American Educational Research Journal
October 2025, Vol. 62, No. 5, pp. 982–1018
DOI: 10.3102/00028312251355990
Article reuse guidelines: sagepub.com/journals-permissions

Summer School as a Learning Loss Recovery Strategy After COVID-19: Evidence from Summer 2022

Ian Callen
University of Virginia
Maria V. Carbonari
University of Pennsylvania
Michael DeArmond
American Institutes for Research, CALDER
Daniel Dewey
Harvard University
Elise Dizon-Ross
Dan Goldhaber
American Institutes for Research, CALDER
Jazmin Isaacs
NWEA
Thomas J. Kane
Harvard University
Megan Kuhfeld
NWEA
Anna McDonald
American Institutes for Research, CALDER
Andrew McEachin
Educational Testing Service
Emily Morton
NWEA
Atsuko Muroga
Harvard University
Douglas O. Staiger
Dartmouth College

To make up for pandemic-related learning losses, many U.S. public school districts have increased enrollment in their summer school programs. We assessed summer school as a strategy for COVID-19 learning recovery by tracking the academic progress of students who attended summer school in

2022 across eight districts serving 400,000 students. Using value-added models that control for students' demographics and prior achievement, we estimated a positive effect of summer school on math test achievement (0.03 standard deviations) but not on reading tests. With only 13% of students participating, these districts' summer programs closed ~2% of their total pandemic-related learning losses in math and none of their losses in reading.

KEYWORDS: academic interventions, achievement, COVID-19 recovery, pandemic recovery, summer school

Introduction

To make up for pandemic-related learning losses, many U.S. public school districts expanded summer school. In their plans for spending the American Rescue Plan's Elementary and Secondary School Emergency Relief funds (ESSER III), for example, districts listed *summer learning* as a top recovery strategy (DiMarco & Jordan, 2022; Roche, 2023). Nationwide, 70% of districts expanded summer school in response to the pandemic, according to a 2022 survey (Diliberti & Schwartz, 2022). System leaders

---

IAN CALLEN is a doctoral student in the School of Education and Human Development at the University of Virginia.

MICHAEL DEARMOND is director of policy at the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research.

DANIEL DEWEY is a research analyst at the Center for Education Policy Research at Harvard University.

ELISE DIZON-ROSS is a researcher at the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research.

DAN GOLDHABER is director of the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research and director of the Center for Education Data and Research (CEDR) at the University of Washington.

JAZMIN ISAACS is a research analyst at NWEA.

THOMAS J. KANE is the Walter H. Gale Professor of Education and Economics and faculty director of the Center for Education Policy Research at Harvard University.

MEGAN KUHFELD is director of growth modeling and data analytics at NWEA.

ANNA MCDONALD is a research assistant at the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research.

ANDREW MCEACHIN is senior research director of policy research at the Educational Testing Service.

EMILY MORTON is a lead research scientist at NWEA.

ATSUKO MUROGA is a postdoctoral fellow at the Center for Education Policy Research at Harvard University.

DOUGLAS O. STAIGER is the John Sloan Dickey Third Century Professor in the Department of Economics at Dartmouth College.

bolstered summer school in order to "address learning loss, provide individualized instructional support and offer . . . acceleration," explained Alberto Carvalho, superintendent of the Los Angeles Unified School District (Harter, 2023).

Summer school's popularity as a COVID-19 academic recovery strategy is unsurprising. In contrast to strategies that might require new programming (e.g., high-dosage tutoring or an extended school year), many districts have prior experience delivering summer school, and leaders can draw on preexisting administrative and delivery routines to expand their programming. Because summer programming is a familiar concept and is mostly voluntary, it also is less contentious than catch-up strategies that, for instance, add mandatory extra days to the school year (MacGillis, 2023). But whether any of this makes summer school an effective response to COVID-19 learning loss is unclear—most of the literature on summer school predates the pandemic (Augustine et al., 2016; Cooper, 2001; Cooper et al., 1996; Kim & Quinn, 2013; Schwartz et al., 2018).

This study contributes the first large-scale evaluation of ESSER-funded summer learning programs implemented after most schools had returned to in-person learning in 2021–22. The paper uses value-added models to evaluate the effect of summer school on student achievement and places the results in the broader context of COVID-19-related learning loss and the adequacy of the system's response.

To estimate the effect of summer school on COVID-19 learning recovery, we tracked the academic progress of individual students who attended summer school in 2022 across eight districts serving 400,000 students. Controlling for student characteristics and spring 2022 achievement, we found an average value-added effect of summer school participation of 0.03 SD on math.[1] We did not find statistically significant effects on reading.

To put these results in context, consider that districts' spring 2022 state test scores in math lagged their 2019 scores by an average of 0.15 SD across the United States[2] and by an average of 0.19 SD in our sample (Reardon et al., 2023). Given the fact that 13% of students in the average district in our sample participated in summer school, we estimate that summer school closed about 2% of our sample's pandemic-related learning losses in math. With increased duration and attendance, summer school programs may help more students recover, but it is clear that summer school alone is insufficient for full recovery in most districts.

Background

Multiple assessments show that student achievement declined during the pandemic. Results from the National Assessment of Educational Progress (U.S. Department of Education, 2022, 2023), NWEA's MAP Growth tests (Lewis & Kuhfeld, 2022, 2023), and Curriculum Associates' i-Ready assessments

(Curriculum Associates, 2020) all document pandemic-era losses. The declines were especially pronounced for students with lower test scores and students from historically marginalized groups (Callen et al., 2024; Dorn et al., 2021; Kilbride et al., 2021; Lewis et al., 2021).

School districts nationwide responded to the pandemic with a range of interventions to help students catch up academically, including expanded summer school programming. Summer school's potential as a recovery strategy is supported by prepandemic research. The broadest prepandemic evidence about the efficacy of summer interventions came from two meta-analyses that, respectively, examined reading and math outcomes. Kim and Quinn's (2013) meta-analysis included 41 studies of summer programs that provided reading instruction to students in grades K–8. This meta-analysis found wide variation in effects across programs but concluded that classroom-based summer programs improved reading test achievement for students from low-income households by ~0.09 SD.[3] The other meta-analysis, by Lynch et al. (2023), examined 37 studies of summer programs with math instruction. The authors found positive effects (+ 0.10 SD) on math achievement across income levels (Lynch et al., 2023).[4] Neither meta-analysis detected significant differences in effects across grade levels.

Additional support for summer school's potential as a recovery strategy comes from a randomized, controlled trial (RCT) of voluntary district-run summer programs conducted by researchers at the RAND Corporation (Augustine et al., 2016). The RCT examined the effects of offering 2 years of voluntary classroom-based summer programming for ~3,000 rising fourth grade students across five districts. It found that students' math achievement improved after the first summer (+ 0.08 SD), although the effect dissipated by the fall of the next year; the study found no impact on reading achievement. The RCT also highlighted the difficulty districts can have persuading students to enroll and participate in voluntary summer school. In the study's first year, researchers reported that 21% of students did not show up to the program at all. Among those who did, 29% had low attendance (attended <80% of days), and about half had high attendance (>80%). During the second summer of programming, the share of students with high attendance dropped to 30%. Given the study's intent-to-treat analysis, these attendance issues suggest that the results likely understate the program's benefits for students who attended.

In contrast to the prepandemic literature, evidence on the effectiveness of *pandemic-era* summer programs is sparse and mixed. A study of one metro-Atlanta district's 2021 summer program found low rates of participation (18% of invited students) and no effect of being invited to the program on achievement (Barry & Sass, 2022). However, Gambi and De Witte (2024) found that summer programs in Belgium in 2020 boosted achievement for vulnerable students and stopped achievement gaps between schools from widening. It is unclear, however, whether these international findings generalize to academic recovery programs in the United States during the time period of this study (summer of 2022), when COVID-19 infection rates were lower and students were mostly back to in-person schooling.

It is important to study the effectiveness of pandemic-era summer programs because they may systematically differ from prepandemic programs in ways that matter for their impacts. Prepandemic evidence suggests that variation in eligibility and invitation rules, curricula, class sizes, amounts of daily instructional time, and overall durations all may influence program impacts on achievement (e.g., Bell & Carrillo, 2007; Boss & Railsback, 2002; McEachin et al., 2018). The researchers involved in the RAND RCT suggested a few key strategies—investing in personalized program recruitment, setting firm enrollment deadlines, having clear attendance policies (and offering incentives, if possible), and establishing systems for monitoring enrollment and attendance—to maximize attendance rates and program effects (Schwartz et al., 2018). They also emphasized the importance of prioritizing high-quality planning and curricula, hiring the district's most effective teachers, offering academic instruction (3–4 hours daily) *and* enrichment activities, and scheduling programming 5 days per week for 5–6 weeks (Schwartz et al., 2018). But implementing a summer program with all or many of these features was difficult even during prepandemic times; despite receiving guidance on program design, several programs in the RAND study failed to meet the recommended criteria.

Different aspects of the pandemic context may further support or hinder summer program effectiveness. On the one hand, pandemic-era summer programs have the advantage of increased access to funds through ESSER. These extra resources could have helped districts recruit more effective teachers for summer school, increase program duration, reduce class sizes, or expand enrollment efforts. On the other hand, the pandemic created significant implementation challenges for many systems. Teacher burnout (Steiner et al., 2022) and spikes in chronic absenteeism (Dee, 2024) may have prevented districts from increasing the scope or effectiveness of their summer programs during and after the pandemic (Carbonari et al., 2024a; Dee, 2024).

Indeed, recent evidence from a nationally representative survey in 2023 suggested that pandemic-era summer school programs nationwide fell short of the best-practice recommendations from the prepandemic literature (Schwartz et al., 2018). Nationwide, just 18% of districts indicated that they ran a summer program lasting the recommended 5 weeks and offering 3 hours of academic instruction daily (Diliberti & Schwartz, 2024). According to the survey, districts enrolled[5] around 42% of eligible elementary school students in programs that restricted participation to those performing below grade level (about half of all programs). In districts with open-enrollment policies, ~31% of eligible elementary school students enrolled (Diliberti & Schwartz, 2024).

While Diliberti and Schwartz (2024) interpreted this 31% enrollment rate as discouragingly low, this rate was substantially higher than participation rates for other popular recovery initiatives. Recent evidence on large-scale tutoring during the pandemic, for example, found that none of the programs studied reached more than 13% of the districts' K–8 students; all had null effects (Carbonari et al., 2024b). The scale of these summer programs also was substantially larger than those of tutoring programs studied prior to the pandemic. In large districts with 15,000 or more K–12 students, enrolling 31% of elementary school students in summer school translates to running a program for >2,000 students (and many programs also served secondary school students). By comparison, the broader evidence on prepandemic tutoring included few studies of programs that served >400 students and found that program effects declined drastically as program size increased (Kraft et al., 2024b; Nickow et al., 2024).

In short, even if pandemic-era summer school programs yield smaller impacts on achievement to what was found in the prepandemic literature, they still may offer a plausible academic recovery strategy to reach meaningful numbers of students. The question of scale is especially important in the context of the pandemic because learning loss was often a district-level phenomenon that affected students similarly across schools (Callen et al., 2024; Goldhaber et al., 2023). All students in impacted districts likely need at least some supplemental support to recover to prepandemic achievement levels. Delivering academic recovery interventions to as many students as possible is key for districtwide recovery. Summer school is a practical strategy on this account.

Interestingly, recent evidence suggests that students have gained more ground academically during summers than during school years in the wake of the pandemic (Lewis & Kuhfeld, 2023). Relative to the "summer slide" in 2019, for example, the slide in 2022 decreased, on average, from −0.09 to −0.07 SD in reading and from −0.20 to −0.17 SD in math across grades 3–8 (Lewis & Kuhfeld, 2022). Whether this attenuated summer slide is related to expanded summer programming, however, is unclear.

Research Questions

In this study, we examined how, if at all, summer learning programs supported students' academic recovery from the pandemic. We used a detailed dataset from eight districts that included information about program characteristics, student eligibility and attendance, and academic outcomes to answer the following questions:

1. How did 2022 summer interventions affect academic achievement?
2. To what extent did these 2022 summer programs contribute to districts' academic recovery from the pandemic?

Table 1
Sample demographics

Demographics	R2R project districts	U.S. public schools
Average district enrollment	50,084	2,674
Average school enrollment	678	497
FRPL eligible (%)	55.2%	45.4%
Race (%)		
Asian	4.7%	2.9%
Latino/a	32.3%	20.1%
Black	23.2%	12.7%
White	33.5%	59.4%
School locale (%)		
City	87.37%	18.5%
Suburb	7.02%	22.7%
Town	0.00%	14.2%
Rural	5.62%	42.1%

R2R, Road to Recovery; FRPL, free or reduced-priced lunch.
Note. Data are from the Common Core of Data collected by the National Center for Education Statistics during the 2020–21 school year.

Methods

Data and Measures

This report draws on data from eight districts participating in the Road to Recovery (R2R) project, an ongoing partnership between researchers and school districts that aims to provide districts with timely feedback on their academic recovery interventions. Of these districts, eight provided data to participate in the summer 2022 analysis. These eight districts comprise the sample for this report.[6] Together they collectively enroll ~400,000 students across seven states. As displayed in Table 1, the study districts were mostly large and urban, and they served higher percentages of Black and Latino/a students (56%) and students eligible for free and reduced-price lunch (55%) compared with national averages (33 and 45%, respectively).[7] Table 2 shows estimates of these districts' academic recovery as of spring 2022 in math (panel A) and reading (panel B) based on the Education Recovery Scorecard[8] (Reardon et al., 2023). We see substantial variation in recovery across the districts because estimates of the change in their achievement from 2019 to 2022 ranged from −0.02 to −0.40 SD in math and from + 0.01 to −0.37 SD in reading. However, on average, the remaining learning loss as of spring 2022 in our sample was substantial (−0.19 SD in math and −0.15 SD in reading), and it was greater than the national district average (−0.15 SD in math and −0.09 SD in reading).

Table 2
Estimated achievement loss and recovery from spring 2019 to 2022, grades 3–8

District	Spring 2019 (SD)	Spring 2022 (SD)	Change from spring 2019 to spring 2022 (SD)	Change from spring 2019 to spring 2022 (grade levels)
Panel A: Math				
Alexandria	−0.11	−0.51	−0.40	−1.30
Anonymous[a]	0.00	−0.03	−0.02	−0.10
Dallas	−0.06	−0.21	−0.15	−0.49
Guilford	−0.09	−0.25	−0.16	−0.56
Portland	0.04	−0.04	−0.09	−0.33
Richardson	0.19	0.04	−0.15	−0.51
Tulsa	−0.68	−1.06	−0.39	−1.33
Study district average	−0.10	−0.29	−0.19	−0.66
National district average	0.04	−0.10	−0.15	−0.51
Panel B: Reading				
Alexandria	−0.13	−0.40	−0.27	−0.99
Anonymous[a]	0.03	0.03	0.01	+ 0.03
Dallas	−0.31	−0.33	−0.02	−0.08
Guilford	−0.03	−0.20	−0.17	−0.60
Richardson	−0.02	−0.10	−0.08	−0.29
Tulsa	−0.60	−0.96	−0.37	–1.34
Study district average	–0.18	–0.33	–0.15	–0.55
National district average	0.37	–0.05	–0.09	–0.32

Note. All estimates are from the Stanford Education Data Archive (Version SEDA 2022 2.0; Reardon et al., 2023) and are scaled such that a 0 in this metric is equal to the National Assessment of Educational Progress average (in grade 5.5) in spring 2019, and 1 unit in this metric is equal to one student-level SD. Estimates in this scale are comparable across the whole country and over time, but they are not comparable across subjects.
[a]Enrollment/racial composition changed significantly in this district from 2019 to 2022, and results should be interpreted with caution.

For our analysis, we constructed a sample of students from these eight districts that met the following criteria: a student was expected to be entering grade 1–grade 8 in fall 2022 and had NWEA MAP Growth scores for both the spring 2022 and fall 2022 tests in reading or math.[9] Our combined district analytic samples included 129,721 students, with student counts ranging from 1,804 to 39,248 across districts and subjects.

Data for this study came from three primary sources: (a) interviews about program characteristics with district leaders, (b) student-level eligibility and program participation data provided by the districts, and (c) NWEA MAP Growth assessments. We describe each in more detail below. For this study, we defined a summer learning program as a program the district considered to be an academic recovery program and that delivered formal academic instruction in math and/or English language arts over the summer.[10]

Interviews

To begin, we collected information about the design and implementation of summer learning programs in fall 2022 via semistructured interviews with summer programming leaders in each district. We conducted nine interviews, all of which were conducted virtually and lasted 60 minutes. Our interview questions focused on the design of each program, including student eligibility criteria, invitation processes, program duration and intensity (i.e., hours per day), daily hours of instruction in each subject, delivery mode (e.g., virtual or in person), and staffing.[11] We also asked whether the summer program included a tutoring component; if yes, we asked how students were identified to receive the tutoring, when it happened and for how long, and what students who did not receive tutoring did during the tutoring time. If time allowed at the end of each interview, we asked the leaders to reflect on the implementation successes and challenges they experienced. After the interviews, the research team reviewed any documents shared by program leaders and followed up via email to resolve remaining questions. Notes from the interviews were captured in a template that was shared with the leaders during, and then again after, the interview to ensure accuracy.[12] We summarize the key design elements of each program in the "Results" section and Table 3.

Student-Level Eligibility, Participation, and Dosage

Following the interviews, each district received a data request tailored to their district and programs. These data requests asked about the summer school enrollment and attendance of individual students as well as details about student demographics for the 2021–22 and 2022–23 school years. The requests yielded a variety of summer program–specific eligibility and participation student-level variables, including whether a student was recommended or invited to attend summer school, their summer school enrollment status, summer school site attended, and daily attendance. We also collected data on whether students were offered multitiered systems of support in math and English language arts and students' previous state test scores. For our analysis, we counted students as attending summer school if they attended at least 1 day. We measured summer school dosage as the average number of days a student attended the program (contingent on any attendance), the number of hours of instruction they received each day, and the percentage of total program days they attended.

Table 3
Designs of summer programs

Factor	District 1	2	3	4	5	6	7	8
Subject	Math and ELA	Math and ELA	Math and ELA	Math and ELA	Math and ELA	Math and ELA	Math and ELA	Math and ELA
Grade level	Rising K–12	Rising pre-K–8	Rising K–8	Rising pre-K–8	Rising 1–8	Rising K–12	Rising 1–8	Rising 1–9
Eligible schools	All schools	Regular calendar schools	All schools	All schools	All schools	All schools	All schools	All schools
Participation: opt in or by invitation?	Opt in	Opt in and by invitation	Opt in and by invitation	Opt in and by invitation	Opt in	By invitation	Opt in and by invitation	Opt in and by invitation
Invitation: to whom?	None	Students scoring below grade-by-subject threshold on state tests	Students with academic, SEL, or other needs, based on own prioritization matrix	Students scoring below grade-by-subject threshold on state or MAP Growth tests	None	Low-scoring students	Low-performing and historically underserved students	Students identified as academically behind
Location	Hub sites	Hub sites	Hub sites	Hub sites	Hub sites	Hub sites	Hub sites	Hub sites
Providers	District teachers	District teachers	District teachers	District teachers	District teachers	District teachers	District teachers	District teachers
Delivery	In person	In person	In person	In person	In person	In person	In person	In person
Intended frequency	5 days/week	4 days/week	4 days/week	4 days/week	5 days/week	3–4 days/week	5 days/week	4 days/week
Intended dosage	4 weeks, 19 days	4 weeks, 15 days	4 weeks, 15 days	5 weeks, 20 days	3 weeks, 15 days	12–18 days	4 weeks, 20 days	6 weeks, 17 days
Academic time per day per subject	90 minutes	90 minutes	45–90 minutes	90–100 minutes	90 minutes	60–120 minutes	60 minutes	120 minutes
Operating hours	Extended hours	Full school day	Short school day	Short school day	Short school day	Short school day	Short school day	Full school day
Other programming	—	Tutoring	Virtual summer program	Tutoring	—	Tutoring	—	—

ELA, English language arts; SEL, social-emotional learning.

Note. Data are based on interview notes with school district leaders conducted between fall 2022 and early spring 2023.

NWEA MAP Growth Assessments

The achievement outcome data in this study were math and reading test scores from the NWEA MAP Growth longitudinal student achievement database. School districts use MAP Growth assessments to monitor student achievement and growth in reading and math over the course of the school year. In most districts, the tests are administered three times each year: in the fall, winter, and spring.[13] Relative to state tests administered each spring, MAP Growth assessments are particularly well suited for this study because they can more narrowly isolate changes in student achievement over the summer (i.e., between spring and fall assessments between 2 school years) rather than from spring to spring. Relative to a fixed-form test, the computer-adaptive format of the MAP Growth assessment increases precision at the high and low ends of the distribution; this increased precision is pertinent in the context of the pandemic because many students are performing below grade level.

We standardized NWEA MAP Growth scores by subject and grade level using the NWEA 2020 MAP Growth norms (Thum & Kuhfeld, 2020), which are based on a nationally representative sample of students in the prepandemic school years (i.e., 2015–16, 2016–17, and 2017–18). We defined *grade level* as students' expected rising grade in fall 2022 based on the grade they were enrolled in as of spring 2022 to have a consistent measure of grade across students regardless of whether a student repeated or skipped a grade in the following year.[14] Normalizing the scores enabled us to assess students' academic performance relative to a prepandemic nationwide distribution of test scores.[15] The NWEA dataset also includes student-level demographic data on race/ethnicity and gender. School-level demographic and enrollment data linked to the NWEA dataset were from the 2020–21 Common Core of Data collected by the National Center for Education Statistics.

Empirical Approach

We used value-added models (VAMs) to estimate the effect of each of the eight summer programs on MAP Growth test scores, using the previous spring as the baseline and the subsequent fall score as the outcome:

$$\begin{aligned} MAP_{math,ijt} = {} & \beta_0 + \beta_1 SS_i + \beta_2 Other_{i,math} + \beta_3 Other_{i,ela} + \tau \mathbf{X}_{it} \times \; Grade_{it} \\ & + \; \psi Sch\_Grade_{ij,t-1} + \; \varepsilon_{ijt} \end{aligned} \tag{1}$$

where $MAP_{math,ijt}$ denotes math achievement for student i in school j in term t as measured by the standardized math MAP Growth score and where t is fall 2022. Our main treatment variable was the SS_i term, a binary indicator equal to 1 if student i attended at least 1 day of summer school during summer 2022. The estimated effect of participating in the summer school program therefore was the estimated coefficient $\hat{\beta}_1$. The term $Other_i$ is an indicator variable equal to 1 if student i attended other academic COVID-19 recovery interventions during the summer, such as tutoring. Three of the eight districts provided tutoring sessions during the summer, mostly at summer school sites. In those districts, we considered the receipt of tutoring and summer school as an additional intervention, distinct from summer school alone or tutoring alone. The vector $\mathbf{X}_{it}$ includes students' prior achievement, demographics, and missing-data indicators. Specifically, we included student i's baseline MAP Growth math achievement in term $t-1$ as a cubic polynomial, where $t-1$ is spring 2022. We additionally include prior MAP growth math test scores from winter 2022 and fall 2021 as well as spring 2022 MAP growth reading test scores. We interacted all MAP Growth scores with a categorical variable (i.e., *MAP_missflag*) flagging possible combinations of missingness within each subject.[16]

When available, we also include in $\mathbf{X}_i$ any prior tests, such as state standardized tests, the district used to identify students to prioritize for summer programs. In all specifications, $\mathbf{X}_i$ additionally included all the following available pretreatment student covariates: a student's race/ethnicity, gender, Individualized Education Program status, English language learner status, 504 Plan status, and economic disadvantage status as of term $t-1$. Finally, we included in $\mathbf{X}_i$ indicators for the calendar weeks during which student i took each math MAP Growth test in terms t and $t-1$, respectively. The entire vector was interacted by grade level to account for across-grade differences in the relationship of covariates, treatment, and the outcome. $Sch\_Grade_{ij,t-1}$ represents school-by-grade fixed effects based on the school and grade of student i in spring 2022. ε_{ijt} denotes idiosyncratic error. We estimated a linear model and clustered the standard error at the school-by-grade level (Abadie et al., 2022). When reading achievement was the outcome, we reversed the reading and math subscripts in Equation (1).

We adapted the preceding general specification based on each district's available data and summer school program design. For districts that used data and/or created decision rules that combined multiple data sources to target or prioritize students for participation (e.g., scores below a certain threshold on MAP Growth assessments, state tests, and/or other academic assessments), we controlled for a binary measure of prioritization in addition to any other indicators or test scores (in cubic form) related to prioritization. As with the MAP Growth scores described earlier, we imputed missing values of these additional achievement measures and interacted them with imputation flags and expected rising grade level.

The primary coefficient of interest was β_1, the coefficient on the summer school treatment. Interpreting β_1 as a causal relationship relies on the assumption that assignment to treatment (i.e., participation in summer school) is as good as random, conditional on the covariates included in our model. Our models were aligned with value-added approaches that had been validated as producing causal estimates for the effects of teachers and schools (e.g., the inclusion of multiple prior test scores; see, e.g., Abdulkadiroglu et al., 2011; Bacher-Hicks et al., 2019; Chetty et al., 2014; Deming, 2014); however, the assumption for causality is likely stronger in the context of evaluating an intervention such as summer school.[17] Indeed, our selection-on-observables research design left room for potential selection bias if participation was correlated with unobservable factors that were also associated with subsequent achievement. The direction of this potential bias was difficult to predict a priori. For example, it was possible that students with more motivation or greater family resources (e.g., an adult available to get them to school) were more likely to attend summer school than observably similar students who did not, positively biasing our VAM estimates. In contrast, families with fewer resources may have been more likely to enroll their students in the district's summer program because it was less costly than other summer childcare alternatives, negatively biasing the estimates.

We cannot say with certainty whether the observable covariates we included in our models adequately controlled for these selection issues. To test whether the inclusion of covariates affected our VAM estimates, we estimated versions of the models that gradually included the complete set of covariates, starting with (a) school-by-grade fixed effects only, (b) adding baseline test scores, (c) adding demographic and other student characteristics, and (d) adding any additional prior test scores and relevant eligibility indicators (see Appendix Table A1 in the online version of this paper). Our findings from these supplemental models consistently suggested that a naive model without covariates would result in a sizable negative selection bias, which was gradually reduced as covariates were added. While this pattern of reducing selection bias as we approach the full specification of our model is promising, we recognize that the potential for selection bias remains. As such, while we use causal language throughout this paper for simplicity (e.g., "treatment effects" and "value-added effects"), and we acknowledge that our estimates may in part reflect unobservable differences between treated and control students.

In addition to estimating the effect of summer school participation, we estimated the effect per hour of math (or reading) instruction during summer school on the fall 2022 MAP Growth scores. We first calculated the average hours of instruction received in each program using the daily attendance data from the district and information from district interviews on the intended hours of subject-level instruction per day. To estimate the effect of the program per hour, we then divided the average treatment effect by the average hours of instruction received by participating students in the sample. These estimates allowed us to make comparisons of each program's effectiveness without conflating the average treatment effect of the program with its duration and attendance rate, assuming that dosage is one of many factors (e.g., curriculum and teacher quality) that can influence the effectiveness of summer programming (see Schwartz et al., 2018).

Finally, we synthesized our estimates of summer program effects by conducting a meta-analysis across the eight districts. For each subject, we made meta-analytic estimates with a random-effects model with restricted maximum likelihood estimation (DerSimonian & Laird 1986; Hedges, 1983; Raudenbush, 2009).

Putting Our Estimates in Context

To help contextualize our findings, we compared our estimated effects with those in the prepandemic summer school literature. More specifically, we compared our estimates to what we would expect to see given the observed dosage (in instructional hours) of summer school that students received (i.e., attended) and benchmark effect sizes from the existing literature on summer school's academic impacts. Neither the Kim and Quinn (2013) meta-analysis of summer programs' impacts on reading scores nor the Lynch et al. (2023) meta-analysis of summer programs' impacts on math scores included the measures of student attendance necessary to calculate the summer school dosage that students actually received. Therefore, we relied on the McCombs et al. (2014) study of the first year of implementation of the summer programs participating in the RAND RCT study to estimate the expected effect per subject-specific instructional hour of summer programming *attended* on test scores the following fall. The study estimated a significant average effect of 0.11 SD of attending at least 1 day on math scores but did not detect an effect on reading, so we estimated the *expected* effect per hour of summer instruction only for math. Using the information about the programs and student attendance provided in the study, we found that students who attended at least 1 day of programming received an average of 21.9 hours of math instruction and 32.8 hours of reading instruction over the summer.[18] Assuming a linear relationship between the instructional time students received in math during summer school and the average effect[19] of the program, we backed out a 0.005 SD (i.e., 0.11 SD/21.9) hourly expected effect of summer school instruction in math, which we then multiplied by the average dosage of instruction received in each subject in each of the R2R district programs to arrive at the "expected effect" of summer school participation in each district. Although evidence supports a positive relationship between instructional time and achievement (Kraft & Novicoff, 2024), we note that our expected effect estimates should not be interpreted as a precise measure because they are based on our assumption of linearity, which may not hold true in practice.

Results

Program Design

The features of each summer program are displayed in Table 3. All programs served students in rising grades 1–8, and some additionally served students in earlier or later grades. Six of the eight districts targeted their program toward students based on low test scores and/or other risk factors by sending invitations to those students and families. Only one district limited participation to targeted students; the other seven districts allowed—and often encouraged—all students from eligible grades to participate.[20] All districts selected a subset of schools to operate as summer school sites (dubbed *hub sites*) that served students from two or more neighboring schools.

Each of the programs provided in-person instruction in both math and English language arts and recruited teachers from their own district to serve as classroom teachers. When possible, program leaders placed teachers at their regular school (or the hub site assigned to their regular school). All districts provided some professional development for teachers in advance of the program. All programs were designed to deliver grade-level instruction (based on a student's grade as of 2021–22) to address gaps in students' learning from the prior year. Finally, all programs included some time for enrichment activities (e.g., physical education, art, music, dance, and gardening), which were led by school staff and/or community partners.

Program durations ranged from 12 to 20 instructional days, with anywhere between 45 minutes and 2 hours of daily academic instructional time per subject (i.e., math and reading). Six districts convened 4 days per week. The other two districts offered programming 5 days per week. Five of the eight programs ran an abbreviated day, from 8 or 9 a.m. to 12 or 1 p.m. Two programs ran for a full day, and one program had hours that extended beyond a typical school day. The total hours of academic instruction offered in the programs ranged from 23 to 67 hours and was generally allocated equally to math and reading.

Overall, the planned instructional time we observed in our sample was lower than the recommended minimum of 75 hours of academic instruction for summer programs (Schwartz et al., 2018) and was less than that of the summer programs evaluated in the prior literature. For instance, the five district-led summer programs evaluated by the team at RAND (Augustine et al., 2016; McCombs et al., 2014) offered between 23 and 29 days of instruction. Similarly, the average length of summer programs reviewed by Lynch et al. (2023) was ~5.2 weeks, or 26 days (assuming 5 days per week of programming).

In addition to regular instruction, three districts (districts 2, 4, and 6) offered tutoring in math and/or reading to a subset of lower-performing students during their summer programs. District 6 designed its tutoring program with the goal of delivering two to three 30-minute tutoring sessions per week (i.e., 3–7 total hours over the course of the program) while nontargeted students were participating in enrichment activities. Alternatively, students in districts 2 and 4 were pulled out of their academic summer program classes to receive tutoring. In these cases, tutoring substituted for other academic time and did not increase the overall dose of instruction.

Table 4
Summer programs participation and dosage

District	Sample rising grade level	Sample size	% Treated: Overall	% Treated: Hub sites	% Treated: Not hub	Intended program length (d)	No. of days attended (average)	% Total days attended	Total hours of instruction per student: Math	Total hours of instruction per student: Reading
1	3–8	11,841	22.6%	28.1%	20.3%	19	12.5	65.7%	18.7	18.7
2	3–8	39,248	15.0%	19.8%	11.3%	15	10.8	72.1%	16.2	16.2
3	3–8	5,364	10.2%	9.1%	10.8%	15	9.8	65.0%	14.6	14.6
4	3–8	14,689	15.1%	17.7%	14.6%	20	11.6	58.0%	23.2	23.2
5	1–8	1,804	13.6%	17.4%	14.8%	15	9.9	66.1%	14.9	14.9
6	1–8 (math) 4–8 (reading)	33,504	12.1%	13.4%	10.5%	12–18	10.9	65.2%	20.2	21.0
7	4–8	13,748	8.5%	14.3%	5.8%	20	14.2	70.9%	14.2	14.2
8	6–8	9,371	4.8%	5.9%	4.5%	17	13.6	80.0%	27.2	27.2

Note. Students included in the analytic sample were expected to be entering grades 1–8 in fall 2022 and had NWEA MAP Growth scores for both the spring 2022 and fall 2022 tests in reading or math. The percent treated at hub sites shows the participation rate among students who were enrolled at a school that served as a summer school site. The percentage of total days attended is equal to the total number of days attended by students who attended any summer school divided by the intended program length (in days).

Participation and Dosage

The percentage of students who attended at least 1 day of summer programming for each of the eight R2R districts in the sample is shown in Table 4.[21] Participation rates varied substantially across the districts, ranging from 4.8 to 22.6%. The average participation rate across the eight districts was ~12.7%, similar to national figures from the 2022 Household Pulse Survey, which found that roughly 10% of households with children under the age of 18 years enrolled them in summer learning programs.[22] Across the R2R districts, average participation rates were higher among students whose schools were host (hub) sites for the summer program (15.7%) relative to students whose schools were not hub sites (11.6%).

For the six districts that targeted enrollment, the targeted students comprised 28–55% of the districts' populations, and participation rates among targeted students were higher than among nontargeted students (see Appendix Table A2 in the online version of this paper). In district 7, enrollment was open only to prioritized students, and the participation data confirmed that all participants were in this group. The higher participation rates of targeted students aligned with our finding that students who attended summer school in districts that used targeting tended to score substantially lower on MAP Growth tests administered prior to the summer (see Appendix Table A2 in the online version of this paper). Participation rates in these districts also were higher for students in disadvantaged subgroups, including students receiving special education services, students who were English language

Table 5
Summer programs with tutoring participation and dosage

District	Sample rising grade level	Sample size	% Attended summer school	% Summer school attendees tutored			Intended program length (d)	No. of days attended (average) by tutees	% Total days attended by tutees	Total hours of summer instruction per tutee	
				Overall	Math	Reading				Math	Reading
2	3–8	39,248	15.0%	59.7%	45.4%	43.8%	15	12	77.5%	21.9	23.1
4	3–8	14,689	15.1%	51.5%	42.4%	36.8%	20	12	58.9%	23.4	23.2
6	1–8 (math) 4–8 (reading)	33,504	12.1%	34.2%	19.8%	25.0%	12–18	12	71.2%	26.8	27.5

Note. Tutees refer to students who received tutoring during summer school. The percentage of total days attended by tutees is equal to the total number of days of summer school attended by tutees who attended any summer school divided by the intended program length (in days). District 6 provided tutoring as a supplement to core instruction in the summer program, whereas students were pulled out of core instruction to receive tutoring in districts 2 and 4. To make reasonable comparisons of instructional time provided to tutees across districts, we report the total instructional hours students received in each program, but we note that that the instructional time tutees received in district 6 included 23.3 hours of core instruction and 3.5 hours of tutoring time in math and 24.5 hours of core instruction and 3.0 hours of tutoring time in reading, on average.

learners, students who were economically disadvantaged, and students who were Black or Latino/a. In the two districts (districts 1 and 5) that did not use targeting, spring MAP Growth scores were more similar across students who participated and those who did not. See Appendix Table A2 in the online version of this paper for a breakdown of participation rates for all student subgroups.

Relative to the other districts in our sample, district 1 had a notably higher rate of participation. According to the district leaders, district 1's extended summer program hours helped to drive its relatively high participation rate (see Table 3). The program operated from 8 a.m. to 5:30 p.m. for elementary school grades and from 9 a.m. to 4 p.m. for secondary school grades, plausibly providing childcare for working parents. In this district (and others), district leaders said that framing their summer programs as summer camp—an exciting learning and enrichment program during the summer—as opposed to summer school may have helped persuade students and families to sign up for the program.

As shown in Table 5, far fewer students received tutoring during summer school than attended summer school without extra tutoring. Across the three districts that offered tutoring, the percentage of summer school attendees participating in math or reading tutoring ranged from 34.2 to 59.7%. The total hours of instruction per tutee ranged from 21.9 to 26.8 hours in math and from 23.1 to 27.5 hours in reading.[23]

Table 4 shows the proportion of days students attended relative to days the program was offered. Attendance varied from 58 to 80%, with an average of 67.9%. This rate is slightly below the 75% attendance rate Schwartz et al.

(2018) advised districts to expect when planning their programs. Across the districts, the average number of days attended ranged from 9.8 to 14.5 days, and the average total hours of instructional time students received per subject ranged from 14.2 to 27.2 hours.[24] In the three districts that offered summer tutoring, tutored students attended a slightly higher proportion of days (~4 percentage points) and received slightly more hours of instruction (~4 hours more in math and reading) than their peers in the summer program who were not receiving tutoring (see Table 5).

Overall, these descriptive results highlight two findings: (a) there was substantial variation in program duration, enrollment, and attendance across districts, and (b) the planned durations (the total days and hours of instruction) of programs in our sample were considerably shorter than both the durations of programs studied in prior research and earlier benchmark recommendations.

Academic Effects of Summer School

Figures 1 and 2 (reported numerically in Tables 6 and 7) show the estimated effects of attending at least 1 day of summer school on fall MAP Growth 2022 math and reading scores, respectively. For comparison, the figures also include the expected effect for each program, based on the average dosage received in each district and prior research. We focused on the effect of attending at least 1 day of summer school because it was difficult to account for selection biases that could lead to differential uptake of the treatment in observational data.[25] At the top of both figures, the *overall* estimate reflects results from a cross-district meta-analysis using a restricted maximum likelihood model. While the analytic samples and model specifications vary somewhat across districts because of summer program design and data availability, a data limitation in one district is worth noting. In district 8, representative MAP Growth scores were available only in middle school grades, limiting our analysis to those grades, even though elementary students also took part in the summer program. While the estimates across districts varied, besides District 8, all the estimates for math test scores were positive, and four of the eight districts' estimates were statistically significant, ranging between 0.027 and 0.044 SD.[26] The estimated effect in each of these four districts was larger than the aggregated (across-districts) point estimate of 0.027 SD, which itself was just under one third of the aggregated effect size we would expect for these programs based on their durations and prior research. As displayed in Figure 3 (and reported numerically in Table 6), the districts with significant positive effects in math also had more comparable hourly effects (i.e., districts with larger effects generally also had higher dosage), around 0.001 to 0.002 SD. All the districts' estimated hourly effects were smaller than the expected hourly effect (0.005 SD).

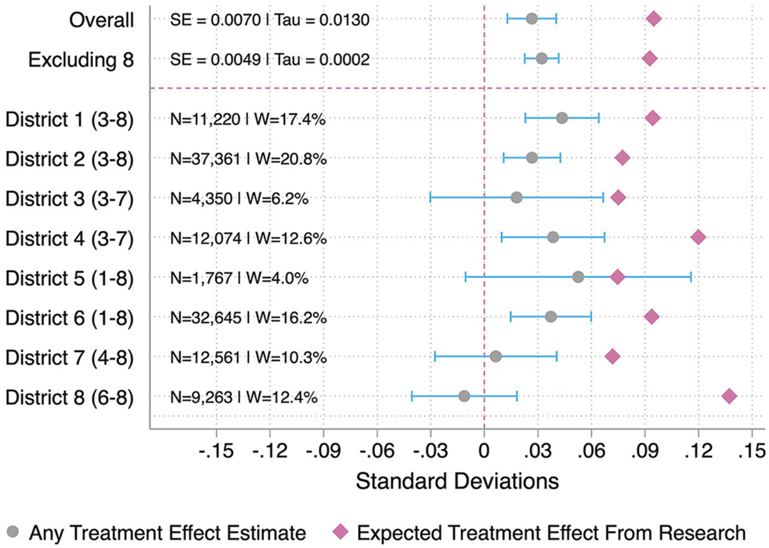

Figure 1. **Effects of attending summer programs on fall MAP Growth test, math.**

Note. The overall estimates were constructed using a random-effects model with a restricted maximum likelihood estimation. n indicates the number of students in the analysis. W indicates the weight given to each district analysis. SE indicates the standard error, which comes from each estimate's weighted errors. Tau indicates the error arising from interestimate heterogeneity. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in math. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in math. Blue bars represent the 95% confidence interval of our estimate. The expected treatment effect from research was based on a 0.11 SD increase in math achievement associated with attending an average of 21.9 hours of math instruction in a summer program (McCombs et al., 2014).

The findings for reading (displayed in Figure 2) are less promising. The estimated effect of attending at least 1 day of summer school was positive and statistically significant only in district 6 (0.029 SD) and district 7 (0.037 SD). In the remaining districts, the effects were statistically indistinguishable from zero. The estimates of hourly effects of summer school on reading, shown in Figure 4 (also reported numerically in Table 7), also were small, between −0.001 and 0.001 SD, for all districts except districts 6 (0.002 SD)

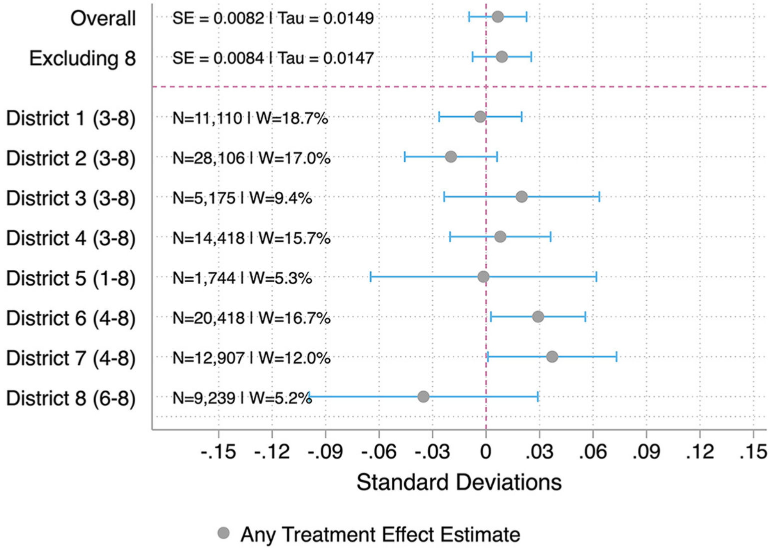

Figure 2. **Effects of attending summer programs on fall MAP Growth test, reading.**
Note. The overall estimates were constructed using a random-effects model with a restricted maximum likelihood estimation. n indicates the number of students in the analysis. W indicates the weight given to each district analysis. SE indicates the standard error, which comes from each estimate's weighted errors. Tau indicates the error arising from interestimate heterogeneity. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in reading. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in reading. Blue bars represent the 95% confidence interval of our estimate.

and 7 (0.003 SD). While overall these results were smaller in magnitude to what we would expect based on the meta-analysis by Kim and Quinn (2013), they were consistent with the findings from the multisite RCT of district-led summer learning programs conducted by the RAND Corporation (Augustine et al., 2016), which found improvements to math achievement following summer school but not to reading.

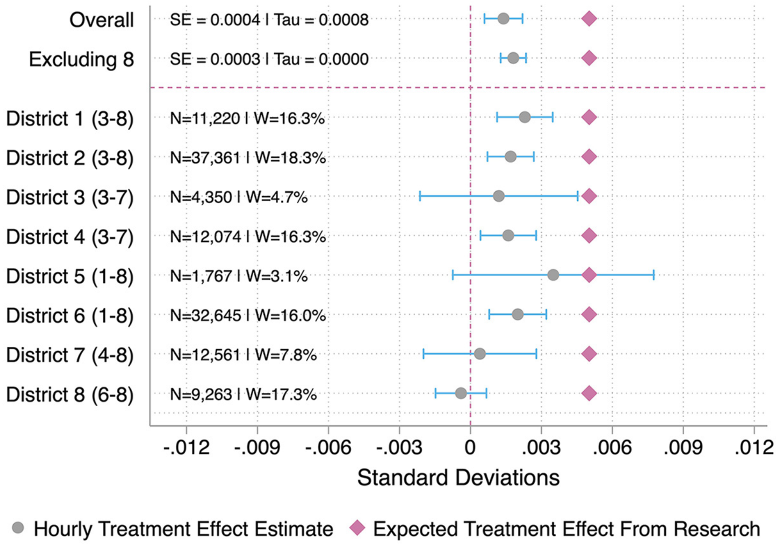

Figure 3. **Effects per summer program instructional hour on fall MAP Growth test, math.**

Note. The overall estimates were constructed using a random-effects model with a restricted maximum likelihood estimation. n indicates the number of students in the analysis. W indicates the weight given to each district analysis. SE indicates the standard error, which comes from each estimate's weighted errors. Tau indicates the error arising from interestimate heterogeneity. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in math. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in math. Blue bars represent the 95% confidence interval of our estimate. The expected treatment effect from research is based on a 0.11 SD increase in math achievement associated with attending an average of 21.9 hours of math instruction in a summer program (McCombs et al., 2014).

Subgroup Analyses

In addition to estimating the effects of summer programming across all grade levels included in our analysis, we examined effects by elementary and middle school grade ranges. The results of these grade-range analyses appear in Tables 8 and 9 for math and reading, respectively. When broken out by grade range, we found suggestive evidence that younger

Table 6
Effects of attending summer programs on fall MAP Growth test, math

District (rising grades)	Estimate: Any participation (SE)	Estimate: Hourly (SE)	Average Dosage (d)	Average dosage (h)	Expected effect from research	% Treated	n
Overall	0.0266* (0.0070)	0.0014* (0.0004)	11.59	18.30	0.0919	10.06%	121,241
Overall (omitting district 8)	0.0321* (0.0049)	0.0018* (0.0003)	11.51	17.97	0.0902	10.51%	111,978
1 (3–8)	0.0435* (0.0105)	0.0023* (0.0006)	12.52	18.78	0.0943	23.03%	11,220
2 (3–8)	0.0267* (0.0081)	0.0017* (0.0005)	10.84	15.42	0.0774	8.19%	37,361
3 (3–7)	0.0181 (0.0247)	0.0012 (0.0017)	9.95	14.93	0.0750	10.64%	4,350
4 (3–7)	0.0385* (0.0147)	0.0016* (0.0006)	11.86	23.89	0.1199	9.57%	12,074
5 (1–8)	0.0526 (0.0322)	0.0035 (0.00217)	9.91	14.86	0.0746	13.75%	1,767
6 (1–8)	0.0373* (0.0115)	0.0019* (0.0006)	10.59	19.56	0.0978	9.70%	32,645
7 (4–8)	0.0064 (0.0174)	0.0004 (0.0012)	14.30	14.30	0.0715	8.73%	12,561
8 (6–8)	−0.0112 (0.015)	−0.0004 (0.0005)	13.66	27.32	0.1366	4.64%	9,263

Note. The overall estimate refers to the meta-analytic estimates of the eight coefficients. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in math. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in math.
*$p < .01$.

students—those in rising grade 5 and below—drove the positive effects in math, although most differences were not statistically significant. The program length and attendance rates among participants were similar across grade ranges within each district, suggesting that any differences in effect sizes across grades did not result from different amounts of dosage. Nevertheless, we are hesitant about overinterpreting any larger effects in earlier grades because research has shown that average standardized test score gains per year are larger in elementary school than in middle school (Cascio &

Table 7
Effects of Attending Summer Programs on Fall MAP Growth Test, Reading

District (rising grades)	Estimate: Any participation (SE)	Estimate: Hourly (SE)	Average Dosage (d)	Average Dosage (h)	% Treated	n
Overall	0.0066	0.0005	11.60	18.05	9.95%	103,117
	(0.0082)	(0.0005)				
Overall (omitting district 8)	0.0089	0.0003	11.51	17.64	10.46%	93,878
	(0.0084)	(0.0005)				
1 (3–8)	−0.0032	−0.0002	12.51	18.76	22.39%	11,110
	(0.0118)	(0.0006)				
2 (3–8)	−0.0197	−0.0013	10.61	14.88	8.57%	28,106
	(0.0132)	(0.0009)				
3 (3–8)	0.0200	0.0014	9.76	14.65	9.97%	5,175
	(0.0222)	(0.0015)				
4 (3–8)	0.0080	0.0003	11.65	23.23	9.56%	14,418
	(0.0144)	(0.0006)				
5 (1–8)	−0.0015	−0.0001	9.95	14.93	13.70%	1,744
	(0.0323)	(0.0022)				
6 (4–8)	0.0292*	0.0015*	10.2	18.86	8.32%	20,418
	(0.0135)	(0.0007)				
7 (4–8)	0.0371*	0.0026*	14.22	14.22	8.45%	12,907
	(0.0184)	(0.0013)				
8 (6–8)	−0.0351	−0.0013	13.59	27.19	4.73%	9,239
	(0.0327)	(0.0012)				

Note. The overall estimate refers to the meta-analytic estimates of the eight coefficients. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in reading. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in reading.
$^{*}p < .05$.

Staiger, 2012; Lipsey et al., 2012). Smaller effect sizes for middle school students may be equivalent to larger effect sizes in elementary schools in terms of the equivalent weeks of learning (Lewis & Kuhfeld, 2023).

We also estimated the effects of summer school by different student subgroups, including by race and ethnicity, free or reduced-price lunch (FRPL) status, special education status, English learner status, prior performance levels, and eligibility for targeting for enrollment in summer school (see Appendix Tables A3–A9 in the online version of this paper).[27] While we

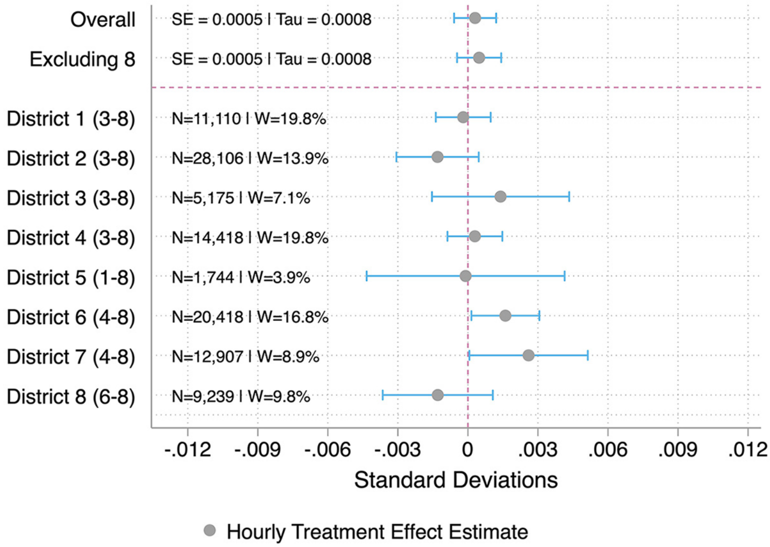

Figure 4. **Effects per summer program instructional hour on fall MAP growth test score, reading.**

Note. The overall estimates were constructed using a random-effects model with a restricted maximum likelihood estimation. n indicates the number of students in the analysis. W indicates the weight given to each district analysis. SE indicates the standard error, which comes from each estimate's weighted errors. Tau indicates the error arising from interestimate heterogeneity. The second overall estimate excluded district 8 because the district's analysis sample included only middle school grades with low participation rates, which tended to have lower effects overall. The analytic sample included all students entering grades 1–8 in fall 2022 who had a valid spring 2022 and fall 2022 MAP Growth test score in reading. For district 7, where spring 2022 MAP Growth testing was very limited, the analytic sample included students who had a nonmissing winter 2022 MAP Growth score and spring 2022 state standardized test score in addition to the fall 2022 MAP Growth score in reading. Blue bars represent the 95% confidence interval of our estimate.

caution that subgroup sample sizes often limited our ability to detect differences between groups, across all districts, we found little consistent evidence of heterogeneous effects across these subgroups. Comparing results within and between districts, the evidence was mixed on whether more or less advantaged students were the primary beneficiaries of summer school. For example, in district 1, we found positive and significant effects on math for students above and below the 50th percentile on baseline MAP Growth scores and for both FRPL- and non-FRPL-eligible students; at the same time, White students were the only racial subgroup for whom we found significant

Table 8
Effects of summer programs on fall MAP growth test score by grade bands, math

	(A) Rising grades 1–5					(B) Rising grades 6–8					
District	Grades in sample	Any participation (SE)	Average dosage (h)	% Treated	n	Grades in sample	Any participation (SE)	Average dosage (h)	% Treated	n	Difference, (A) – (B)
1	3–5	0.0539** (0.0149)	19.09	27.37%	6,368	6–8	0.0247* (0.0106)	18.14	17.33%	4,852	0.0292
2	3–5	0.0387** (0.0112)	15.65	8.85%	19,265	6–8	0.0113 (0.0117)	15.13	7.49%	18,096	0.0274
3	3–5	0.0031 (0.0316)	15.20	10.58%	2,958	6–8	0.0518 (0.0362)	14.35	10.78%	1,392	−0.0487
4	3–5	0.0590** (0.0210)	24.79	10.28%	7,354	6–8	−0.0040 (0.0150)	22.20	8.47%	4,720	0.0630*
5	1–5	0.0527 (0.0429)	17.02	16.54%	1,028	6–8	n/a	9.84	9.88%	739	n/a
6	1–5	0.0511** (0.0147)	21.68	11.40%	20,525	6–8	0.0064 (0.0172)	13.54	6.81%	12,120	0.0447*
7	4–5	0.0116 (0.0251)	14.16	11.85%	5,393	6–8	0.00032 (0.0241)	14.51	6.38%	7,168	0.0113
8	n/a					6–8	−0.0112 (0.0150)	27.32	4.64%	9,263	n/a

Note. Value-added estimates are reported for a subgroup only when the analytic sample for the subgroup was >1,000 and the number of treated students within the subgroup was >100.
*$p < .05$; **$p < .01$.

Table 9
Effects of summer programs on fall MAP Growth test score by grade bands, reading

	(A) Rising grades 1–5					(B) Rising grades 6–8					
District	Grades in sample	Any participation (SE)	Average dosage (h)	% Treated	n	Grades in sample	Any participation (SE)	Average dosage (h)	% Treated	n	Difference, (A) – (B)
1	3–5	−0.0005 (0.0143)	19.00	26.69%	6,318	6–8	−0.0076 (0.0210)	18.25	16.74%	4,792	0.0071
2	3–5	−0.0118 (0.0161)	14.72	8.47%	12,015	6–8	−0.0266 (0.0204)	15.00	8.45%	16,091	0.0148
3	3–5	0.0004 (0.0281)	15.25	10.45%	2,939	6–8	0.0508 (0.0383)	13.76	9.35%	2,236	−0.0504
4	3–5	−0.0075 (0.0205)	25.05	11.80%	7,315	6–8	0.0321 (0.0170)	20.19	7.25%	7,103	−0.0396
5	1–5	n/a	17.08	16.88%	995	6–8	n/a	9.85	9.48%	749	n/a
6	4–5	0.0269 (0.0213)	20.89	10.04%	8,457	6–8	0.0308 (0.0171)	16.83	7.11%	11,961	0.0039
7	4–5	0.0310 (0.0259)	14.15	11.75%	5,419	6–8	0.0444 (0.0260)	14.34	6.05%	7,488	−0.0134
8	n/a	—	—	—	—	6–8	−0.0351 (0.0327)	27.19	4.73%	9,239	n/a

Note. Value-added estimates are reported for a subgroup only when the analytic sample for the subgroup was >1,000 and the number of treated students within the subgroup was >100.

positive effects. Meanwhile, in district 4, we found positive and significant effects on math for students below the 50th percentile on baseline MAP Growth scores and for FRPL-eligible students but not for their counterparts; we also saw positive and significant effects for Latino/a students but not for any other racial subgroup. Whether participation was concentrated among certain subgroups (e.g., district 6) or relatively consistent across all groups (e.g., district 1) also varied between districts. Collectively, the lack of clear patterns across all the districts in our sample suggests that summer school benefits are not limited to participants in either advantaged or disadvantaged student subgroups.

Academic Effects of Summer School with Tutoring

Figure 5 (reported numerically in Tables 10 and 11) shows the estimated effects in math and reading of attending any amount of summer school that replaced or supplemented classroom instruction with tutoring in math, reading, or both. Our VAM estimates of the effects of these programs are similar to the effects we estimated for receiving the general summer learning program in each of the three districts that offered summer school with tutoring. Among these districts, districts 2 and 4 (where tutoring time typically replaced classroom instruction) had positive and statistically significant effects of the combined program with math tutoring on math achievement (0.038 SD in district 2 and 0.047 SD in district 4). The point estimate for district 6 (where tutoring time supplemented classroom instruction and provided an additional 3.5 hours of math instruction and 3.0 hours of reading instruction, on average) was of similar magnitude (0.046 SD), although not statistically significant. In reading, none of the estimated effects of combined summer programming and reading tutoring were statistically distinguishable from zero, and the magnitudes were considerably smaller (or negative) relative to math. Because fewer students participated in combined summer school with tutoring than in summer school alone, our statistical power for this analysis was limited, particularly in district 6, where only 4.1% of the student population in the relevant grade levels participated in any tutoring. Overall, the estimated average effects of our sample's standard summer learning programs on math (0.027 SD) and reading (no effect) achievement are statistically equivalent to those of our sample's programs that combined summer school with tutoring.

Discussion and Conclusion

Across the eight districts in this study, we found that the magnitude (e.g., size relative to district enrollment, number of days, and hours per day) of summer programs and take-up rates varied substantially. Students who attended summer programs were found to have made small but significant improvements in math achievement but no significant improvements on reading tests

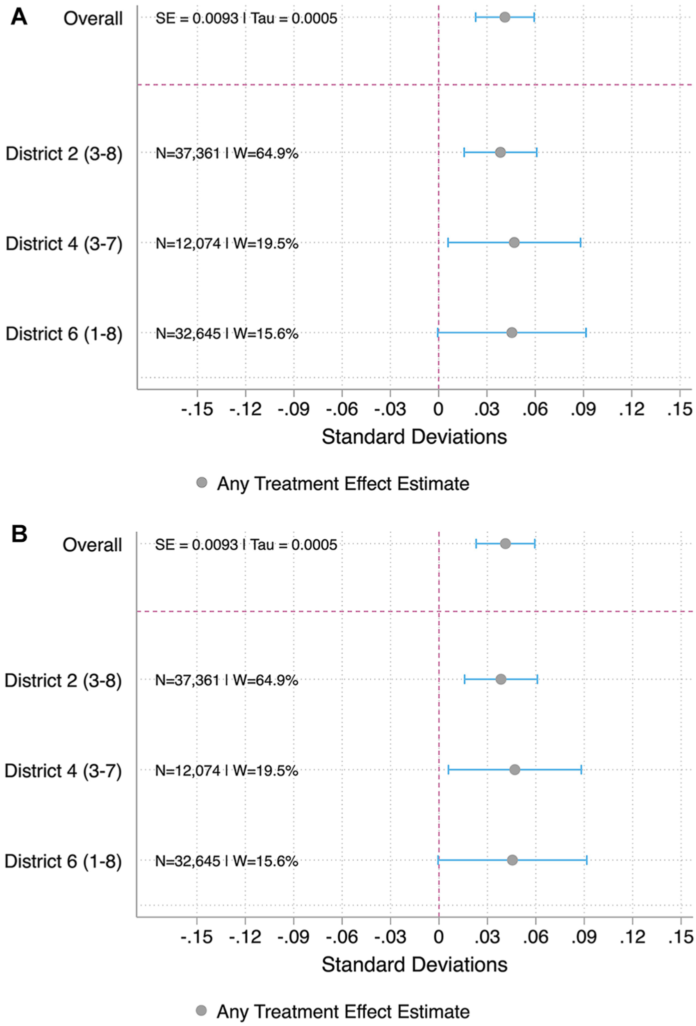

Figure 5. **Effects of attending summer programs with tutoring on fall MAP growth tests: (A) math; (B) reading.**

Note. The two parts of this figure show the effects of attending at least 1 day of a summer program with tutoring on math and reading scores, respectively. The overall estimates were constructed using a random-effects model with a restricted maximum likelihood estimation. n indicates the number of students in the analysis. W indicates the weight given to each district analysis. SE indicates the standard error, which comes from each estimate's weighted errors. Tau indicates the error arising from interestimate heterogeneity. Blue bars represent the 95% confidence interval of our estimate.

Table 10
Effects of summer programs and tutoring on fall MAP Growth test score, math

District (rising grades)	Any participation (SE)	Average tutoring dosage (h)	Average summer program dosage (h)	% Treated	n
Overall	0.0412** (0.0093)	18.73	6.88	4.99%	82,080
2 (3–8)	0.0384** (0.0115)	22.01	3.92	6.63%	37,361
4 (3–7)	0.0470* (0.0210)	23.46	0.00	6.84%	12,074
6 (1–8)	0.0455 (0.0235)	3.51	23.33	2.42%	32,645

*$p < .05$; **$p < .01$.

Table 11
Effects of summer programs and tutoring on fall MAP test score, reading

District (rising grades)	Any participation (SE)	Average tutoring dosage (h)	Average summer program dosage (h)	% Treated	n
Overall	−0.0089 (0.0151)	21.03	5.42	5.09%	62,942
2 (3–8)	−0.0180 (0.0210)	24.00	4.10	6.00%	28,106
4 (3–7)	−0.0197 (0.0302)	23.41	0.00	7.59%	14,418
6 (4–8)	0.0225	3.12	24.64	2.08%	20,418

relative to similar peers who did not attend. To be sure, summer learning programs may have goals beyond improving achievement (e.g., increasing student engagement, providing safe and accessible childcare, and providing meals). Examining broader (nontest) outcomes and the impacts of various program features (e.g., dosage, curriculum, and teacher quality) are important directions for future research but are beyond the scope of this study. In this study, we examined how summer learning programs in large districts affected academic achievement and the extent to which they could meaningfully address the learning losses caused by the COVID-19 pandemic.

For reading, we estimated null effects of attending a program in all but two districts and could rule out even small positive effects of at least 0.03

SD for most programs. The lack of reading effects, on average, was consistent with the null reading effects estimated in the RAND RCT study by Schwartz et al. (2018) but contrasts with the + 0.09 SD average effect of classroom-based summer reading interventions estimated in the meta-analysis by Kim and Quinn (2013). More broadly, the idea that summer school would have smaller effects on reading than math is consistent with research that suggested that school inputs have larger effects on math than reading (Jackson et al., 2014; Riehl & Welch, 2022). While there is no definitive explanation for the difference, researchers hypothesize that school has a greater impact on math learning because math instruction happens primarily at school, whereas reading and language development are more likely to happen both at school and at home. During the summer, it also may be the case that families are more likely to engage in activities at home that develop reading and language skills than math skills. If this were the case, the treatment–control contrast of participating in summer school would be smaller for reading than for math, potentially explaining some of our results. Future research should examine the experiences of students who were eligible or invited to attend summer school but did not take part in order to inform our interpretation of the relationship between summer programs and math and reading achievement.

Our analysis of math outcomes found that the effects per hour of academic programming were positive and remarkably similar across districts. The math effects were prevalent even in districts that served larger proportions of their students (up to 23% of their students in rising grades 3–8). The consistency in math effects across different student subgroups suggests that when summer program enrollment is limited, increasing the targeted recruitment and attendance of struggling students could be an effective strategy for boosting achievement for students with the greatest academic needs (we found that programs targeting struggling students enrolled greater proportions of low-performing and disadvantaged students than programs that encouraged all students to participate). That said, districts that targeted recruitment still served a relatively small proportion of these students, with only one in four targeted students attending summer school, on average. These attendance rates are consistent with the enrollment rates across summer programs found by Diliberti and Schwartz (2024) in summer 2023.

For the three districts whose summer programs provided tutoring to a subset of students, our VAM estimates suggest that the effect of receiving summer school *with* tutoring was not distinguishable from that of receiving standard summer school. This finding was surprising at first given the large effects of tutoring demonstrated in earlier research (Nickow et al., 2024). However, it was less surprising given the smaller (or null) impacts of pandemic-era tutoring programs to date (Carbonari et al., 2024a, 2024b; Kraft et al., 2024a). The results also may reflect the fact that tutoring sessions replaced classroom instruction in two districts and that tutored students in the other district received only 3.5 hours of math tutoring and 3.0 hours of reading tutoring, on average. We also know little from district interviews about the implementation of the tutoring that occurred during these programs and therefore are limited in our ability to speculate as to why it had limited additional benefits to achievement.

As ESSER funding lapses, district resources for summer programming likely will decline. Indeed, Diliberti and Schwartz (2024) found that nearly half of urban and suburban districts expect their funding for summer programs to decrease from summer 2023 to summer 2024. Districts facing reduced funds should consider replacing open-enrollment policies with targeted recruitment of lower-achieving students. Given that participation rates were higher among students whose schools hosted summer programming, assigning schools that serve higher proportions of targeted students as hub sites may be a promising strategy to boost targeted enrollment.

To put our overall findings into context, it is helpful to compare our estimated math effects with both the prepandemic literature and the scale of COVID-19 learning loss. First, the magnitude of the average effect of the R2R summer programs on math was only about one third of what we would expect based on the prepandemic literature (McCombs et al., 2014). One optimistic explanation for the smaller effects is that students who did not attend summer school during the pandemic experienced less summer slide than similar students prepandemic, perhaps because parents were more equipped to support learning at home after experiencing remote schooling. This explanation would be consistent with national data showing overall reductions in summer slide after the pandemic (Lewis & Kuhfeld, 2022). In fact, there is some evidence in our sample that supports this hypothesis. Increases in the average normed MAP Growth scores in both subjects from the spring to fall of 2022 for summer school participants and nonparticipants across all our districts indicate that *both* groups experienced less than typical summer slide (see Appendix A2 in the online version of this paper). In some districts, and especially for math, both participants and nonparticipants even made gains in terms of their raw MAP Growth test scores (i.e., RIT points) from spring to fall, suggesting that they did not experience the typical summer slide during summer 2023.[28] The summer gains among nonparticipants further suggest that increases and/or improvements in summer learning experiences outside of district-provided summer school could have reduced the treatment–control contrast, resulting in smaller estimated effects of participating in summer school.

However, less optimistic explanations for the reduced effects of pandemic-era programs related to program quality and implementation (e.g., differences in implementation related to curriculum, teacher quality, class size, etc.) align with evidence that suggests that many school-year academic recovery interventions are failing to have the desired (or anticipated) impacts on student achievement (Carbonari et al., 2024a; Kraft et al., 2024a; Robinson et al., 2025). Qualitative evidence suggests that schools found it difficult to implement evidence-based academic interventions during the pandemic for a host of reasons (Carbonari et al., 2024a). To combat these challenges and help students recover lost learning, districts may need to adjust and/or layer implementation strategies, such as incentivizing student and teacher attendance or increasing instructional time, to increase their impact.

To further contextualize the program effects against the scale of learning loss, a simple back-of-the-envelope calculation suggests that the math effects found in this study are small relative to the scale of the study districts' recovery needs. Specifically, multiplying the weighted average value-added effect of summer programs on *student* achievement (0.027 SD) by the average share of students attending them (~13%) yields a total effect of 0.0035 SD on *district*-level achievement (for the targeted grades). This is roughly 2% of the overall magnitude of pandemic-related learning loss in math in these districts and 3% of the average U.S. district's math losses as of spring 2022 (Reardon et al., 2023).[29] Even if districts facing average COVID-19-related learning losses of 0.2 SD in math delivered best-case scenarios in summer school programs (i.e., >5 weeks, with >2 hours of daily math instruction, or metrics that boost math gains by 0.10 SD), they would need to send *every student to summer school for 2–3 years in a row* to get back to prepandemic math test levels (Lynch et al., 2023).

These results may seem discouraging. But in the scope of COVID-19 recovery research, they stand out for having a significant effect on math achievement *and* for serving a substantial share of students in large districts—>2,000 students per district, on average. In contrast, recent studies of academic recovery programs during the 2022–23 school year (in many of the same districts) found no positive effects for any school-year recovery intervention serving >1% of a district's K–8 students (Carbonari et al., 2024b). Most of the large-scale tutoring programs implemented during the school year in these districts targeted a subset of grades within K–8, and none served >13% of the district's K–8 students (Carbonari et al., 2024b). Summer school's ability to impact math achievement at scale is notable given the finding of Kraft et al. (2024b) that tutoring program effect sizes estimated in RCTs substantially decline as program size increases. However, the tutoring RCTs that serve 1,000 or more students still have larger effect sizes (~0.16 SD, pooled across subjects) than our summer school estimates.

Overall, our findings suggest that summer school can be a valuable part of a district's academic recovery strategy for math, even if average effects on achievement are small. Expanding student participation in summer school and increasing the duration of summer programs are promising strategies for recovery. But districts undoubtedly will need other supports and interventions as well if they are going to make substantial progress toward recovery for all students. Evidence from this study and related work (Carbonari et al., 2024a, 2024b) underscores the need for continued commitment to delivering high-quality recovery interventions at the scale and intensity needed to address the pandemic's academic impact, whether in the summer or regular school year.

Acknowledgements

We could not have drafted this report without the district leaders who generously gave their time and attention to us during a challenging school year. We are very grateful for their help. We are also grateful to Erin Fahle and Lisa Sanbonmatsu for assistance with various aspects of this work.

Funding

This research was supported by the Carnegie Corporation of New York, the Walton Family Foundation, Kenneth C. Griffin, and an anonymous foundation. CALDER is funded by a consortium of foundations (for more information about CALDER funders, go to www.caldercenter.org/about-calder).

ORCID iDs

Ian Callen https://orcid.org/0009-0008-8924-0401
Dan Goldhaber https://orcid.org/0000-0003-4260-4040
Jazmin Isaacs https://orcid.org/0009-0006-8321-1306
Megan Kuhfeld https://orcid.org/0000-0002-2231-5228
Andrew McEachin https://orcid.org/0000-0002-5113-6616
Emily Morton https://orcid.org/0000-0003-3662-8556

Notes

[1]As we note in the "Methods" section, we attempted to minimize any selection bias in our estimates by using a selection-on-observables research design. Although we use causal language through this report (e.g., "impact"), we cannot rule out the possibility that our results are biased in an unknown direction.

[2]Analyses of changes in scores on the National Assessment of Educational Progress (NAEP) found that student achievement declined from 2019 to 2022 by 0.16 and 0.20 *SD* in grades 4 and 8 math, respectively (U.S. Department of Education, 2023).

[3]Fourteen of the studies (40%) employed a regression discontinuity design or leveraged a RCT to estimate program effects on reading outcomes (Kim & Quinn, 2013).

[4]Eleven of the studies (30%) employed a regression discontinuity design or leveraged a RCT to estimate program effects on math outcomes (Lynch et al., 2023).

[5]Note that enrollment rates represent the percentage of students who enrolled in a program. These rates provide an upper bound on the percentage of students who actually attend the program because some (or many) students may enroll in the program but never attend it (Augustine et al., 2016).

[6]These districts include Alexandria City Public Schools (Virginia), Dallas Independent School District (Texas), Guilford County Schools (North Carolina), Portland Public Schools (Oregon), Richardson Independent School District (Texas), Suffern Central School District (New York), Tulsa Public Schools (Oklahoma), and one district that asked to remain anonymous. For more about the Road to Recovery project, go to https://caldercenter.org/road-covid-recovery.

[7]In alignment with our agreements with each of the R2R districts, we protected districts' anonymity with respect to their results by masking district names and by being purposely ambiguous about the details of specific programs.

[8]These data are available through the Educational Opportunity Project at Stanford University and can be accessed at https://edopportunity.org/. Math and reading estimates for the Suffern Central School District and reading estimates for Portland Public Schools were not available at the time of this report.

[9]An exception to this rule is one district (district 7) with optional MAP growth testing and very low testing rates in spring 2022. For this district, we included spring 2022 state standardized test scores as a proxy for spring 2022 MAP growth scores, and we defined our analytic sample based on the combined availability of state standardized test scores along with winter 2022 and fall 2022 MAP growth scores. Our model specification for this district included cubic polynomials of both winter 2022 MAP growth and state standardized test scores. Less restrictive specifications that allowed for missingness in either of these variables yielded consistent results.

[10]This definition excluded some summer programs offered by R2R districts from our analysis because they were focused exclusively on enrichment activities, such as art, karate, or drama. We excluded these programs from our analysis because the districts did not consider these programs to be academic recovery programs with the explicit goal of improving student achievement.

[11]We did not collect detailed information about curriculum due to the limited duration and scope of our interviews.

[12]The interview notes template is available at https://caldercenter.org/road-covid-recovery.

[13]Schools typically administer fall tests between August and November, winter tests between December and mid-March, and spring tests between late March and June.

[14]For each district, we confirmed that atypical grade progression patterns were rare and that including or excluding these students in our sample did not affect our results. Of note, grade progression/retention was not contingent on summer school attendance for any district herein.

[15]$z(Y_{igst}) = (Y_{igt} - \bar{Y}_{gt})/SD(Y_{gt})$.

[16]This is commonly referred to as the *missingness-indicator method*. For a recent theoretical support of this method, see Zhao and Ding (2022).

[17]Note that ours is not the first study to use value-added models as an empirical strategy to estimate the effects of educational interventions. See, for example, prior studies of math textbooks (Blazar et al., 2020) and teacher training programs (Henry et al., 2014; Plecki et al., 2012).

[18]We first calculated the total hours of instruction received in each subject by multiplying the hours of math or reading offered per day by the total days attended by all students who attended at least 1 day for each program. The hours of instruction offered per day were not available for one of the five programs, so we substitute the average of the other four programs (1.21 hours of math and 1.75 hours of reading) for this program for our other calculations. To get the average hours of instruction received per student and subject, we divided the number of students who attended at least 1 day of programming by the total hours of math or reading instruction.

[19]We used the TOT sample (n = 2,518) and effect size (0.11 SD) to estimate the expected hourly effect of math instruction received on math scores, but using the ITT sample (n = 3,194) and effect size (0.09 SD) generated the same 0.005 SD hourly effect.

[20]One district that implemented extended-year calendars at a subset of schools excluded those schools from participating in summer school because their school year overlapped with the program. For this district, "all" students refers to students in schools without an extended-year calendar.

[21]It is important to note that students in some grade levels were omitted from the analytic sample due to sparse MAP growth score availability (which varied by district), even though they were offered summer programming. If we include these grades, participation rates in some districts were slightly higher than described in this study.

[22]We used data from the Week 49 Household Pulse Survey Public Use File release (https://www.census.gov/programs-surveys/household-pulse-survey/datasets.html) and limited the sample to respondents with children under age 18 years enrolled in public schools.

[23]Summer tutees are a subset of students in the analytic sample of each district, meaning that they are in the relevant grade levels for attending summer school and that they have valid spring and fall 2022 MAP Growth scores.

[24]Proportion of scheduled days students attended in district 4 was low, at 58%. The summer program was separated into two blocks, with a 1-week break in between the blocks, which may have affected the number of days students attended.

[25]In supplemental analyses, we estimated separate VAM estimates for students who were high attenders (>80% of days attended) and students who were low attenders of summer school. The observed relationship between dosage and estimated effects varied widely across districts; in some cases, effects were significantly greater for low attenders, whereas in others, they were significantly greater for high attenders. The high likelihood that attendance rates are endogenous and that these estimates reflect varying degrees of selection bias make these models difficult to interpret; to avoid overinterpretation, we do not include them in this paper. Results are available on request.

[26]We note that we may be statistically underpowered to detect significant effects for the programs in districts 3 and 5 because their sample sizes were notably small.

[27]We do not present subgroup results for district 5 due to sample size.

[28]See Appendix Table A2 in the online version of this paper for descriptive summaries, including average MAP Growth scores (normed by subject, term, and grade and in RIT points), of treated and untreated students in each district. For all the districts where we estimated significant and positive effects of summer school on either math or reading achievement, the average gains from spring to fall for treated students exceeded the average gains for untreated students in the subject where we detected a positive effect.

[29]Unfortunately, recent evidence suggests that academic recovery during the 2022–23 school year was very limited and that the learning losses estimated as of spring 2022 are largely consistent with those estimated as of spring 2023 (Lewis & Kuhfeld, 2023).

References

Abadie, A., Athey, S, Imbens, G. W., & Wooldridge, J. M. (2022). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Abdulkadiroglu, A., Angrist, J. D., Dynarski, S., Kane, T., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics*, *126*, 699–748. https://doi.org/10.1093/qje/qjr017

Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., & Siler-Evans, K. (2016). *Learning from summer: Effects of voluntary summer learning programs on low-income urban youth* (Report RR-1557-WF). RAND Corporation.

Bacher-Hicks, A., Chin, M., Kane, T. J., & Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, *73*, 101919. https://doi.org/10.1016/j.econedurev.2019.101919

Barry, S. S., & Sass, T. R. (2022). *The impact of a 2021 summer school program on student achievement*. Georgia Policy Labs. https://doi.org/10.57709/FAJ9-8597

Bell, S. R., & Carrillo, N. (2007) Characteristics of effective summer learning programs in practice. *New Directions for Youth Development*, *114*, 45–63. https://doi.org/10.1002/yd.212

Blazar, D., Heller, B., Kane, T. J., Polikoff, M., Staiger, D. O., Carrell, S., Goldhaber, D., Harris, D. N., Hitch, R., Holden, K. L., & Kurlaendar, M. (2020). Curriculum reform in the Common Core era: Evaluating elementary math textbooks across six U.S. states. *Journal of Policy Analysis and Management*, *39*(4), 966–1019. https://doi.org/10.1002/pam.22257

Boss, S., & Railsback, J. (2002). *Summer school programs: A look at the research, implications for practice, and program sampler*. Northwest Regional Educational Laboratory.

Callen, I., Goldhaber, D., Kane, T. J., McDonald, A., McEachin, A., & Morton, E. (2024). *Pandemic learning loss by student baseline achievement: Extent and sources of heterogeneity* (Working Paper No. 292-0224). Calder Foundation

Carbonari, M. V., Davison, M., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Hashim, A., Kane, T. J., McEachin, A., Morton, E., Muroga, A., Patterson, T., & Staiger, D. O. (2024a). The impact and implementation of academic interventions during COVID-19: Evidence from the road to recovery project. *AERA Open*. Advance online publication. https://doi.org/10.1177/23328584241281286

Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Kane, T. J., McDonald, A., McEachin, A., Morton, E., Muroga, A., Salazar, A., & Staiger, D. O. (2024b). *Impacts of academic recovery interventions on student achievement in 2022-23* (Working Paper No. 303-0724). Calder Foundation.

Cascio, E. U., & Staiger, D. O. (2012). *Knowledge, tests, and fadeout in educational interventions* (NBER Working Paper No. 18038). National Bureau of Economic Research. https://www.nber.org/papers/w18038

Chetty, R., Friedman, J. N., & Rockoff, J. E., (2014). Measuring the impacts of teachers: I. Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Cooper, H. (2001). *Summer school: Research-based recommendations for policymakers*. SERVE Policy Brief. http://files.eric.ed.gov/fulltext/ED456557.pdf

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227–268. https://doi.org/10.3102/00346543066003227

Curriculum Associates. (2020). *Understanding student needs: Early results from fall assessments.* https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-diagnostic-results-understanding-student-needs-paper-2020.pdf

Dee, T. S. (2024). Higher chronic absenteeism threatens academic recovery from the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *121*(3), e2312249121. https://doi.org/10.1073/pnas.2312249121

Deming, D. J. (2014). *Using school choice lotteries to test measures of school effectiveness* (NBER Working Paper No. 19803). National Bureau of Economic Research.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Diliberti, M. K., & Schwartz, H. L. (2024). *School was in session this summer, but less than half of eligible students enrolled: Selected findings from the fifth American school district panel survey.* RAND Corporation. https://doi.org/10.7249/RRA956-22

Diliberti, M. K., & Schwartz, H. L. (2022). *Districts continue to struggle with staffing, political polarization, and unfinished instruction: Selected findings from the fifth American School District Panel Survey*. RAND Corporation. https://doi.org/10.7249/RRA956-13

DiMarco, B., & Jordan, P. W. (2022). *How local educators plan to spend billions in federal COVID aid*. FutureEd Explainer. https://www.future-ed.org/local-covid-relief-spending/

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2021). *COVID-19 and education: The lingering effects of unfinished learning* (Vol. 27). McKinsey & Company.

Gambi, L., & De Witte, K. (2024). The resilience in test scores and the mitigating effect of summer schools one year after the COVID-19 school closures. *Labour Economics*. Advance online publication. https://doi.org/10.1016/j.labeco.2024.102549

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights*, *5*(3), 377–392. https://doi.org/10.1257/aeri.20220180

Harter, C. (2023, June 26). This summer LAUSD tackles learning loss in a fun-filled environment. *Los Angeles Daily News*. https://www.dailynews.com/2023/06/26/this-summer-lausd-tackles-learning-loss-in-a-fun-filled-environment/

Hedges, L. V. (1983). Combining independent estimators in research synthesis. *British Journal of Mathematical and Statistical Psychology*, *36*, 123–131. https://doi.org/10.1111/j.2044-8317.1983.tb00768.x

Henry, G., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education*, *65*(1), 7–23. https://doi.org/10.1177/0022487113503871

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), 801–825. https://doi.org/10.1146/annurev-economics-080213-040845

Kilbride, T., Hopkins, B., Strunk, K. O., & Imberman, S. (2021). *K–8 student achievement and achievement gaps on Michigan's 2020–21 benchmark and summative assessments* [Research brief.] Education Policy Innovation Collaborative.

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*(3), 386–431. https://doi.org/10.3102/0034654313483906

Kraft, M. A., & Novicoff, S. (2024). Time in school: A conceptual framework, synthesis of the causal research, and empirical exploration. *American Educational Research Journal*, *61*(4), 724–766. https://doi.org/10.3102/00028312241251857

Kraft, M. A., Edwards, D. S., & Cannata, M. (2024a). *The scaling dynamics and causal effects of a district-operated tutoring program* (EdWorkingPaper No. 24-1030). https://doi.org/10.26300/zcw7-4547

Kraft, M. A., Schueler, B. E., & Falken, G. (2024b). *What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability* (EdWorkingPaper No. 24 -1031). https://doi.org/10.26300/zygj-m525

Lewis, K., & Kuhfeld, M. (2023). *Education's long COVID: 2022–23 achievement data reveal stalled progress toward pandemic recovery* [Research brief.] NWEA.

Lewis, K., & Kuhfeld, M. (2022). *Progress toward pandemic recovery: Continued signs of rebounding achievement at the start of the 2022–23 school year* [Research brief.] NWEA.

Lewis, K., Kuhfeld, M., Ruzek, E., & McEachin, A. (2021). *Learning during COVID-19: Reading and math achievement in the 2020–21 school year* [Research brief.] NWEA.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012), *Translating the statistical representation of the effects of education intervention into more readily interpretable forms* (NCSER 2013–3000). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

Lynch, K., An, L., & Mancenido, Z. (2023). The impact of summer programs on student mathematics achievement: A meta-analysis. *Review of Educational Research*, *93*(2), 275–315. https://doi.org/10.3102/00346543221105543

MacGillis, A. (2023, June 19). What can we do about pandemic-related learning loss? *The New Yorker*. https://www.newyorker.com/magazine/2023/06/26/what-can-we-do-about-pandemic-related-learning-loss

McCombs, J. S., Pane, J. F., Augustine, C. H., Schwartz, H. L., Martorell, P., & Zakaras, L. (2014). *Ready for fall? Near-term effects of voluntary summer learning programs on low-income students' learning opportunities and outcomes*. RAND Corporation. https://doi.org/10.7249/RR815

McEachin, A., Augustine, C. H., & McCombs, J. (2018). Effective summer programming: What educators and policymakers should know. *American Educator*, *42*(1), 10–11.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for pre-K–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, *61*(1), 74–107. https://doi.org/10.3102/00028312231208687

Plecki, M. L., Elfers, A. M., & Nakamura, Y. (2012). Using evidence for teacher education program improvement and accountability: An illustrative case of the role of value-added measures. *Journal of Teacher Education*, *63*(5), 318–334. https://doi.org/10.1177/0022487112447110

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–315). Sage.

Reardon, S. F., Fahle, E. M., Ho, A. D., Shear, B. R., Kalogrides, D., Saliba, J., & Kane, T. J. (2023). *Stanford education data archive* (Version SEDA 2022 2.0). Stanford University Libraries. http://purl.stanford.edu/dt080zr0625

Riehl, E., & Welch, M. (2022). Accountability, test prep incentives, and the design of math and English exams. *Journal of Policy Analysis and Management*, *42*(1), 60–96. https://doi.org/10.1002/pam.22417

Robinson, C. D., Bisht, B., & Loeb, S. (2025). *The inequity of opt-in educational resources and an intervention to increase equitable access*. *Educational Researcher*. Advance online publication. https://doi.org/10.3102/0013189X251331518

Roche, D. (2023, June 20). *75 + ESSER III spending categories*. Burbio School Tracker. https://about.burbio.com/weekly-updates/burbio-school-tracker-6/20-75-esser-iii-spending-categories

Schwartz, J. L., McCombs, J. S., Augustine, C. H., & Leschitz, J. T. (2018). *Getting to work on summer learning: Recommended practices for success* (2nd ed). RAND Corporation.

Steiner, E. D., Doan, S., Woo, A., Gittens, A. D., Lawrence, R. A., Berdie, L., Wolfe, R. L., Greer, L., & Schwartz, H. L. (2022). *Restoring teacher and principal well-being is an essential step for rebuilding schools: Findings from the State of the American Teachers and State of the American Principal Surveys*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1108-4.html

Thum, Y. M., & Kuhfeld, M. (2020). *NWEA 2020 MAP Growth achievement status and growth norms for students and schools*. NWEA.

U.S. Department of Education. (2023). *National Assessment of Educational Progress (NAEP) long-term trend assessment results: Reading and mathematics*. Institute of Education Sciences, National Center for Education Statistics. https://www.nationsreportcard.gov/highlights/ltt/2023/

U.S. Department of Education. (2022). *National Assessment of Educational Progress (NAEP) 2022 long-term trend assessment results: Reading and mathematics*. Institute of Education Sciences, National Center for Education Statistics. https://www.nationsreportcard.gov/highlights/ltt/2022/

Zhao, A., & Ding, P. (2022). To adjust or not to adjust? Estimating the average treatment effect in randomized experiments with missing covariates. *Journal of the American Statistical Association*, *119*(545), 450–460. https://doi.org/10.1080/01621459.2022.2123814

Manuscript received May 29, 2024
Final revision received May 5, 2025
Accepted June 16, 2025