Differences in 2019–2022 COVID-Related NAEP Urban District Score Declines in Grade 4 Based on County-Level Health and Socioeconomic Data

AIR-NAEP Working Paper 2024-02

Lisa M. Yarnell, Thomas D. Snyder, George W. Bohrnstedt, Markus Broer

November 2024

The research contained in this working paper was commissioned by the National Center for Education Statistics (NCES). It was conducted by the American Institutes for Research (AIR) in the previous framework of the Education Statistics Services Institute Network (ESSIN) Task Order 14: Assessment Division Support (Contract No. ED-IES-12-D-0002/0004) which supported NCES with expert advice and technical assistance on issues related to the National Assessment of Educational Progress (NAEP). The following working paper was completed under a follow-on contract (#91990022C0053). It is disseminated under the current NAEP Technical Support contract (#91990023D0006/91990023F0350). AIR is responsible for any error that this report may contain. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Advancing Evidence. Improving Lives.

In Memoriam – Lisa Yarnell

Lisa Yarnell passed away unexpectedly in December 2024.

If you knew Lisa, you undoubtedly felt, read, and heard about her love for applied statistics and psychometrics. Her intellectual curiosity was evident in the way she approached statistical challenges. On email threads, you could witness her working through challenges, thinking aloud as she typed—questioning assumptions, exploring model choices, and evaluating their appropriateness. Her curiosity was inspiring, sparking enthusiastic statistical discussions with colleagues via Teams, in meetings, and as an engaged member of AIR's Methodological Communities of Practice.

Lisa joined AIR in 2014 after earning her PhD in Educational Psychology and completing postdoctoral work at the University of Texas at Austin. Her academic focus was on human development and applied statistical modeling. Her dissertation, titled *The Nature of Socioeconomic Status Among Young Adults, and Its Effect on Health: A Multi-Group SEM Analysis by Gender and Race/Ethnicity*, laid the foundation for much of her work at AIR.

Lisa played a pivotal role in the technical review and research studies for the National Assessment of Educational Progress (NAEP), the only ongoing measure of what U.S. students know and can do. When she was part of the technical review team, we knew the product would meet the rigorous standards expected from a statistical agency. Lisa also contributed to several research teams, where she tackled the methodological challenge of creating a set of weights for teacher-centric reporting for both NAEP and the Trends in International Mathematics and Science Study. This innovative methodology advanced research possibilities and provided valuable insights to inform policy discussions. Lisa also contributed to the field of research exploring the racial match between students and teachers and its relationship with student performance.

Her most recent contribution—the research underlying this working paper—tied together her love of statistics, her deep knowledge of NAEP, and her PhD-level expertise in health and socioeconomic status. The working paper builds on several discussions, including the merging of multiple datasets to increase their utility, the connection between health and academic performance, and the role of socioeconomic status in relation to quality-of-life indicators for our students.

This working paper is dedicated to the memory of our dear colleague, Lisa Yarnell.

AIR

Established in 1946, with headquarters in Arlington, VA, the American Institutes for Research (AIR) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of health, education, and workforce productivity. For more information, visit www.air.org.

NCES

The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education in the United States and other nations. NCES is located within the U.S. Department of Education and the Institute of Education Sciences. NCES fulfills a congressional mandate to collect, collate, analyze, and report complete statistics on the condition of American education; conduct and publish reports; and review and report on education activities internationally.

NCES support contracts

The analysis for this working paper was completed under the previous Education Statistics Services Institute Network (ESSIN) Task Order 14: Assessment Division Support (Contract No. ED-IES-12-D-0002/0004), and the working paper was completed under a follow on contract (#91990022C0053). It is disseminated under the current NAEP Technical Support contract (#91990023D0006/91990023F0350).

William Tirre, a program director in the NCES Assessment Division, oversaw the Research, Analysis, and Psychometric Support subcomponent of the previous ESSIN Task Order 14 (Contract No. ED-IES-12-D-0002/0004) and current NAEP Support contract #91990022C0053. The Contracting Officer's Representative of the current NAEP Technical Support contract (#91990023D0006/91990023F0350) is Grady Wilburn.

Suggested citation:

Yarnell, L., Snyder, T., Bohrnstedt, G., and Broer, M. (2024). *Differences in 2019–2022 COVID-Related NAEP Urban District Score Declines in Grade 4 Based on County-Level Health and Socioeconomic Data.* [AIR-NAEP Working Paper # 2024-02]. Washington, DC: American Institutes for Research.

For inquiries, contact:

Markus Broer, Project Director for Research under NAEP Support Contracts: mbroer@air.org

Mary Ann Fox, AIR Vice President and Project Director of NAEP Support Contracts: mafox@air.org

Executive Summary

The COVID-19 pandemic resulted in many schools making difficult decisions about remote instruction, masking, and social distancing policies, as well as how and when to transition students back to classroom learning. In light of these challenges, the 2022 state National Assessment of Educational Progress (NAEP) assessments revealed declines for 4th- and 8th-grade reading and mathematics assessment scores compared to the pre-pandemic levels of 2019. These declines were reflected by significant decreases in 9 of the 26 NAEP Trial Urban District Assessments (TUDAs). This memo explores the relationship of classes of health-related indicators and changes in grade 4 reading and mathematics scores at a time of heightened health and social risk associated with the COVID-19 pandemic.

For this study, NAEP data were merged with data from the 2022 County Health Rankings & Roadmap, provided by the University of Wisconsin, to obtain relevant indicators of community health. Based on health data for the counties in which the TUDA districts were located, a latent class analysis (LCA) model was developed that divided the schools into 3 classes, or groups, based on their level of community health. For example, schools in class 1 were the "healthiest" in that their counties had lower rates of infant mortality, child mortality, and low-weight births; fewer years lost due to premature death; and lower rates of frequent mental distress among adults relative to the other classes of schools. Furthermore, schools in class 1 were in counties with the longest life expectancies, the highest median incomes, and the lowest rates of firearm fatalities. In contrast, class 3 schools were located in counties that were notably worse according to the totality of these indicators, and schools in class 2 fell between those in classes 1 and 3 on these indicators.

The LCA model results suggested that TUDA schools in class 1 (i.e., those located in "healthier" surrounding communities with higher median household incomes) had higher grade 4 reading and mathematics scores in 2019 and 2022 compared to schools in less healthy and lower-SES counties (classes 2 and 3). These schools' scores also fell less during the height of the pandemic compared to others. On the other hand, TUDA schools with less healthy community indicators and lower median household incomes not only performed relatively worse on the grade 4 assessments in both years, but also experienced a greater decline in scores from 2019 to 2022. A student-level analysis of the data provided similar results and also revealed a counterintuitive finding: there were relatively large increases in student absences in schools in the more healthy than in the less healthy communities; however, these higher rates of absence did not translate into drops in assessment scores as large as those seen among schools in less healthy communities.

While the analyses presented here are exploratory, it would be beneficial to explore the availability of health-related data on air pollution and other health factors more aligned to the specific school location, rather than relying on health data available at the county level. This study highlights the potential value of integrating NAEP with other datasets to inform policies aimed at improving student academic performance.

Contents

List of Tables

Background

In wake of the disruptions caused by COVID-19, there is interest in ascertaining the changes in school performance that have occurred, given the challenges it presented—both health- and learning-related—to students, families, teachers, and schools. The pandemic resulted in many schools conducting instruction remotely between 2020 and 2022, whether prepared to do so or not, and making difficult decisions about masking and social distancing policies, as well as how and when to transition students back to classroom learning. Likewise, families were challenged by their various social, health, and economic needs, which may have impacted how much they were able to support students' learning while schools were closed. In addition, many students were challenged by a sudden shift to learning at a distance from their peers and teachers. These challenges were further complicated by debates about optimal solutions for learning under these conditions. Understanding why some schools fared relatively better despite these challenges is of interest to education practitioners, policymakers, and families alike.

Given that the National Assessment of Educational Progress (NAEP) represents our nation's best tool for measuring what our students know and can do across time and location and the availability of NAEP data collected both before and after the height of the pandemic, it represents an optimal data source for studying changes in academic performance associated with the disruptions associated with the pandemic. Indeed, the results of the 2022 state NAEP assessments revealed statistically significant score declines among students in the United States for both reading and mathematics at both grades 4 and grade 8, compared to the prior state NAEP assessments conducted in the spring of 2019. In fact, the declines reduced the average scores for each of these assessments to the lowest levels seen in over a decade.

However, many large school districts exhibited patterns that were not consistent with these national trends. For instance, among the 26 large urban school districts that participated in both the 2019 and 2022 NAEP Trial Urban District Assessments (TUDAs), 17 did not have a measurable score change on the grade 4 reading assessment between the two years, while the remaining 9 had declines in their average scores, some substantially: Atlanta's public schools fell 8 points in grade 4 reading, Charlotte-Mecklenberg's fell by 10 points, and Cleveland's fell by 15 points. Similarly, all but three districts experienced a statistically significant decline in their average score on the grade 4 mathematics assessment. Looking across TUDA districts, score declines ranged from nonsignificant changes of 1 point for Hillsborough County (FL)'s public schools and 3 points for Dallas's to declines of 15 points in public schools in Cleveland and Baltimore City.[1]

Health-related aspects of the school districts may be particularly relevant to understanding the variations in these changes at a time of heightened health and social risk, as was induced by the pandemic. The robust associations between community health and children's school readiness and achievement have already been established in the research literature (see Leventhal and Brooks-Gunn [2000] for a review). Defined broadly, health as defined by the World Health Organization in its constitution refers to "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."[2] Within a school district, community health can encompass various neighborhood features reflecting access to education, health care, information, and social resources that help ensure physical, mental, and social well-being. Many of these features may be strongly correlated with socioeconomic status (SES) yet not accounted for by its traditional indicators and unique in their impact on student achievement, especially in the context of COVID-19.

For instance, limited access to healthy foods can impact learning by impeding brain development (Georgieff, Ramel, and Cusick 2018; Tanner and Finn-Stevenson 2002), and lower-SES children are more likely to experience iron deficiency (Andriastuti, Ilmana, Nawangwulan, & Kosasih, 2020). Likewise, children from lower-SES families have a higher incidence of lead in the blood, often due to lead-containing paint in older homes, which can result in poorer cognition and attention issues in affected children (Marshall, Betts, Kan, McConnell, Lanphear, & Sowell (2020). These examples highlight how various factors in the residential environment, including housing conditions and relative access to healthy foods, can impact student achievement. This appears to be particularly significant when large numbers of children were attending school remotely and spending greater amounts of time at home, thus lacking their usual access to school resources, such as school lunch programs and a potentially healthier learning environment.

Similarly, both the overcrowding in urban environments and limited access to health care in rural environments had the potential to impact vulnerability to the COVID-19 virus among students and their families (OECD, 2023). Students may have also been affected more strongly by the physical and mental health conditions of their caregivers during this time of residential isolation, depending on their caregivers' own preexisting health conditions, their healthy or

---

[1] The 2022 state NAEP results for grade 4 reading are available at https://www.nationsreportcard.gov/reading/districts/scores/?grade=4. Those for grade 4 mathematics are available at https://www.nationsreportcard.gov/mathematics/districts/scores/?grae=4.

[2] https://www.who.int/about/governance/constitution#:~:text=Health%20is%20a%20state%20of,belief%2C%20economic%20or%20social%20condition

unhealthy coping mechanisms at a time of heightened stress (e.g., excessive drinking or smoking), and families' access to physical and mental health resources.

Thus, by linking data on pandemic-related aspects of community health and socioeconomic resources with results from NAEP, it may be possible to shed light on why some schools performed better than others during this time and what community health and socioeconomic features are associated with various degrees of performance change.

To examine this issue, we merged the 2019 and 2022 NAEP grade 4 reading and mathematics data collected from the TUDA districts with county-level heath data available for this same time period. More specifically, we examined changes in average grade 4 reading and mathematics performance between 2019 (prior to the COVID pandemic) and 2022 (largely post-pandemic) among the set of TUDA schools that participated in both the 2019 and 2022 NAEP assessments (i.e., the overlap sample of schools) and the association of those changes with community health characteristics.

Our use of data for the 26 TUDA districts that participated in both the 2019 and 2022 assessments allowed us to explore a variety of community characteristics, resources, and challenges in relation to performance changes during this period. TUDA schools are sampled for NAEP at a higher rate than schools in the rest of the state in which they are located to enable more accurate district reporting. As a result, a common panel of schools can be readily established, enabling an analysis of change over time with a design that controls for many location-based covariates. In the work presented here, we examine which subsets of these schools performed relatively better, and which performed relatively worse, in 2022 compared to 2019. Our use of grade 4 data may be particularly fruitful for this analysis given that younger students are less skilled in independent learning relative to the older students who participate in NAEP at grades 8 or 12; younger students may also be more sensitive to community and familial health risks given their developmental vulnerability. For instance, they may have been more affected by their parents' physical and mental well-being during the pandemic relative to older children, given their greater dependence, thereby potentially leading to stronger associations between community health and social resources and changes in their performance on NAEP.

This study explores how external health data can be analyzed to contextualize school-level differences in NAEP performance change during the period when COVID peaked, an approach that highlights the value of linking NAEP with other datasets. Additionally, we examine how the health profiles of schools' communities relate to changes in rates of chronic absenteeism, as reported by students in responses to the NAEP background questionnaire, given the strong, lasting toll that COVID-19 appears to have taken on student attendance, post-pandemic (Vázquez Toness 2023). While the analyses presented here are exploratory, our results appear to reveal useful information about which clusters of schools performed relatively better and which performed worse during the pandemic—information that is useful for contextualizing NAEP performance.

Methods

Data

The NAEP data were drawn from the 2019 and 2022 NAEP grade 4 reading and mathematics assessments, limited to the set of TUDA schools that were sampled and participated in both years for each of the two subjects (approximately 700 schools). Note that while most schools participated in both subjects, some schools may have students who were assessed in one subject only (for example, this may be the case for a school with lower student enrollment where the entirety of its students are excluded, are absent, or opt out on the day of the assessment). For each dataset, the student data were limited to the reporting sample—the set of students who participated in the assessment and therefore contributed to school-average performance scores. This involved excluding students who were not assessed yet whose other demographic characteristics, such as their gender, race/ethnicity, and eligibility for the National School Lunch Program [NSLP], may have been known. To conduct analyses at the school level, we used a Stata macro program to produce two sets of school-average performance values for these two sets of TUDA schools: one set for reading and a second set for mathematics.

We then merged the NAEP data for each subject with data from the 2022 County Health Rankings & Roadmap to obtain relevant indicators of community health. The County Health Rankings data are provided in a series of annual datasets provided by the University of Wisconsin that contain numerous measures of county-level health and socioeconomic features for every county in the United States, including measures of child health risk factors, adult health and health behaviors, mortality rates, health care access, environmental health factors, economic factors, crime/safety, and county population composition, as well as rates of COVID deaths in 2020.[3] We merged the County Health Rankings data with NAEP by converting the zip codes of NAEP schools into county Federal Information Processing Standards (FIPS) codes, which are used as county identifiers in the County Health Rankings data. This conversion was facilitated by a zip-to-county FIPS crosswalk provided by the U.S. Department of Housing and

[3] The County Health Rankings data are available at https://www.countyhealthrankings.org/.

Urban Development for the fall 2018 quarter, which corresponds to the time of the 2019 NAEP school sampling.[4,5]

Some NAEP schools, while possessing distinct zip codes, are located in the same county. In these cases, multiple schools have the same county health data (although student performance may differ). The opposite situation also occurs, where a single school zip code maps onto multiple counties, which may provide more efficient postal mailing in some locations. In this case, it is unclear which county's health and socioeconomic values to assign to the school. For the approximately 200 schools in each dataset whose 5-digit zip codes mapped to multiple counties in the County Health Rankings data, we obtained more specific 9-digit zip codes and precise county locations from the 2018-19 Education Demographic and Geographic Estimates (EDGE) file corresponding with the 2018 Common Core of Data (CCD), which allowed us to decide which was the appropriate health data to link to these schools.[6]

Table 1 shows the distribution of school locations by TUDA district for each overlap sample. For both subjects, the TUDA with the greatest representation was the District of Columbia, reflecting its high rate of school sampling for NAEP, implemented to ensure adequate student sample sizes.

---

[4] The HUD crosswalks are available at https://www.huduser.gov/portal/datasets/usps_crosswalk.html.
[5] We also explored the use of data from the NCES 2021-22 School Pulse Panel (SPP) survey, which provides monthly data for approximately 2,400 U.S. public schools on various COVID-related topics. However, the overlap with the NAEP TUDA samples used here was insufficient for doing so.
[6] The NCES EDGE files are available at https://nces.ed.gov/programs/edge/geographic/schoollocations.

Table 1. School Sample Sizes and Frequencies (Percentages) for the 2019/2022 TUDA Overlap Samples, by District: 2019 and 2022 Grade 4 NAEP Reading and Mathematics Assessments

District	Reading		Mathematics	
	School Sample Size (*n*)	Frequency (%)	School Sample Size (*n*)	Frequency (%)
Atlanta	40	5.01	30	4.88
Chicago	20	2.58	20	2.58
Houston	40	5.59	40	5.60
Los Angeles	10	2.01	10	2.01
New York City	10	1.43	10	1.43
District of Columbia (DCPS)	50	7.59	50	7.60
Boston	40	6.45	50	6.46
Charlotte	20	3.01	20	3.01
Cleveland	50	6.45	50	6.46
San Diego	20	3.44	20	3.44
Austin	30	4.15	20	4.16
Baltimore	20	3.30	20	3.30
Detroit	40	5.44	40	5.45
Jefferson County (KY)	20	3.30	20	3.30
Miami	20	2.72	20	2.73
Milwaukee	30	4.30	30	4.30
Philadelphia	20	2.72	20	2.73
Albuquerque	30	4.30	30	4.30
Dallas	10	1.58	10	1.58
Hillsborough County (FL)	10	1.86	10	1.87
Duval County (FL)	20	3.30	20	3.30
Clark County (NV)	20	3.44	20	3.44
Denver	20	3.01	20	3.01
Fort Worth	30	4.58	30	4.59
Guilford County (NC)	40	5.01	40	5.02
Shelby County (TN)	20	3.44	20	3.44

NOTE: Total school sample size for reading is approximately 700; estimated student population size is approximately 67,900. Total school sample size for mathematics is approximately 700; estimated student population size is approximately 66,700. Sample size is rounded to the nearest ten.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Reading and Mathematics Assessments.

Indicator selection and treatment

We chose community health indicators from the County Health Rankings & Roadmap data for use in modeling, based on their meeting all three of the following criteria:

(1) Having a theoretically or otherwise plausible COVID-related association with changes in NAEP performance with a focus on health and socioeconomic risks, resources, and other characteristics of students, families, and their communities;
(2) Having less than 20 percent missingness for the overlap samples of schools; and
(3) Demonstrating a statistically significant association ($p < .01$) in the theoretically expected direction with school-average change scores between the 2019 and 2022 reading assessments, after accounting for school-average NSLP eligibility among the school's enrolled students in the 2022 NAEP data.

Community health variables demonstrating upper outlier values (z-score > 4.0) were treated by winsorizing these values to the next-highest value within 4 standard deviations (*SD*s) of the mean (Tukey 1962). We also transformed a measure of the county homicide rate by taking its square root, given high kurtosis for the variable in its original distribution. We also divided a variable marking the total number of county deaths due to COVID-19 in 2020 into quartiles for the TUDA school sample used here. This was done given the extremity of upper outliers as well as the high kurtosis in the original distribution for this variable. Finally, we used survey-weighted regression analysis to determine the direction and significance of these potential indicators' associations with the 2019 to 2022 NAEP change scores for grade 4 reading. The regression models were weighted using the 2022 NAEP reading school weights.

Table A-1 in Appendix A displays the majority of variables available in the 2022 County Health Rankings data to detail this process, omitting only variables that

- are redundant in meaning with those that are displayed or less specific in their meaning (e.g., the county rate of informally processed juvenile delinquency cases is omitted, while the rate of formally processed cases is included); or
- are specific to a subpopulation, such as Black, Hispanic, or White persons (which generally have higher rates of missing data and do not capture the experience of all families in the TUDA schools); or
- are measures of county racial/ethnic composition (as these may drive the results and distract from the intended focus on county health and socioeconomic characteristics).

Table A-1's three columns indicate whether a variable met each criterion for retention as a potential indicator. The discussion following this appendix table summarizes the selection process, which resulted in using the indicators presented in Table 2 below.

More specifically, Table 2 provides brief descriptions of the variables that met the indicator selection criteria as detailed in the County Health Rankings 2022 Data Dictionary and the supplemental online documentation of health outcomes. The indicators are organized into three topical groups: child/adolescent health and well-being, adult health, and social and economic factors.

Table 2. Descriptions of Variables Selected for Use as Indicators in Latent Class Analysis: 2022 County Health Rankings & Roadmap Data

Variable/Indicator (original scale)	Description
Child/adolescent health and well-being	
Infant mortality (rate)	Number of infant deaths (within 1 year) per 1,000 live births.
Child mortality (rate)	Number of deaths among residents under age 18 per 100,000 population.
Low-weight births (rate)	Percentage of live births with low birthweight (< 2,500 grams).
Disconnected youth (rate)	Percentage of teens and young adults ages 16-19 who are neither working nor in school.
Adult health	
Premature death (years)	Years of potential life lost before age 75 per 100,000 population (age-adjusted). Also referred to as "years of potential life lost (YPLL)," it is a widely used measure of the rate and distribution of premature mortality. Measuring premature mortality, rather than overall mortality, focuses attention on deaths that might have been prevented. YPLL emphasizes deaths of younger persons, whereas statistics that include all mortality are dominated by deaths of the elderly. For example, in an age-adjusted measure of premature mortality, a death at age 55 counts twice as much as a death at age 65 (because twice as many years of life have been lost compared to a death at age 75), and a death at age 35 counts eight times as much as a death at age 70.
Life expectancy (years)	Average number of years a person can expect to live.
Frequent mental distress (rate)	Percentage of adults reporting 14 or more days of poor mental health per month (age-adjusted; see above).
Social and economic factors	
Median household income (dollars)	The income at which half of the in a county earn more and half of the households earn less.
Firearm fatalities (rate)	Number of deaths due to firearms per 100,000 population.

SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap.

Primary analysis: Latent class analysis

Our primary analytic method for examining the health profiles of the NAEP TUDA schools was latent class analysis (LCA) (Muthén & Muthén 2017), which is a statistical method for identifying latent, unmeasured class membership among subjects or units (such as persons or schools) using observed variables termed “indicators.”[7] The nature of these latent classes (or subpopulations) is not known but may be hypothesized *a priori* such that the indicators are chosen based on theory or previous research. LCA explores “constellations” of factors reflected by the indicators as they naturally cohere in the data in order to characterize the cases into a number of subsets or “classes.” While fit statistics generated by the LCA help decide on the number of latent populations (classes), informed judgment is needed to choose the “best” of the several models concerning the numbers and compositions of the latent classes, each presenting different potential solutions (see, e.g., Lanza, Bray, & Collins [2013] and Bucholz, Heath & Collins [2000] for details and demonstrations).

In the LCA models, the schools analyzed were classified using the “most likely class” method (Clogg 1995). For each subject, we examined the results of models with varying numbers of classes (ranging from 2 to 5) for relative fit and interpretability. The relative fit indices included the Akaike information criterion (AIC) and Bayesian information criterion (BIC), for which smaller relative values indicate better fit (Keith 2006). We also used the Lo-Mendell-Rubin likelihood ratio test (LRT) (Lo, Mendell, & Rubin 2001), using alpha = 0.05, for which a statistically significant value signifies that the model at hand provides better probable fit to the data relative to the model with one fewer class (e.g., rejecting a null 4-class model in favor of a more complex 5-class model). We also examined entropy values, with an entropy value of 0.80 or higher reflecting adequate classification (at least 90% of subjects correctly classified) (Ramaswamy et al. 1993).

Finally, we examined the proportional sizes of each class in each model solution, given that some solutions may suggest classes that are too small to be practically meaningful for consideration by education decisionmakers (e.g., a group of schools representing only 5% of the data). Together, these indices provide information about whether models with greater numbers of classes are useful in light of the downside of additional model complexity, relative to more parsimonious model solutions. After identifying the best potential model based on these criteria for reading and mathematics separately, we examined each solution for substantive interpretability.

While the final LCA models used 2022 NAEP school weights, some methods literature (e.g., Vermunt 2002) suggests that when using LCA with complex survey data, unweighted solutions should also be examined to understand any potential dependency between factors used in the weights and indicators in the model. Therefore, we also examined the results of unweighted LCA models to ascertain the potential existence of additional subpopulations of schools that

[7] Item-level missingness below the 20% threshold was treated in the models via Mplus’s default Full Information Maximum Likelihood (FIML) methods in the LCA models run for each subject (i.e., estimated based on other model covariates) (see Muthén & Muthén 2017).

were weighted downward via the weights. In considering potential solutions, we favored solutions that did not have associated warnings about potential dependencies in the output.

After confirming that the solution indicated by the fit indices also provided a substantively interpretable profile of schools for each class, we analyzed the school-average 2022 NAEP scores and 2019–2022 change scores estimated for each class from the two models (one for reading and one for mathematics). This analysis was performed using Mplus's AUXILIARY commands, meaning that the NAEP change scores were not used as indicators in the model but were examined as auxiliary variables. To assess statistical differences between the latent classes in terms of their 2022 scores and 2019–2022 score changes, we utilized the Mplus DE3STEP command.

Post-hoc analyses

Following the primary LCA models run for each subject, we conducted several analyses to garner additional information about the classes of schools suggested by the results. First, we explored additional characteristics of the classes of schools using data available in NAEP, focusing on two student-level demographic variables (race/ethnicity and NSLP status) as well as absenteeism, which has been shown to be a strong indicator of student achievement.[8] The rationale for their inclusion was to determine whether some latent class or classes had more minority students and absenteeism than others. The effects of these variables were examined using Mplus's AUXILIARY commands along with the performance data, as noted above. Additionally, we examined the distribution of schools in each TUDA district into the latent classes in terms of their proportional membership. Finally, we estimated the average performance as well as the 2019–2022 change scores for the students enrolled in each class of schools to examine whether the same pattern of performance and performance changes were seen at the student level as were suggested in the school-level results.

School-level Analyses Results

Latent class analysis for grade 4 reading

Table 3 shows the relative fit statistics and class proportions for the unweighted and weighted LCA model results for grade 4 reading. Following the criteria described above (see Appendix A for a more detailed explanation), the 3-class solution emerged as the best fit to the data according to both the unweighted and weighted models. These solutions exhibited notably smaller AIC and BIC values than the corresponding 2-class solutions, and their LRTs indicated

[8] See Center for Research in Education and Social Policy (2018) for a summary of research on the effects of absenteeism.

a better fit to the data relative to the more parsimonious 2-class models. While the LRT indicated that the 4-class weighted model offered a better fit to the data than the 3-class weighted solution, it suggested the existence of a fourth class of schools, constituting only 3% of the sample—a proportion too small to be considered practically significant, similar to what was found in the 5-class weighted model. We also did not consider the 4- or 5-class unweighted models to be viable solutions, given estimation warnings in the output even when run with additional random starts and stages of optimization in Mplus. We thus chose the 3-class weighted solution, although future analyses may suggest other designs for improved interpretability.

The very high entropy value of .995 for the weighted 3-class model suggests near-perfect separation of the school data based on the indicators used. This may be logical given that schools are homogeneous in terms of their surrounding community, whereas they are heterogeneous in terms of their internal features. As such, county features may categorize school-level performance data quite well.

Table 3. Model Comparison Statistics and Class Prevalence (Proportions) for Latent Class Model Solutions, Weighted and Unweighted, for 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Reading Assessments

	Model Comparison Statistics					
Model	**Parameters**	**AIC**	**BIC**	**Entropy**	**LRT**	**Class Prevalence (Proportions)**
	Unweighted					
2 Classes	28	10218.3	10345.7	.979	< .001	.64 / .36
3 Classes	**38**	**7968.4**	**8141.2**	**.992**	**< .001**	**.47 / .31 / .22**
4 Classes[a,b]	48	6612.8	6831.2	.997	< .001	.22 / .24 / .12 / .43
5 Classes[a,b]	58	5661.6	5925.4	.996	< .001	.30 / .24 / .22 / .12 / .12
	Weighted					
2 Classes	28	9796.0	9923.4	.987	< .001	.71 / .29
3 Classes	38	**7326.7**	**7499.5**	**.995**	**< .001**	**.26 / .47 / .26**
4 Classes	48	5849.3	6067.6	~1.000	.001	.47 / .24 / .26 / .03
5 Classes	58	5207.3	5471.1	.999	< .001	.03 / .26 / .47 / .06 / .17

[a] The best loglikelihood value was not replicated.

[b] Standard errors of model parameter estimates may not be trustworthy for some parameters due to a non-positive definite first-order derivative product matrix.

NOTE: School sample size is approximately 700. Estimated student population size is approximately 67,900. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LRT = Lo-Mendell-Rubin Likelihood Ratio Test. Smaller AIC and BIC values indicate better fit. Class prevalence shown in order obtained in output; order may differ between unweighted and weighted solutions. Estimation messages in Mplus output indicated with superscripts. Tentatively selected solutions shown in bold. School sampling weights for 2022 applied.

SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Reading Assessments.

Table 4 presents the LCA results, including the prevalence of each class suggested by the weighted 3-class solution, along with the characteristics of the 3 classes in terms of means for the utilized indicators. Note that the selected indicators were rescaled for optimal interpretation in terms of the digits displayed (for example, average levels for median household income are presented in thousands of dollars). The results indicate that latent class 2 comprises approximately one-half of the sample, with latent class 1 and latent class 3 comprising about a quarter each of the data. In terms of the averages for the county health indicators used, one of the smaller two classes (class 1) was notably better than the others, while class 3 was notably worse.

Schools in class 1 were the "healthiest" in that their counties had lower rates of infant mortality (.41%), child mortality (.04%), low-weight births (7.6%), and disconnected youth (5.9%); fewer years lost due to premature death (5.4 years); and lower rates of frequent mental distress (12.8%) among adults on average, relative to the other classes of schools.[9] In addition, schools in class 1 were in counties with the longest life expectancies (81.3 years), the highest median income ($78,000), and the lowest rates of firearms fatalities (7.35 per 100,000 persons).

In contrast, class 3 schools were located in counties that were notably worse according to the totality of these indicators. For example, schools in class 3 were located in counties where about 10.7% of births were low weight, about 9% of youth were disconnected, the average adult was expected to live about 75 years, and the average median household income was less than $54,000. Also, the rate of firearm fatalities was about three times as great relative to class 1 schools. Schools in class 2, the largest class, were in counties with indicator values that fell between those for classes 1 and 3.

In terms of performance on NAEP (displayed in the middle section of Table 4), class 1 schools also had the highest 2022 school-average reading score (approximately 213 on the scale of 0 to 500 for grade 4 reading) relative to schools in classes 2 and 3 (which had average reading scores of approximately 206 and 189, respectively). Schools in class 1 also experienced the smallest drop in average performance between 2019 and 2022 (1.5 points) compared to classes 2 and 3 (which fell about 4.6 and 8.4 points on average, respectively). By comparison, public schools as well as "large city" schools in the nation as a whole both fell by 3 points, and students eligible for the NSLP fell by 4 points, on average (see the 2022 Nation's Report Card).

In sum, the LCA results suggest that TUDA schools with "healthier" surrounding communities and higher median household incomes had higher grade 4 reading scores in 2019 and 2022 compared to schools in less healthy and lower-SES counties. These schools' scores also "fell less" during the period of the pandemic's height compared to others. In turn, TUDA schools

[9] When the terms "heathiest,"" heathier," and "healthy" are used in this memo, their meaning is limited by the indicators used in the latent class analyses as reported in Tables 4 and 6.

with less healthy community indicators and lower median household incomes not only performed relatively worse on the grade 4 reading assessment in both years, but they experienced a greater loss in the 2019–2022 period as well.[10]

Table 4. Class Prevalence, Indicator Means, and Means for 2022 Reading Performance and 2019–2022 Performance Change for the 3-Class Weighted Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Reading Assessments

	Class 1	Class 2	Class 3
Class prevalence	26%	47%	26%
		Indicator Means	
Child/adolescent health and well-being			
Infant mortality (% of births)	.41%	.62%	.85%
Child mortality (% of youth < 18 years)	.04%	.05%	.07%
Low-weight births (% of births)	7.6%	9.2%	10.7%
Disconnected youth (% of youth ages 16-19 years)	5.9%	7.5%	8.9%
Adult health			
Premature death (years lost before age 75 per 1,000 persons)	5.402	7.473	10.856
Life expectancy (years)	81.262	78.524	74.978
Frequent mental distress (% of adults)	12.8%	13.5%	16.2%
Social and economic factors			
Median household income (1,000s of dollars)	78.061	67.819	53.866
Firearm fatalities (per 100,000 persons)	7.348	15.359	24.676
Average reading performance and change		**NAEP Means (*SE*)**	
2022 grade 4 NAEP reading score	213.10[a] (2.452)	206.19[b] (1.314)	189.44[c] (1.585)
2019–2022 change	-1.53[a] (1.446)	-4.58[a] (0.746)	-8.35[b] (1.137)
Student group		**2022 Enrollment Means**	
White	16.0%	17.2%	14.4%
Black	13.4%[a]	31.5%[b]	60.1%[c]
Hispanic	56.8%[a]	44.7%[b]	18.9%[c]
Asian	9.5%[a]	2.7%[b]	2.2%[b]
NSLP-eligible	68.5%[a]	74.2%[a,b]	77.0%[b]
		Student Absence Mean Responses	
Absent 5 or more days in the past month, 2022	26.0%	24.0%	26.0%
Absent 5 or more days in the past month, 2019–2022 change	14.4%[a]	12.2%[a]	9.4%[b]

NOTE: School sample size is approximately 700. Estimated student population size is approximately 67,900. Missingness for indicators treated via Full Information Maximum Likelihood (FIML). Scaling of indicators noted in parentheses. *SE* = standard error and is omitted except for performance means. Superscripts that differ indicate means that differ significantly ($p < .05$), so, for example, when looking at the row for 2019–2022 change, the Class 1 and Class 2 means are not significantly different from one another, but both the Class 1 and Class 2 means are significantly different from the Class 3 mean. NAEP school sampling weights for 2022 were applied.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Reading Assessments.

[10] Roughly a 3.7- to 3.9-point difference reflects a .10 *SD* difference on the scale for grade 4 reading. As such, class 3 schools were suggested to have fallen by nearly a quarter *SD* in average performance between the two assessment years. NAEP *SD* can be found at the NAEP Data Explorer: https://www.nationsreportcard.gov/ndecore/xplore/nde.

School characteristics for grade 4 reading

Post-hoc to the LCA, we examined the characteristics of the three groups of schools as they were classified in the analysis using NAEP reading data, focusing on rates of student group enrollment and rates of absences as reported by students in the NAEP student questionnaires in the 2019 and 2022 reading data (specifically, the proportion of students who reported extensive absenteeism, defined as being absent at least 5 days in the past month). These rates are displayed at the bottom of Table 4, along with the changes in rates of absences among schools in each of the latent classes between 2019 and 2022. (Note that changes in student group enrollment for the three classes of schools were minimal and hence are omitted from display.)

Among these results, it can be observed that the class of schools with the poorest community health profile (class 3 schools) also had the greatest proportion of Black and NSLP-eligible students on average, relative to the healthier classes of schools, while the class of schools with the “healthiest” community profile (class 1 schools) had the highest proportion of Asian students.

In terms of students’ average reported absenteeism in the past month, class 1 schools experienced the greatest increase between 2019 and 2022 in the proportion of students reporting having missed 5 or more days of school (14.4%). In contrast, schools in class 2 (12.2%) and class 3 (9.4%) experienced a 12.2% and 9.4% increase, respectively. Specifically, class 1 schools experienced an increase in the proportion of students who reported having missed 5 or more days of school (from 11.6% in 2019 to 26.0% in 2022) such that the proportion of students with this higher rate of absenteeism no longer differed from that seen in schools with much poorer community characteristics, as represented by the class 3 schools. This underscores that even schools located in comparatively healthier communities were affected in their conditions for learning across the COVID-19 period, aligning them more closely with schools with less healthy community features in terms of absences, which have the potential to detract from learning and connectedness to school.

Latent class analysis for grade 4 mathematics

Table 5 shows the relative fit statistics and class proportions for the unweighted and weighted LCA model results for grade 4 mathematics. As shown, the weighted 3-class solution was again suggested as providing the best fit to the data according to the statistical indices and in terms of converging to a solution that was free from warning messages in the Mplus output. Briefly, this solution fit relatively better to the data than the weighted 2-class solution according to the

AIC and BIC, had excellent entropy (.999), and its LRT indicated better probable fit than the weighted 2-class solution.[11]

Additionally, the weighted 3-class solution suggested three latent classes of schools that comprised a reasonable share of the data: one larger class comprised 48% of the sample of schools, along with two smaller classes that each comprised about a quarter (25% and 27%, respectively) of the remaining sample.

This solution mirrors what was found in the prior analysis for grade 4 reading, which is logical given that most schools are assessed in both subjects. Consequently, the school sample used here is very similar, with a high degree of similarity in the surrounding community characteristics in both samples. We thus again chose the 3-class weighted solution on a tentative basis, pending examination of its substantive interpretability.

Table 5. Model Comparison Statistics and Class Prevalence (Proportions) for Latent Class Model Solutions, Weighted and Unweighted, for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Mathematics Assessments

	Model Comparison Statistics					
Model	Parameters	AIC	BIC	Entropy	LRT	Class Prevalence (Proportions)
	Unweighted					
2 Classes[a]	28	9846.584	9973.894	.983	< .001	.37 / .63
3 Classes[a]	38	7672.953	7845.731	.999	< .001	.20 / .48 / .32
4 Classes[a]	48	6172.289	6390.534	.999	< .001	.12 / .25 / .20 / .43
5 Classes[a,b]	58	5157.744	5421.458	.999	< .001	.20 / .20 / .48 / .03 / .09
	Weighted					
2 Classes	28	9529.810	9657.120	.988	< .001	.31 / .69
3 Classes	38	**7115.731**	**7288.508**	**.999**	**.001**	**.25 / .48 / .27**
4 Classes[b]	48	5535.257	5753.503	1.000	< .001	.21 / .25 / .48 / .06
5 Classes[b]	58	4872.800	5136.514	.999	< .001	.38 / .07 / .25 / .21 / .10

a Standard errors of model parameter estimates may not be trustworthy for some parameters due to a non-positive definite first-order derivative product matrix.
b The best log likelihood value was not replicated.
NOTE: School sample size is approximately 700. Estimated student population size is approximately 66,700. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. LRT = Lo-Mendell-Rubin Likelihood Ratio Test. Smaller AIC and BIC values indicate better fit. Class prevalence shown in order obtained in output; order may differ between unweighted and weighted solutions. Estimation messages in Mplus output indicated with superscripts. Tentatively selected solution shown in bold. School sampling weights for 2022 applied.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Mathematics Assessments.

[11] Many other model solutions were accompanied by warning solutions in the Mplus output—for instance, that the model parameters may not be trustworthy—even when increased random starts and stages of optimization were implemented in Mplus programming.

Table 6 presents the LCA results for the 3-class weighted model run for grade 4 mathematics, again displaying the prevalence of each class, along with the characteristics of the three classes in terms of means for the utilized indicators. Note again that the selected indicators were rescaled for optimal interpretation in terms of the number of digits displayed.

As found in the prior results for grade 4 reading, one of the smaller two classes (class 1) was notably better than the others, while the other was notably worse (class 3). Schools in class 1 were the “healthiest” in that their counties had lower rates of infant mortality (.41%), child mortality (.04%), low-weight births (7.7%), and disconnected youth (6.0%); fewer years lost due to premature death (5.4 years); and lower rates of frequent mental distress (12.8%) among adults on average, relative to the other classes of schools. In addition, schools in class 1 were in counties with the longest life expectancies (81.3 years), the highest median income (about $77,000), and the lowest rate of firearms fatalities (7.19 per 100,000 persons).

In contrast, class 3 schools were located in counties that were notably worse according to the totality of these indicators. For example, schools in class 3 were located in counties where about 10.7% of births were low weight, about 9% of youth were disconnected, the average adult was expected to live about 75 years, and the average median household income was about $54,000. Schools in class 2, the largest class, were in counties with indicator values that fell between those for classes 1 and 3.

In terms of performance on NAEP (displayed in the middle section of Table 6), class 1 schools also had the highest average reading scores in 2022 (approximately 228 on the scale of 0 to 500 for grade 4 mathematics). While schools in class 2 performed similarly (with an average score of about 226), schools in class 3 performed significantly worse (with an average score of 209). Schools in classes 1 and 2 also experienced smaller drops in average performance between 2019 and 2022 (by 6.5 points and 8.3 points, respectively) compared to schools in class 3 (which fell by almost 11 points on average). For purposes of comparison, scores for public schools in the nation, “large city” schools, and students eligible for the NSLP fell by 5 points, 8 points, and 6 points on average (see the 2022 Nation’s Report Card).

Note that these results differ slightly from the results found for grade 4 reading, where there was greater distinction between the average performance and performance changes between class 1 (“healthiest county”) schools and class 2 (“moderately healthy”) schools, although the direction of the findings are the same. In sum, as in the prior work, these results suggest that TUDA schools with healthier surrounding communities and higher median household incomes had higher grade 4 mathematics scores in 2019 and 2022 compared to schools in less healthy and lower-SES counties. These schools’ scores also fell less during the period of the pandemic’s height compared to others. While TUDA schools in moderately healthy counties appear to have fared only slightly worse in terms of average performance and the decline in performance between 2019 and 2022, TUDA schools with considerably poorer healthy community indicators and lower median household incomes performed much worse on the grade 4 mathematics assessment in both years. They also experienced a greater decline in the 2019–2022 period, falling an average of 4 points below the *NAEP Basic* achievement level of 213.[12]

Table 6. Class Prevalence, Indicator Means, and Means for 2022 Mathematics Performance and 2019–2022 Performance Change for the 3-Class Weighted Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Mathematics Assessments

	Class 1	Class 2	Class 3
Class prevalence	25%	48%	27%
		Indicator Means	
Child/adolescent health and well-being			
Infant mortality (% of births)	.41%	.62%	.84%
Child mortality (% of youth < 18 years)	.04%	.05%	.08%
Low-weight births (% of births)	7.7%	9.2%	10.7%
Disconnected youth (% of youth ages 16-19)	6.0%	7.5%	8.9%
Adult health			
Premature death (years lost before age 75 per 1,000 persons)	5.400	7.432	10.889
Life expectancy (years)	81.303	78.578	74.978
Frequent mental distress (% of adults)	12.8%	13.4%	16.2%
Social and economic factors			
Median household income (1,000s of dollars)	76.782	67.685	54.081
Firearm fatalities (per 100,000 persons)	7.194	15.386	24.919
Average mathematics performance and change		**NAEP Means (*SE*)**	
2022 grade 4 NAEP mathematics score	227.91[a] (2.663)	225.92[a] (1.164)	208.78[b] (1.388)
2019–2022 change	-6.48[a] (1.239)	-8.30[a] (0.631)	-10.98[b] (0.763)
Student group		**2022 Enrollment Means**	
White	15.0%	17.4%	15.1%
Black	13.1%[a]	30.8%[b]	59.9%[c]
Hispanic	59.1%[a]	44.7%[b]	18.3%[c]
Asian	9.5%[a]	2.8%[b]	2.2%[b]
NSLP-eligible	69.0%[a]	73.0%[a,b]	76.7%[b]

Table notes on next page.

[12] Roughly a 3.1- to 3.3-point difference reflects a .10 *SD* difference on the scale for grade 4 mathematics. As such, class 3 schools were suggested to have fallen by nearly a quarter *SD* in average performance between the two assessment years. NAEP *SD* can be found at the NAEP Data Explorer: https://www.nationsreportcard.gov/ndecore/xplore/nde.

Table 6. Class Prevalence, Indicator Means, and Means for 2022 Mathematics Performance and 2019–2022 Performance Change for the 3-Class Weighted Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Mathematics Assessments—Continued

	Class 1	Class 2	Class 3
	Student Absence Mean Responses		
Absent 5 or more days in the past month, 2022	26.9%[a,b]	23.1%[a]	26.5%[b]
Absent 5 or more days in the past month, 2019–2022 change	15.1%[a]	10.9%[b]	8.3%[b]

NOTE: School sample size is approximately 700. Estimated student population size is approximately 66,700. NSLP = National School Lunch Program. Missingness for LCA indicators treated via Full Information Maximum Likelihood (FIML). Scaling of indicators noted in parentheses. School enrollment reflects NAEP school administrative data. NSLP eligibility refers to eligibility for a free or reduced-price lunch as provided in the NAEP student data, aggregated to the school level. Rates of absences are self-reported by students in responses to the 2019 and 2022 NAEP Student Questionnaires. *SE* = standard error and is omitted except for performance means. Superscripts that differ indicate means that differ significantly from each other ($p < .05$), so, for example, when looking at the row for 2019–2022 change, the Class 1 and Class 2 means are not significantly different from one another, but both the Class 1 and Class 2 means are significantly different from the Class 3 mean. NAEP school sampling weights for 2022 were applied.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Mathematics Assessments.

School characteristics for grade 4 mathematics

Similar to the analysis for reading, after conducting the LCA we examined the characteristics of the three groups of schools as they were classified in the analysis using NAEP mathematics data, focusing as was done for reading on two student-level demographic variables (race-ethnicity and NSLP status) as well as on absenteeism. Absenteeism was assessed using the rate of absence reported by students in the NAEP student questionnaires for the 2019 and 2022 mathematics assessments (again, specifically, the proportion of students who reported being absent at least 5 days in the past month). The resulting rates, along with the changes in rates of absences among schools in each of the latent classes between 2019 and 2022, are displayed at the bottom of Table 6. (Note that changes in student group enrollment for the three classes of schools were again minimal and are omitted from display.)

Among these results, it can again be observed that the class of schools with the poorest community health profile (class 3 schools) also had the greatest proportion of Black and NSLP-eligible students on average, relative to the healthier classes of schools, while the class of schools with the healthiest community profile (class 1 schools) had the highest proportion of Asian and Hispanic students.

In terms of students' average reported absences in the past month, class 1 schools experienced the greatest increase in the proportion of students reporting missing 5 or more days of school (15.1%), compared to class 2 (10.9%) and class 3 (8.3%) schools. Specifically, class 1 schools experienced an increase in the proportion of students reporting 5 or more days off from 11.8% in 2019 to 26.9% in 2022. This proportion did not differ from the rate seen in schools with much poorer community characteristics, represented by the class 3 schools. This again underscores that even schools located in relatively healthier communities were affected in their conditions for learning across COVID-19. Interestingly, the association with NAEP achievement appears to be greater in class 1 for mathematics (-6.5) than it was for reading (-1.5). However, it is important to exercise caution in drawing this inference, as NAEP mathematics and reading are on different scales.

TUDA distributions, reading and mathematics

Next, Tables 7 and 8 display the estimated percentages of schools that fell into the three classes for each TUDA district, ordered by modal class, for reading and mathematics, respectively. We conducted these analyses unweighted, with the understanding that the schools in our overlap samples are simply examples of schools with one of the health profiles suggested by the LCA results and that different sets of schools may be present in other NAEP assessment years. Additionally, NAEP samples some schools with particular features at higher probability (for instance, schools that are higher in student enrollment or that have higher proportions of minority students), such that schools without these features may have been less represented in our overlap sample for each subject, thus affecting the proportional distributions of TUDA schools into the three latent classes.

Note that the LCA class assignment of schools by TUDA district is fairly deterministic, given that all schools in a particular county have the same health variables. For instance, 100% of schools in San Diego and Austin were categorized as class 1 ("most healthy county") schools for both subjects, while 100% of schools in Cleveland and Milwaukee were categorized as class 3 ("least healthy county") schools. While the proportional classification of schools may make sense for some districts, as in the examples here—given their well-known health and socioeconomic characteristics—several points are relevant.

First, if all of a district's schools are located in the same county, they will all receive the same class assignment since they have identical county data. This was true for Los Angeles and the District of Columbia, for example, with both districts displaying percentages of 100% in the results above. It is also possible for a district's schools to be located in a collection of two or more counties, as was true for Austin and New York City. In such cases, the mix of classifications also depends on the heterogeneity of health and socioeconomic features across the county locations, as well as in comparison to schools in other TUDAs.

For instance, despite 14% of Austin's schools being located in counties distinct from the location of the majority of its schools, which stood at 86%, all of Austin's schools received the same classification in the analyses (class 1). Although not all of Austin's schools are located in healthy communities, a relatively higher proportion may be compared to schools in other TUDAs. In contrast, public schools sampled from New York City were located across four counties, with 50% in one county and the remaining 50% located elsewhere. Nonetheless, 80% of its schools received the same classification (class 1). These examples demonstrate that when a district's schools are located across multiple counties, class assignment depends on both the proportional location of schools across the counties and the heterogeneity of their counties' features. To distinguish among these scenarios and aid in interpreting the results, districts that have their sample of schools positioned in multiple counties are marked with an asterisk in Tables 7 and 8.

Related to this point, recall that the classification method used in the LCAs presented here is the "most likely class" method (Clogg 1995)—i.e., each school is assigned to the class with the highest probability as produced by the analysis. As such, even districts shown as having 100% of their schools being given the same classification (such as Austin) may have other potential classifications for their schools—simply at a lower probability.

Last, the indicators are reflective of students' schools rather than their homes. Some students may live in relatively unhealthy neighborhoods but travel to schools with healthier county features; the reverse may also be true. As such, some students may be affected by conditions of their residential neighborhood in ways not observed in the current analysis.

Table 7. Frequency Distribution (Percentages) of Schools' Most Probable District Location for the Weighed 3-Class Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Reading Assessments

District	*n*	Class 1 Frequency (%)	Class 2 Frequency (%)	Class 3 Frequency (%)
Austin*	30	100.0	‡	‡
Boston*	50	100.0	‡	‡
Los Angeles	10	100.0	‡	‡
Miami*	20	100.0	‡	‡
New York City*	10	80.0	‡	‡
San Diego	20	100.0	‡	‡
Albuquerque	30	‡	100.0	‡
Atlanta*	40	‡	97.1	‡
Charlotte*	20	‡	95.2	‡
Chicago	20	‡	100.0	‡
Clark County (NV)	20	‡	100.0	‡
Dallas	10	‡	100.0	‡
Denver*	20	‡	76.2	‡
District of Columbia (DCPS)	50	‡	100.0	‡
Fort Worth*	30	‡	93.8	‡
Guilford County (NC)*	40	‡	91.4	‡

Table notes on next page.

Table 7. Frequency Distribution (Percentages) of Schools' Most Probable District Location for the Weighed 3-Class Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Reading Assessments—Continued

District	*n*	Class 1 Frequency (%)	Class 2 Frequency (%)	Class 3 Frequency (%)
Hillsborough County (FL)*	10	‡	100.0	‡
Houston	40	‡	100.0	‡
Baltimore*	20	‡	‡	91.3
Cleveland	50	‡	‡	100.0
Detroit*	40	‡	‡	94.7
Duval County (FL)*	20	‡	‡	95.7
Jefferson County (KY)*	20	‡	‡	78.3
Milwaukee	30	‡	‡	100.0
Philadelphia*	20	‡	‡	89.5
Shelby County (TN)	20	‡	‡	100.0

‡ Reporting sample not met.
NOTE: Total school sample size is approximately 700. Estimated student population size is approximately 67,900. DCPS = District of Columbia Public Schools. School classification based on the "most likely class" method (Clogg 1995). Districts with schools located in multiple counties in the sample used here are marked with an asterisk. Sample size is rounded to the nearest ten.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Reading Assessments.

Table 8. Frequency Distribution (Percentages) of Schools' Most Probable District Location for the Weighed 3-Class Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Mathematics Assessments

District	*n*	Class 1 Frequency (%)	Class 2 Frequency (%)	Class 3 Frequency (%)
Austin*	30	100.0	‡	‡
Boston*	50	100.0	‡	‡
Los Angeles	10	100.0	‡	‡
Miami*	20	100.0	‡	‡
New York City*	10	80.0	‡	‡
San Diego	20	100.0	‡	‡
Albuquerque	30	‡	100.0	‡
Atlanta*	30	‡	100.0	‡
Charlotte*	20	‡	100.0	‡
Chicago	20	‡	100.0	‡
Clark County (NV)	20	‡	100.0	‡
Dallas	10	‡	100.0	‡
Denver*	20	‡	100.0	‡
District of Columbia (DCPS)	50	‡	100.0	‡
Fort Worth*	30	‡	100.0	‡
Guilford County (NC)*	40	‡	100.0	‡

Table notes on next page.

Table 8. Frequency Distribution (Percentages) of Schools' Most Probable District Location for the Weighed 3-Class Solution for the 2019/2022 TUDA Overlap Sample: 2019 and 2022 Grade 4 NAEP Mathematics Assessments—Continued

District	*n*	Class 1 Frequency (%)	Class 2 Frequency (%)	Class 3 Frequency (%)
Hillsborough County (FL)*	10	‡	100.0	‡
Houston	40	‡	100.0	‡
Baltimore*	20	‡	‡	100.0
Cleveland	50	‡	‡	100.0
Detroit*	40	‡	‡	100.0
Duval County (FL)*	20	‡	‡	100.0
Jefferson County (KY)*	20	‡	‡	100.0
Milwaukee	30	‡	‡	100.0
Philadelphia*	20	‡	‡	100.0
Shelby County (TN)	20	‡	‡	100.0

‡ Reporting sample not met.
NOTE: School sample size is approximately 700. Estimated student population size is approximately 66,700. DCPS = District of Columbia Public Schools. School classification based on the "most likely class" method (Clogg 1995). Districts with schools located in multiple counties in the sample used here are marked with an asterisk. Sample size is rounded to the nearest ten.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Mathematics Assessments.

Student-level analyses, reading and mathematics

Finally, when estimating the average performance and 2019 to 2022 performance changes at the student level in each latent class of schools, using the 20 plausible values available in the NAEP student data for each subject along with student-survey weights, the same overall pattern of results was observed as seen for school-average performance levels and changes in the LCA results.

Specifically, for grade 4 reading, the student-level analysis showed that in 2022, students attending schools in class 1 scored an average of about 213 on the NAEP scale, which was statistically significantly higher than the average scores of students in schools in classes 2 or 3, where students scored about 208 and 196, respectively, on average. These average scores were lower than the scores of students attending schools in each class in 2019 by approximately 2 points, 5 points, and 8 points, respectively.

For mathematics, students enrolled in the class 1 schools had an average mathematics score of 231 on the NAEP mathematics assessment in 2022, compared to an average score of 228 among students in class 2 schools and an average score of 214 among students enrolled in class 3 schools. In terms of performance changes between 2019 and 2022, the 2022 scores were lower than the average scores seen among students in the same sets of schools in 2019 by about 6 points, 8 points, and 11 points, respectively.

While the scores of students in each class of schools are not identical to those generated in the school-level LCAs, the student-level analyses position the three classes in a very similar manner in terms of their relative performance on NAEP in 2022 and their average decline across the height of COVID-19. That is, students in class 1 schools scored the highest in 2022 and fell the least from the scores of their 2019 counterparts. By comparison, students in class 3 schools scored the lowest in 2022 and fell the most from the scores of their 2019 counterparts, while students in class 2 schools were positioned in the middle in terms of their 2022 scores and relative change between the 2019 and 2022 assessment years. In summary, these analyses suggest that when considered at the student level, we observe a consistent pattern: a positive association between community health and students' relative performance on NAEP, even amid the challenges of COVID-19.

Discussion

This paper provided an examination of how health data can be used to contextualize changes in performance on NAEP reading and mathematics assessments in the wake of COVID-19 using latent class analysis (LCA). We focused on NAEP's TUDA sample schools, given that individual TUDA schools have greater overlap between administrations of NAEP in terms of their sampling relative to non-TUDA schools. While this limits the generalizability of the results, it allowed for an analysis of changes in school-level performance for a common sample of schools across the 2019 and 2022 administrations of grade 4 NAEP—conducted just prior to and after the height of the pandemic—when used in conjunction with data on their county-level health and socioeconomic features. Additionally, schools in a given TUDA may be more likely to share similar county-level features than non-TUDA schools, even if located in close proximity, since county borders may be used to define school districts. This use of a common sample of schools effectively controlled for other contextual features by design.

In comparing the relative performance of schools belonging to 26 distinct TUDAs, we were able to consider which county health and socioeconomic features appeared to relate to school performance during the peak of COVID-19. While the indicators included in the LCA provided only a snapshot of the health status of the locations represented by these schools, they painted a coherent portrait of three underlying subpopulations of schools in the TUDA overlap sample for each subject, each with a unique community health profile. These profiles were associated with grade 4 student performance in reading and mathematics in meaningful ways.

For example, for both the reading and mathematics assessments, the class 3 schools were observed to have had double the rates of child and infant mortality, twice as many years lost due to premature death among its adults, a quarter less median annual household income, and three times as many firearm fatalities in the period of the pandemic as schools in class 1. Furthermore, schools in class 3 not only had lower average NAEP scores among their grade 4 students in both 2019 and 2022 than schools in the other two classes, but also experienced a more pronounced decline in average performance from 2019 to 2022: by over 8 points on average in the LCA for reading and by nearly 11 points in the LCA for mathematics. In contrast, the “healthiest” groups of schools in each analysis not only had higher grade 4 reading scores in both 2019 and 2022 but fell less drastically in their scores (1.5 points on average for reading and 6.5 points on average for mathematics). These findings comport with the literature on community health and its association with student learning and performance, demonstrating the applicability of this literature to NAEP.

Additionally, the analysis of the relatively large increases in student absences in schools in healthier surrounding communities did not translate into drops in average assessment scores as substantial as those seen among schools in less healthy communities. The definitive reasons for these changes in absence rates cannot be known from the NAEP and County Health Rankings data alone—for instance, whether they are related to higher rates of COVID or other physical or mental health conditions in some locations or to differences in attitudes toward schooling, such as the importance of attendance or students’ feelings of “school belonging.” For instance, prior research has shown that risk factors at school (e.g., low feelings of connectedness to school, disengagement from class, mistreatment) and at home (e.g., health problems, family obligations, housing instability, neighborhood violence) all contribute to chronic absenteeism (Balfanz and Chang 2016; Chang et al. 2019; Stempel et al. 2017).

Overall, this finding for NAEP data is meaningful because previous research has shown that chronic absenteeism (missing 10% or more days of school) often leads to academic declines, Moreover, chronic absenteeism doesn’t just impact individual students; it also leads to lower achievement for students who have chronically absent classmates (Gottfried 2019). This highlights the role of chronic absenteeism as an aspect of the school environment that may be related to the declines in school-average performance shown in the NAEP state results. Additional qualitative data collected from students and their families may be beneficial in understanding the causes (and consequences) of the higher rates of absences seen in each group of schools.

A limitation of the study was that the data were collected at the county level. As a next step, it would be useful to explore the availability of data at a finer-grained level (e.g., district, school, or residential), such as latitude and longitude coordinates. It might also be useful to examine other ways of parceling NAEP data, such as the use of random forest analysis (e.g., Ho 1995; Breiman 2001). Lastly, path and multilevel models examining the mechanisms by which health factors might influence school performance might also be valuable—for instance, when considering the more “distal” nature of county features relative to more “proximal” school variables.

The exploratory analyses presented here point to the value of linking NAEP with other datasets—county-level health data in this case—to identify correlates of the nation’s educational performance. The analyses also show the value of latent class analysis as a way to create meaningful clusters of schools based on the linked data in that the classes derived showed relationships with children’s academic performance over time in believable ways. Using latent class analysis, health (as measured in this study) was shown to be related to the decline in grade 4 NAEP reading and mathematics performance that occurred during the period of the COVID-19 pandemic. Again, this highlights the value of linking health data with NAEP data, with the COVID-19 learning era as a particularly appropriate case in point.

References

Andriastuti, M., Ilmana, G., Nawangwulan, S. A., & Kosasih, K. A. (2020). Prevalence of anemia and iron profile among children and adolescent with low socio-economic status. *International Journal of Pediatrics & Adolescent Medicine, 7*(2), 88–92.

Balfanz, R., & Chang, H. N. (2016*). Preventing missed opportunity: Taking collective action to confront chronic absence.* Attendance Works and Everyone Graduates Center. https://www.attendanceworks.org/preventing-missed-opportunity/

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Bucholz, K. K., Heath, A. C., & Madden, P. A. (2000). Transitions in drinking in adolescent females: Evidence from the Missouri adolescent female twin study. *Alcoholism, Clinical and Experimental Research, 24*(6), 914–923.

Center for Research in Education and Social Policy. (2018). *Chronic absenteeism and its impact on achievement* (P18-002.5). Newark, DE: University of Delaware.

Chang, H. N., Osher, D., Sundius, J., & Bauer, L. (2019). *Using chronic absence data to improve conditions for learning*. Attendance Works and American Institutes for Research (AIR). https://www.attendanceworks.org/using-chronic-absence-data-to-improve-conditions-for-learning/

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). Plenum Press.

Georgieff, M. K., Ramel, S. E., & Cusick, S. E. (2018). Nutritional influences on brain development. *Acta Paediatrica 107*(8), 1310–1321.

Gottfried, M. A. (2019). Chronic absenteeism in the classroom context: Effects on achievement. *Urban Education, 54*(1), 3–34.

Ho, T. K. (1995). *Random decision forests*. Paper presented at the 3rd International Conference on Document Analysis and Recognition, Montreal, Quebec.

Keith, T. Z. (2006). *Multiple regression and beyond*. Pearson Education.

Lanza, S. T., Bray, B. C., & Collins, L. M. (2013). An introduction to latent class and latent transition analysis. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of psychology: Research methods in psychology* (2nd ed., pp. 691–716). John Wiley & Sons.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes*. Psychological Bulletin, 126*(2), 309–337.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778.

Marshall, A. T., Betts, S., Kan, E. C., McConnell, R., Lanphear, B. P., & Sowell, E. R. (2020). Association of lead-exposure risk and family income with childhood brain outcomes. *Nature medicine, 26*(1), 91–97.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

OECD (2023). *OECD Regional Outlook 2023: The Longstanding Geography of Inequalities*. Paris: OECD.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103–124.

Reardon, S. (2019, November 14). *Segregation and academic achievement gaps*. [Research presentation]. Separate and Unequal: How School Investment and Integration Matter for Educational Opportunity and Attainment. Learning Policy Institute. Washington, DC, United States.

Saegert S., Evans G.W. (2003). Poverty, housing niches, and health in the United States. *Journal of Social Issues*, 59, 569–590.

Stempel, H., Cox-Martin, M., Bronsert, M., Dickinson, L. M., & Allison, M. A. (2017). Chronic school absenteeism and the role of adverse childhood experiences. *Academic Pediatrics, 17*(8), 837–843.

Tanner, E. M., & Finn-Stevenson, M. (2002). Nutrition and brain development: Social policy implications. *American Journal of Orthopsychiatry, 72*(2), 182–193.

Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics, 33*(1), 1–67.

Vázquez Toness, B. (2023, August 11). It's Not Just the Office People Don't Want To Go To: COVID Looks To Have Permanently Severed Something as School Attendance Plummets and Keeps Dropping. *Fortune.* https://fortune.com/2023/08/11/school-chronic-absence-record-rates-skyrockets-kids-stay-home/

Vermunt, J. K. (2002). Comment: Latent class analysis of complex sample survey data: Application to dietary data. *Journal of the American Statistical Association, 97*(459), 736–737.

Appendix A

Table A-1. County Health Indicators Considered for Analysis and Stage of Omission or Retention for LCA: 2022 County Health Rankings & Roadmap and 2019/2022 TUDA Overlap Sample Data

	Met Criterion 1 (Hypothesized Association with Change)	Met Criterion 2 (Rate of Complete Data)	Met Criterion 3 (Empirical Association with Change)
Child/Adolescent Health and Well-being			
Child mortality, rate (children < 18 years)	Yes (-)	Yes (100.0%)	Yes, *p* = .003 (-)
Infant mortality, rate (infants)	Yes (-)	Yes (99.8%)	Yes, *p* = .001 (-)
Low birthweight, rate (infants)	Yes (-)	Yes (100.0%)	Yes, *p* < .001 (-)
Disconnected youth, rate (youth ages 16-19)	Yes (-)	Yes (99.8%)	Yes, < .001 (-)
Teen births, rate (females ages 15-19)	No		
Children in poverty, rate (children < 18 years)	Yes (-)	Yes (100.0%)	No, *p* = .482 (-)
Children in single-parent households, rate	Yes (-)	Yes (100.0%)	No, *p* = .139 (-)
School funding adequacy, gap (dollars) between actual and required	Yes (-)	Yes (99.4%)	No, *p* = .651(+)
Childcare cost burden, percent of median household income	No		
Childcare centers, rate per 1,000 under 5 years	No		
Adult Health (adults age 18+ unless otherwise noted)			
Premature deaths, years lost before age 75 per 1,000 persons	Yes (-)	Yes (100.0%)	Yes, *p* < .001 (-)
Life expectancy, years	Yes (+)	Yes (100.0%)	Yes, *p* < .001 (+)
Poor or fair health, rate	Yes (-)	Yes (100.0%)	No, *p* = .149 (+)
Poor physical health days, number in past 30 days	No		
Poor mental health days, number in past 30 days	No		
Preventable hospital stays, rate	No		
Adult smoking, rate	No		
Excessive drinking, rate	No		
Insufficient sleep, rate	No		
Frequent physical distress, rate	No		
Frequent mental distress, rate	Yes (-)	Yes (100.0%)	Yes, *p* < .001 (-)
Adult obesity, rate	No		
Diabetes, rate (adults age 20+)	No		
General Population Health (among population total)			

Drug overdose deaths, rate	No		
Injury deaths, rate	No		
HIV prevalence, rate	No		
Sexually transmitted infections, rate of chlamydia	No		
Social associations, average number	Yes (-)	Yes (100.0%)	No, *p* = .209 (+)
COVID-19 age-adjusted mortality, total in 2020 (quartiles)	Yes (-)	Yes (100.0%)	No, *p* ³ .488 (for upper quartiles)
Driving Risks			
Motor vehicle crash deaths, rate (population total)	No		
Driving alone to work, rate (among workers)	No		
Long commute, rate (among workers)	No		
Alcohol-impaired driving deaths, rate (population total)	No		
Health Care Access (among population total unless otherwise noted)			
Uninsured persons, rate	No		
Uninsured adults, rate	No		
Uninsured children, rate	No		
Ratio of primary care physicians to population	Yes (+)	Yes (100.0%)	No, *p* = .413 (-)
Ratio of other primary care providers to population	No		
Ratio of mental health providers to population	Yes (+)	Yes (100.0%)	No, *p* = .725 (+)
Ratio of dentists to population	No		
Flu vaccinations, rate	Yes (+)	Yes (100.0%)	No, *p* = .003 (-)
Mammogram screening, rate (females ages 65-74)	No		
Food and Exercise (among population total)			
Food insecurity, rate	Yes (-)	Yes (100.0%)	No, *p* = .114 (-)
Limited access to healthy foods, rate	Yes (-)	Yes (100.0%)	No, *p* = .071 (-)
Food environment index, value on scale from 1 to 10	Yes (+)	Yes (100.0%)	No, *p* = .030 (+)
Physical inactivity, rate	No		
Access to exercise opportunities, rate	No		
Socioeconomics (among population total unless otherwise noted)			
Unemployment, rate (persons age 16+ years)	Yes (-)	Yes (100.0%)	No, *p* = .106 (-)
Severe housing problems, rate	Yes (-)	Yes (100.0%)	No, *p* = .006 (+)
Households with high housing costs, rate	No		
Households with overcrowding, rate	Yes (-)	Yes (100.0%)	No, *p* = .003 (+)
Households with lack of kitchen or plumbing facilities, rate	No		
Broadband access, rate	No		
Median household income, dollars	Yes (+)	Yes (100.0%)	Yes, *p* = .002 (+)
Homeownership, rate	No		
4-year high school graduation, rate	No		

Some college, rate (adults ages 25-44)	Yes (+)	Yes (100.0%)	No, *p* = .893 (+)
High school completion, rate (adults age 25+)	Yes (+)	Yes (100.0%)	No, *p* = .001 (-)
Environment Quality / Pollution			
Air pollution - particulate matter, average mcg per square meter	Yes (-)	Yes (100.0%)	No, *p* = .090 (+)
Drinking water violations, “yes/no”	No		
Traffic volume, average volume per meter of major roadways	No		
Crime and Violence (among population total unless otherwise noted)			
Violent crime, rate	Yes (-)	Yes (100.0%)	No, *p* = .057 (-)
Homicides, rate	Yes (-)	Yes (99.8%)	No, *p* = .025 (-)
Suicides, rate	No		
Firearm fatalities, rate	Yes (-)	Yes (100.0%)	Yes, *p* = .003 (-)
Processed juvenile delinquency cases, rate (among juveniles)	Yes (-)	No (79%)	
Segregation and Inequality			
Residential segregation - Black/White, value from 1 to 100	No		
Residential segregation - non-White/White, value from 1 to 100	No		
School segregation, value from 1 to 100	Yes (-)	Yes (100.0%)	No, *p* = .753 (+)
Income inequality, ratio of 20th to 80th percentiles	No		
Gender pay, ratio of women’s to men’s (among full-time workers)	No		
Population Composition			
Population total, rate	Yes (-)	Yes (100.0%)	No, *p* = .013 (+)
Percent non-proficient in English, rate	Yes (-)	Yes (100.0%)	No, *p* < .001 (+)
Percent rural, rate	No		

NOTE: Total school sample size is approximately 700. Scaling of NAEP data are described at: https://nces.ed.gov/nationsreportcard/tdw/analysis. Empirical associations shown for Criterion 3 obtained using NAEP school-weighted regression analysis, and control for average student eligibility for free or reduced-price lunch via the National School Lunch Program (NSLP) as provided in the 2022 NAEP, aggregated to the school level. All county health characteristics are averages. Gray shading indicates the stage of omission from retaining among the health variables used for the analysis. Further detail on health variables is available at https://www.countyhealthrankings.org; mcg = micrograms.
SOURCE: University of Wisconsin Population Health Institute, 2022 County Health Rankings & Roadmap; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 and 2022 Grade 4 Reading Assessments.

Summary of indicator selection process

Criteria 1 and 2

In Table A-1, the first column indicates whether or not each county measure met the first criterion—having a practical, hypothesized association with changes in school-average performance on NAEP due to COVID-19 (marked as either "Yes" or "No"), along with the direction of the hypothesized association, if applicable (marked as "+" or "-"). In terms of direction, "+" indicates that the variable was thought to result in a positive change or a "less negative decline," while "-" indicates that the variable was thought to result in a negative change or a "more negative decline."

On the basis of meeting the first criterion, we retained numerous measures of *child health and well-being* for their direct influence on children's academic performance. Among these are several measures of preexisting or comorbid health, social, and economic conditions that may influence a child's risk for contracting COVID-19 or their performance in school during the pandemic (e.g., rates of low birthweight, disconnected youth, children in poverty), as well as a measure of the adequacy of school funding. We omitted a measure of teen births, which were infrequent in this sample (with a county maximum of 3%), and measures of childcare center availability and childcare cost (as most children would be more likely to stay in their own home, with their own families, at the height of the pandemic).

Among the measures of *adult and general population health*, we retained variables that most directly and comprehensively capture adult health conditions that in theory relate to adults' ability to support youth in the pandemic period. Among these were measures of premature death, life expectancy, and self-report of poor/fair health or frequent mental distress (as the most comprehensive summary measures of physical and mental well-being) and the average number of social associations (as a measure of adults' degree of social support). We also included county rates of deaths due to COVID-19 in 2020, which may capture adults' health and ability to support youth, specifically throughout the 2019–2022 period, and thus could be related to changes in school-average performance during this time. This variable also serves as a comprehensive measure of counties' relative risk and preparedness for being affected negatively by COVID-19 and thus also relates to these changes. We omitted variables that were redundant with these measures (e.g., self-report of frequent physical distress is redundant with the information captured by life expectancy and overall poor/fair health) or that captured only a more specific aspect of adult health (e.g., rates of smoking and obesity). We also omitted a small group of measures of *driving risk*, which have no strong theoretical association with school performance, particularly in a period in which most people were likely to commute and travel less.

In terms of access to *health care access and access to food and exercise opportunity*, we retained measures of the ratio of primary care physicians and mental health providers to the population (as supports for child and adult physical and mental health in general and during COVID-19 in particular); rates of flu vaccinations (which relate to respiratory health and thus to coronavirus vulnerability); and measures of access to healthy food (since lower-SES students attending school remotely in this period were likely to lack access to meals they may have received via the NSLP, which may affect their academic performance). We excluded measures related to lacking insurance, which are likely redundant with the measures of family socioeconomic well-being described below and had low rates among the schools' counties (e.g., 11% on average for uninsured persons overall). We also omitted ratios of other health care providers to the population, rates of mammogram screening, and access to exercise opportunity, which have weaker practical associations with well-being during COVID-19.

In terms of *socioeconomics*, we retained measures of county-level family employment, income, and education for their positive, direct influence on the ability of families to support their level of well-being during COVID-19. In addition, we retained an index of severe housing problems and the component problem that likely relates most closely to family well-being and to children's academic success when learning from home: the rate of household overcrowding. (See Table A-1 for the remaining component items, which we omitted from further consideration.) We omitted measures of homeownership and 4-year high school graduation (as families' ability to support student learning should not differ notably according to whether persons in the county rent or own their homes or whether or not they graduated high school in 4 years, compared to more direct measures of county-level socioeconomics, such as median household income). We also omitted a measure of county-level broadband access (as the 2022 NAEP data suggest that many schools in this sample provided access for students who lacked this resource).

Among the available measures of *environmental quality/pollution*, we retained only a measure of air pollution for its direct impact on respiratory health. Among measures of *crime and violence*, we retained measures that were most comprehensive and had the highest rates in this sample (omitting a measure of county suicides, which had an average of only 1% and a maximum of 3% across the schools' counties). These measures may have shaped the overall context for safety and well-being at a time when families and youth were likely to work and learn at home and thus be impacted by these factors.

Among the measures of *segregation/inequality*, we retained an index of school segregation, which may provide additional context on the adequacy of school resources for supporting student learning, given its known association with school socioeconomic well-being (e.g., Reardon 2019). Finally, among the measures of *population composition*, we retained two measures: (1) the counties' population total (as a reflection of county density, which may affect the likelihood of contracting the coronavirus as well as competition for available health and social resources); and (2) the rate of non-proficiency in English among the counties' population (due to its impact on health literacy and ability to access these same resources). We omitted measures of residential segregation and income inequality—which likely have weaker associations with family well-being and the ability to support student learning relative to the socioeconomic measures described above—and the county percentage of rural land, which was low on the whole for the TUDA population of "large city" schools (with an average of only 3%).

Among these retained measures, all but one—rates of processed juvenile delinquency cases—also met the second criterion (having at least 80% complete data for the TUDA overlap sample schools). The second column in Table A-1 displays rates of complete data among the retained indicators.

Criterion 3

Measures that met both Criteria 1 and 2 were next examined using regression analysis to determine whether they met the third criterion: demonstrating a statistically significant association (using an alpha level of .01, given the number of indicators examined) in the theoretically expected direction with school-average change scores between the 2019 and 2022 reading assessments, after accounting for the school-average NSLP eligibility of the enrolled students in the 2022 NAEP data. In Table A-1, the third column indicates whether each retained measure met the final criterion, denoted by either "Yes" or "No," along with the corresponding *p* value and direction of empirical association obtained via the regression model (again using "+" and "-" symbols to indicate positive and negative associations, respectively).

Those measures meeting the final criterion are indicated using a green font and are displayed in tables of results for the LCAs in our main text. Note that the measures that met all three selection criteria are grouped neatly into three categories: child/adolescent health and well-being, adult health, and socioeconomic indicators.

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit AIR.ORG.

American Institutes for Research®
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
202.403.5000 | **AIR.ORG**