, gender, eligibility for Free/Reduced-Price Lunch (FRPL), English learner (EL) status, Individualized Education Program (IEP) participation, Homelessness; and 2019 school characteristics: % White, % Black, % Hispanic/Latinx, % Two or more races, % Low-income, % Homeless, % ELs, % IEP, mean attendance rate, indicator for Evidence-based Funding Tier 4, and scores on the *5Essentials* measure for Involved Families.

Appendix C

Two primary approaches for analyzing hierarchically structured data are the use of Hierarchical Linear Models (HLM) and fixed effects. As described above, HLM allows us to understand how the relationship between instructional modality and test scores varies depending on the school that a student attends. Accounting for the hierarchical data structure is also important for ensuring that our estimates are not biased due to the similarity of students within schools compared to students between schools. These models help us separate the proportion of variation in test scores that occurs between schools from that occurring between students within schools.

Fixed effects models are also designed to reduce bias resulting from differences between higher-level units, such as schools or districts. However, rather than modeling the variation at higher levels, fixed effects models control it away. This means that they are able to control for all higher-level confounders–both observed and unobserved. As a result, fixed effects models estimate effects within, but not between, higher-level units such as schools or districts. These models are useful for creating unbiased estimates within higher-level units.

To assess how results of our models might change when controlling for district confounders, we reran analyses with district fixed effects. Our models for these analyses are as follows:

Level-1 Model.

$$Y_{ij} = \pi_{0j} + \pi_{1j}(perInp_{ij}) + \pi_{2j}(2019score_{ij}) + \boldsymbol{\Omega X}_{ij} + e_{ij}$$

Level-2 Model.

$$\pi_{0j} = \beta_{01} + \boldsymbol{\Omega D}_j + \beta_{02}(AttendRate_j) + \beta_{03}(Enrollment_j) + \beta_{04}(EBFTier4_j) + \beta_{05}(5InvolvedFams_j) + \beta_{06}(DistrictFixedEffects_j)\ r_{0j}$$

$$\pi_{1j} = \beta_{11} + r_{0j}$$

$$\pi_{2j} = \beta_{21} + r_{0j}$$

$$\pi_{3j} = \beta_{31}$$

Where interpretation of each term is the same as those presented in the main analysis, except that β_{06} represents school *j*'s intercept relative to the intercept of its district. We present the results of models using district fixed effects in Table C1.

As shown, estimates are slightly larger than those of models without fixed effects, by approximately 0.006 points in ELA and by .003 points in math. These estimates represent the relationship between in-person learning and achievement among students within similar districts. Because modality policy decisions were often made at the district level, students across similar districts would have had similar modality options, such as returning to in-person learning in the spring. Estimates accounting for district fixed effects were larger potentially because they represent the relationship between modality and achievement when driven by the choices of students and families. In other words, given the same set of modality options, students who spent more time learning in-person scored even higher relative to students whose districts gave them the option to learn in person, but stayed remote. Overall, however, results from our models with district fixed effects are very similar to those presented in our main analyses, suggesting that unobserved district characteristics are not creating significant bias in our estimates.

A drawback of district fixed effects models is that they do not allow us to understand the variation *between* districts. These models, in turn, control away much of the variation statewide in modality. Other types of fixed effects, such as student or school fixed effects, present similar issues; namely, they assume that the relationship between modality and achievement is constant across students and schools. For these reasons, we present the results of HLMs that nest students within schools, not controlling for fixed effects, as our main analyses.

Table C1.

Estimated associations between percentage of SY21 in-person and IAR scale scores, grades 5-8, including district-level fixed effects

	ELA	Math
SY21 Scale Score	0.060***	0.078***
	(0.004)	(0.003)
Constant	218.38	240.15
	(13.24)	(12.05)
Random Effect Parameters (Variance)		
In-Person Instruction	0.006	0.006
	(0.001)	(0.001)
SY19 Scores	0.000	0.000
	(0.00)	(0.00)
School (constant)	0.000	0.000
	(0.00)	(0.00)
Residual	469.43	394.47
	(1.13)	(0.95)
N Schools	2,600	
N Students	346,596	

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Note: Controls included the following student characteristics: ELA/math IAR score in 2019, race/ethnicity, gender, eligibility for Free/Reduced-Price Lunch (FRPL), English learner (EL) status, Individualized Education Program (IEP) participation, Homelessness; and 2019 school characteristics: % White, % Black, % Hispanic/Latinx, % Two or more races, % Low-income, % Homeless, % ELs, % IEP, mean attendance rate, indicator for Evidence-based Funding Tier 4, and scores on the *5Essentials* measure for Involved Families.