December 2022

The Challenges of Implementing Academic COVID Recovery Interventions: Evidence From the Road to Recovery Project

Maria V. Carbonari
Miles Davison
Michael DeArmond
Daniel Dewey
Elise Dizon-Ross
Dan Goldhaber
Ayesha K. Hashim
Thomas J. Kane
Andrew McEachin
Emily Morton
Tyler Patterson
Douglas O. Staiger

CALDER
Center for Education Policy Research
HARVARD UNIVERSITY
nwea Research

SUGGESTED CITATION
Carbonari, M., Davison, M., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Hashim, A., Kane, T., McEachin, A., Morton, E., Patterson, T., Staiger, D. (2022) The Challenges of Implementing Academic COVID Recovery Interventions: Evidence From the Road to Recovery Project. Research Report. Cambridge, MA: Center for Education Policy Research, Harvard University

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, see www.caldercenter.org/about-calder). 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 grateful to Anna McDonald, Ian Callen, Andrew Camp, and Andrew Diemer for their assistance with various aspects of this work.

Table of Contents

ABSTRACT

In this paper we examine academic recovery in 12 mid- to large-sized school districts across 10 states during the 2021–22 school year. Our findings highlight the challenges that recovery efforts faced during the 2021–22 school year. Although, on average, math and reading test score gains during the school year reached the pace of pre-pandemic school years, they were not accelerated beyond that pace. This is not surprising given that we found that districts struggled to implement recovery programs at the scale they had planned. In the districts where we had detailed data on student participation in academic interventions, we found that recovery efforts often fell short of original expectations for program scale, intensity of treatment, and impact. Interviews with a subsample of district leaders revealed several implementation challenges, including difficulty engaging targeted students consistently across schools, issues with staffing and limitations to staff capacity, challenges with scheduling, and limited engagement of parents as partners in recovery initiatives. Our findings on the pace and trajectory of recovery and the challenges of implementing recovery initiatives raise important questions about the scale of district recovery efforts.

INTRODUCTION

The pandemic's ongoing educational significance was put in stark relief with the recent release of findings from the National Assessment of Educational Progress (NAEP), which showed math and reading scores decreasing nationwide for fourth and eighth graders (U.S. Department of Education, 2022). The negative impact of the pandemic on students' scores was especially large for students with lower test scores and students from historically marginalized groups, exacerbating preexisting inequities (Dorn et al., 2021; Education Policy Innovation Collaborative, 2021; Lewis et al., 2021).

However, trends on formative assessment data across the 2021–22 school year suggest that students' scores have stopped declining. On average, students' pace of progress during the 2021–22 school year met or slightly exceeded pre-pandemic norms (Kuhfeld & Lewis, 2022). Given the many extraordinary challenges districts continued to face during the 2021-22 school year, such as continued COVID surges and increased mental health needs of students and staff, "typical" academic growth is as an accomplishment. But it is not enough. Based on the pace of recovery during the 2021–22 school year, Kuhfeld & Lewis (2022) estimate that most students in Grades 3 to 8 still need a minimum of 3 school years to fully recover from the negative academic effects of the pandemic, with upper elementary and middle school students potentially needing much longer. They estimate that students' test score gains during the 2021–22 school year were similar for students in low-poverty and high-poverty schools. But the inequitable initial impact of the pandemic on historically marginalized students means the timelines for their academic recovery are even longer.[1] Maintaining the 2021-22 pace of recovery would mean that many middle school students and historically marginalized students would never reach pre-pandemic levels of achievement before they exit the K–12 education system.

Aided by $190 billion from the American Rescue Plan's Elementary and Secondary School Emergency Relief Fund (ESSER), school districts across the country have responded by launching a range of targeted and districtwide academic interventions last year that were designed to help students catch up (e.g., expanded summer school and tutoring programs). Even as districts continue to face challenging operating environments—including tight labor markets (Domash & Summers, 2022; Jordan & DiMarco, 2022) and political polarization (Schwartz, 2022)—they will need to learn quickly from early recovery efforts and urgently apply those lessons to the remaining use of their ESSER funds in the months and years ahead to have hope of returning students to pre-pandemic levels of achievement.[2]

Not all interventions will be equally effective. The evidence on tutoring and summer school programs,[3] for instance, finds that the efficacy of these type of interventions depends greatly on both their design elements and the fidelity of their implementation (e.g., Lynch et al., 2022; McEachin et al., 2018; Nickow et al., 2020).

---

[1] It is important to note that this paper measures recovery and/or rebounding on achievement as measured by NWEA MAP Growth assessments. It is quite possible students are recovering from the pandemic on unmeasured dimensions that are outside the scope of this project.
[2] As of November 2022, all ESSER funds must be allocated by September 30, 2024, and spent by April 2026.
[3] "Program" and "intervention" are used interchangeably to refer to districts' academic recovery initiatives throughout the report.

Recent evidence on an ESSER-funded on-demand virtual tutoring program, for example, shows that the program had low take-up (19% of students in participating schools at baseline) and that the intended recipients of the program, struggling students, were far less likely to opt into the program than higher achieving students. Unsurprisingly, the program had no detectible effect on students' academic outcomes (Robinson et al., 2022).

To inform school districts' ongoing recovery efforts, this study examines academic recovery initiatives during the 2021–22 school year in a consortium of 12 mid- to large-sized school districts across 10 states. The districts are part of the Road to Recovery (R2R) project, which brings together district leaders and researchers to study COVID academic recovery initiatives and their efficacy.[4] The project aims to estimate the efficacy of the academic interventions implemented by the consortium that target subsets of students, making it possible to observe differences in outcomes across treatment and control groups. We also estimate each district's districtwide academic recovery and can parse the extent to which any observed recovery is driven by targeted academic interventions as opposed to districtwide practices or changes (e.g., professional development, curriculum changes). In this mixed-methods study, we focus on the following four questions:

1. To what extent did math and reading test achievement in the R2R districts rebound as of spring 2022?
2. What types of academic interventions did the R2R districts use, and what types of students did they target?
3. What were the initial results of these interventions?
4. How might the implementation of the interventions help explain their initial results?

Our findings highlight the challenges that recovery efforts faced during the 2021–22 school year. Although, on average, math and reading test score gains during the school year reached the pace of pre-pandemic school years, they were not accelerated beyond that pace; we find little evidence of systematic catch-up to pre-pandemic levels of test achievement. This is not surprising given that we found that districts struggled to implement recovery programs at the scale they had planned.[5]

We also found that the labels of popular recovery strategies mask significant differences in program design and take-up. Knowing that a district had a "tutoring" program, for example, revealed surprisingly little about the program or whom it targeted. In the districts where we had detailed data on student participation in academic interventions, we found that recovery efforts often fell short of original expectations for program scale or intensity of treatment.[6]

---

[4] The R2R districts are partnering with researchers from the American Institutes for Research (AIR), Harvard University, and NWEA in the R2R project to understand and improve how school districts are helping students catch up and recover from the learning opportunities they lost during the pandemic. For more details on the project, see www.covidrecovery.us.
[5] As we describe below, we were able to assess recovery in five districts in which we had detailed pre-pandemic test achievement information.
[6] At the time of writing, we were able to conduct implementation and impact analyses in four of the 12 districts based on data availability. We plan to conduct similar analyses in all 12 districts going forward.

Perhaps unsurprisingly, we found few statistically or practically significant effects of the interventions on student math and reading test scores through spring 2022. To better understand the results, we conducted follow-up interviews with leaders of seven programs in three of the districts. These interviews revealed several implementation challenges, including difficulty engaging targeted students consistently across schools, issues with staffing and limitations to staff capacity, challenges with scheduling, and limited engagement of parents as partners in recovery initiatives.

To some readers, the slowing average declines in student achievement during a tumultuous school year may feel like a victory of sorts—and it is a start to the long road to recovery. However, many systems will need to do better to help students catch up in the coming years. In particular, our findings on the pace and trajectory of recovery, and districts' plans, raise fundamental questions about the scale of district recovery efforts. Based on districts' plans during the 2021-22 school year, even if districts had implemented their recovery strategies at the scale they intended and to the students they intended to target, we estimate the associated gains would not have been sufficient to substantively alter the trajectory of student achievement. Given the relationships among test scores, educational attainment, and lifetime earnings—and their implications for opportunity and equity—this possibility should worry us all (Kane et al., 2022).

The rest of this paper is arranged as follows. We provide a brief overview of the sample and then, for each of the four research questions, we describe the data, methods, and findings. We conclude with a discussion of implications.

SAMPLE OVERVIEW

This report draws on a rich set of data from 12 mid- to large-sized school districts.[7] Together these districts enroll more than 600,000 students across 10 states spanning the Northeast, Southeast, Southwest, Midwest, and West. As noted in Table 1, the 12 districts serve higher proportions of students of color and students attending high-poverty schools compared to national averages.

We use different analytic samples to address each research question due to data availability (see Figure 1). To provide additional context for our 12 focal districts, we also present results using multiple years of data from a broader sample of over 3,000 districts who administered the NWEA MAP Growth math and reading tests, comprising more than 2.5 million students each year. The data, samples, and methods used to address each of the four research questions are described in the following sections.

[7] These districts include Alexandria City Public Schools (VA), Dallas Independent School District (TX), Guilford County Schools (NC), Pinellas County Schools (FL), Portland Public Schools (OR), Richardson Independent School District (TX), Santa Ana Unified School District (CA), Suffern Central School District (NY), Syracuse City School District (NY), Tulsa Public Schools (OK), and two districts that asked to remain anonymous.

In accordance with our agreement with partners, we mask the names of districts when reporting quantitative results and are purposely ambiguous when describing programs to protect districts' anonymity with respect to their results.

RQ 1: To what extent did math and reading test achievement in the R2R districts rebound as of spring 2022?

Data

The data used to measure academic achievement in this study come from student test scores on the NWEA Measures of Academic Progress (MAP) Growth math and reading assessments in Grades 3–8. The MAP Growth test has several advantages for our analysis. First, the tests are administered multiple times during the school year (fall, winter, and spring), allowing us to gauge changes in achievement during the school year.[8] Second, the tests are computer adaptive, such that the difficulty of the test items increases or decreases in response to a student's performance. Relative to fixed-form (i.e., nonadaptive) tests, adaptive tests are designed to provide more precise measures of achievement at the high and low ends of the distribution, which is particularly important in the context of the pandemic, when many students are performing below grade level (Kingsbury et al., 2014). Third, because the tests items are linked to a common scale, we can compare student test achievement and growth both within and across districts. Finally, because the scales are equal-interval (i.e. vertically scaled) across grades, we can also compare academic growth across students and time, both within an academic year and over multiple years.

To understand the extent to which students' academic performance has rebounded to pre-pandemic levels, we report the magnitude of achievement changes (in math and reading in Grades 3–5 and Grades 6–8 on the MAP Growth assessment) between fall 2019 and fall 2021, as well as the subsequent change in achievement during the 2021 academic year (i.e., between fall 2021, winter 2022, and spring 2022). Several of the R2R districts did not yet administer or administered limited MAP Growth testing during one or more of these time periods; therefore, we excluded district grades from the analysis if less than 60 percent of the student population had valid MAP Growth scores in any given term to ensure that our estimates are as representative of the districts as possible (we describe this process in more detail in the following section). Consequently, we estimated overall academic loss and recovery for elementary and/or middle school students in five R2R districts, across which more than 80 percent of students were tested on average in fall 2019 and fall, winter, and spring of 2021–22.

To contextualize our results, we also use data from the full set of districts nationwide that utilized the MAP Growth assessments[9] during the same time periods, but again restrict the sample to district-grades where at

[8] Districts choose the week that they administer the assessment each term, and scores are standardized (as described further in the Method section) such that they account for the instructional week that the test was taken.
[9] The nationwide NWEA sample is broadly representative of the population of U.S. public schools in terms of race and gender (Kuhfeld & Lewis, 2022).

least 60 percent of the student population had MAP Growth scores. Tables 2 and 3 show that the samples of students tested in math and reading in each term in the five R2R districts are similar in their race and gender to their enrolled populations of students in fall 2019. Relative to the nationwide NWEA sample, the math and reading R2R district samples on average have a similar percentage of Black students, a higher percentage of Hispanic students, and a smaller percentage of White students.

Method

To make our findings comparable to previous research on the efficacy of various interventions, we measured the losses in terms of the standard deviation of achievement before the pandemic. Specifically, we standardized student scores by grade, subject, and the instructional week in which assessments were administered using the NWEA 2020 MAP Growth norms (Thum & Kuhfeld, 2020). These norms were based on a nationally representative sample of students from the 2015–16, 2016–17, and 2017–18 school years and therefore represent the pre-pandemic nationwide distribution of student achievement. The resulting standardized score, $z(Y_{igt})$, for student *i* in Grade *g* tested in instructional week *t* was calculated in the following way, where $\bar{Y}_{gt}$ the predicted mean and $SD(Y_{gt})$ is the standard deviation based on the 2020 MAP Growth norms model (Thum & Kuhfeld, 2020):

$$z(Y_{igst}) = \frac{(Y_{igt} - \bar{Y}_{gt})}{SD(Y_{gt})}.$$

We averaged these z-scores for each Grade 3–8 in each R2R district and restricted our sample to grades where at least 60 percent of all students enrolled had MAP Growth test scores in all four of the points in time examined: fall 2019, fall 2021, winter 2022, and spring 2022.[10] This sample restriction left us with five districts that had at least one grade with representative testing data in both math and reading. To arrive at estimates of mean standardized achievement in each period for elementary and middle school grade ranges (i.e., Grades 3–5 and 6–8), we took a student weighted average of the z-scores across included grades. We did so for each of the five districts individually, as well as for the five of them collectively.[11]

We defined our estimate of achievement change at each grade level as the difference between the fall 2021 mean standardized MAP Growth score and the fall 2019 mean standardized MAP Growth score. Put differently, our estimate represents the difference in achievement between a post-pandemic grade (for example, 3rd

[10]We used data from the NCES Common Core of Data on the total student enrollment in each grade of each of our R2R districts. Because these data were not yet available for the 2021–22 school year, we used enrollment data from 2019–20 as a proxy for total enrollment in the fall, winter, and spring of 2021–22. We do the same for the full set of districts in the NWEA MAP Growth dataset.
[11]Note that our overall results are robust to variations in this approach. This includes supplemental analyses in which we use a more conservative 75% testing threshold to determine district-grades to retain in the sample, and analyses where we use unweighted averages to aggregate mean achievement in individual grades to the grade range level, or to aggregate mean achievement in individual districts to the combined district level. Across these different specifications, students experienced greater declines in math than in reading from 2019 to 2021, and scores have only seen relative improvement in math, with improvement driven specifically by students in the lower grades.

graders in fall 2021) and a pre-pandemic grade of students (3rd graders two years earlier)—where both are standardized on the same scale, relative to the nationwide distribution of achievement prior to the pandemic. Our estimates of the progress that districts have made toward recovery are similarly the difference between winter 2022 and fall 2021 mean standardized MAP Growth scores, and between spring 2022 and fall 2021 mean standardized MAP Growth scores.

To contextualize the achievement change between fall 2019 and fall 2021, we also approximated the hours of instruction that would be needed to eliminate the pandemic-driven observed declines that students saw during that period. According to a recent meta-analysis of experimental evidence on tutoring programs from preschool through secondary school (Nickow et al., 2020), pooled estimates of the impact of tutoring on student achievement are approximately 0.38 standard deviations in math and 0.35 standard deviations in reading. Although the tutoring programs included in the meta-analysis vary in terms of structure and dosage, for the sake of simplicity, we assumed that these average impacts corresponded to a high-dosage program consisting of 3 hours per week for 36 instructional weeks, or 108 total hours. We subsequently approximated the hours of instruction needed to counteract the pandemic loss by the following: $\frac{(Fall\ 2021 - Fall\ 2019)}{.38} * 108$ for math or $\frac{(Fall\ 2021 - Fall\ 2019)}{.35} * 108$ for reading.

However, if districts continue to face similar barriers to implementing interventions in the coming years as they faced during 2021-22 (described further in the Results section following RQ #4), the interventions may have weaker positive effects per hour of intervention on student outcomes than the pre-pandemic tutoring programs included in Nickow et al.'s (2020) meta-analysis. Therefore, our estimates of the additional hours of tutoring that would be needed for full recovery should be interpreted as a lower bound.

Results

In Panel A of Figure 2, we report the mean standardized score in Fall, Winter, and Spring for the full sample of NWEA districts. The scores are standardized by testing period using the pre-pandemic norms. For example, in Fall 2021, mean achievement in grade 3-5 math was .26 standard deviations below norm. By Spring 2022, students had made up a small amount of ground (.06 standard deviations) to end the year scoring .20 standard deviations below pre-pandemic norms. However, in Grade 6-8 math, students ended the year essentially where they started it, .21 standard deviations below norms. And, in reading, students seemed to lose ground, falling from .11 to .15 standard deviations below norms.

The average trajectory in the five Road to Recovery districts with necessary testing data was similar.[12] In math, students in the five districts saw a very slight recovery throughout the 2021–22 year, gaining an average of 0.01 standard deviations in the first semester and a total of 0.03 standard deviations by the end of the second

[12] See also Tables 4 and 5 for recovery trajectory estimates for R2R districts, as well as the estimated instructional hours to recover the loss.

semester. However, these modest gains were seen only among elementary students (grades 3-5), who had made up 0.07 standard deviations by Spring 2022. Middle schoolers, on the other hand, saw a slight further decline in the first semester, and ultimately had no improvement at the end of the school year compared to where they started. Although this pattern indicates that middle school students' test scores increased at theexpected rate based on pre-pandemic norms, it also means that, at this rate, they are not on a trajectory to recover from the negative impact of the pandemic.

Although the initial loss was smaller in reading than in math, students have not yet seen the same recovery in 2021–22. Overall, student average scores continued to decline by 0.07 standard deviations in the first semester and slowed to a total decline of 0.05 standard deviations across both semesters. Both elementary and middle school students experienced this trajectory, although declines were somewhat greater for middle schoolers. These average estimates of loss and progress toward recovery, however, mask variation across the five districts. In math, District 4's elementary students saw a promising 0.18 standard deviations recovery (although they had also had a significant initial loss of 0.40 standard deviations). Meanwhile, District 5 experienced a concerning continued decline in scores, particularly among its middle school students, whose students scored 0.11 standard deviations lower at the end of the 2021–22 school year than would be expected based on pre-pandemic achievement. Consistent with the variation in loss and recovery across districts, the number of estimated instructional hours to close the achievement loss as of spring 2022 also varied across districts (Tables 4 and 5). District 1 had essentially rebounded from the pandemic, while the rest of the districts needed between 40 and well over 100 instructional hours (or a similar dose of an effective intervention) per student on average to close the pandemic-induced achievement loss. As we discuss below, the levels and intensities of interventions (at least tutoring) occurring in the 2021-22 school year were well below the above estimates. In reading, the trajectories were somewhat more similar across schools, although total fall 2021 to spring 2022 changes varied from a positive 0.02 to a negative 0.16 standard deviations, with District 5 again seeing the most concerning trends. In addition, along with math, the districts varied in estimated hours of instruction needed to offset pandemic-induced losses. These varied from approximately 15 hours in District 1 to approximately 120 hours in District 5.

RQ 2: What types of academic interventions did the R2R districts use, and what types of students did they target?

Data

To understand district intervention plans and target populations, we collected a rich set of programmatic data on academic recovery interventions in all 12 districts. We defined an academic recovery intervention as a program that (a) was new or had expanded since the pandemic, (b) was supported by ESSER funds, and (c) provided targeted students with additional or more intensive learning time relative to core instruction, such that a treatment and control group could be identified within each targeted district-grade.

Method

We iteratively collected qualitative data on interventions throughout the school year via two waves of semi-structured interviews with program leaders that lasted between 60 and 90 minutes.[13] We conducted these interviews virtually in fall 2021 and spring 2022. During the interviews, we shared our notes with participants in real time so they could check our descriptions for accuracy, and we followed up via email and reviewed any documentation shared by district leaders to gain clarity on any outstanding questions. These interviews gave us a detailed description of the design (fall 2021) and implementation (spring 2022) of district academic interventions on several key dimensions: program type (e.g., tutoring, virtual learning), program content (e.g., math or reading), program intensity (sessions per week), program dosage (minutes per session), program duration, delivery mode (e.g., virtual or in-person), and student eligibility criteria. By the end of the school year, we had conducted a total of 29 interviews with 68 district staff.

Following the completion of the interviews, the research team used the interview notes to prepare matrices that captured the design of each intervention. The matrices were then used to produce summaries of the variation in program design and implementation features by program type across and within the R2R districts.

Results

Across the 12 districts, educators used a range of academic interventions to support COVID recovery during the 2021–22 school year.[14] As mentioned previously, all district initiatives examined in this study intended to provide students with additional or more intensive learning time in math and/or reading relative to what they otherwise would receive (i.e., business as usual). Based on districts' descriptions of their interventions, we grouped interventions into five major categories: (a) tutoring programs, (b) small-group push-in and pull-out interventions, (c) after-school and other out-of-school-time programs, (d) virtual learning programs, and (e) extended school year calendars. The number of districts implementing each type of intervention is displayed in Table 6.

As we describe in greater detail below, we matched our programmatic categories to the labels used by the districts, but the categories could also mask considerable variation in program designs between and within districts. Initiatives in each category varied on several dimensions: which students were targeted, when the intervention happened, whether sessions were virtual or in-person, the qualifications and backgrounds of the providers, the student-provider ratio, and the frequency and duration of sessions. In this section, we provide summaries of the variation we observed in the implementation of each type of program across the consortium

[13] Interview template available at https://caldercenter.org/sites/default/files/District-SR-template-summer22.pdf
[14] Districts also used summer programming as a common academic COVID recovery strategy in the summer of 2021. We do not include descriptions of districts' summer programming in this report because only a subset of districts was able to provide the research team with that information in the fall of 2021.

of districts. For more detail on the variation in districts' plans, see Appendix B.

Tutoring

Eight of the 12 districts used tutoring services to give students additional academic support. Most tutoring programs were centered on math and reading but occasionally included other school subjects. The implementation of tutoring programs varied across districts and sometimes across schools within one district. For example, one district offered both in-person and virtual tutoring programs at different schools. There were three general approaches to scheduling tutoring: two districts provided students with tutoring support during normal school hours; four districts providing tutoring outside of school hours; and two districts offered tutoring both during and outside of school. Some programs targeted subsets of at-risk students. But others provided optional, on-demand tutoring for any student who wanted it.

The personnel tasked with providing tutoring varied across districts. In different districts, tutors included credentialed teachers, local college students, national tutoring services, and local nonprofit organizations and/or peers (usually older, high-performing students). Among the programs that were not on demand, students received between 2–5 tutoring sessions per week, with total expected tutoring time ranging from 15–100 hours over the course of the entire year. The planned length of tutoring sessions also varied, from 15 minutes or less for on-demand tutoring services to more than 60 minutes for targeted tutoring programs.

Small-Group Push-In and Pull-Out Interventions

Six of the 12 districts designed small-group interventions to provide students with reading and math support during school hours. These interventions were often indistinguishable from or similar to tutoring programs in terms of their design features. Schools offered them exclusively during the school day, and they were facilitated only by certified teachers and specialists. The districts took multiple approaches to small-group interventions. Five districts used a "pull-out" approach that involved intervention staff removing students from core instruction to work with them in a separate space; one district used a "push-in" approach, in which the interventionist worked with a targeted group of students while their classmates participated in "reading stations" with the classroom teacher. Students were targeted for small-group push-in/pull-out interventions based on a variety of at-risk factors and low academic performance (e.g., test scores, course performance).

Districts varied in the amount of small-group intervention that they provided. Most districts offered these interventions for the entire school year. But a few places did not begin their interventions until midyear. Push-in/pull-out interventions generally happened 4–5 days per week and ranged from 25–50 minutes each day. Two districts capped the total amount of the intervention that students could receive (one district limited participation to 30 hours; the other limited participation to 50 hours). As we note in the section on implementation, student participation in small-group interventions was also influenced by classroom teachers' willingness to release students, resulting in further variation in the amount of small group instruction students received within each district.

After-School/Out-of-School-Time Programs

Five of the 12 districts used after-school, before-school, and Saturday programs to provide additional instruction to students. These programs included time allocated for academic instruction or support (e.g., homework help). Most before- and after-school programs provided students with a mix of enrichment activities, additional instructional time, and/or homework help. By contrast, Saturday school programs were typically focused only on math and reading instruction. Participation was optional for all programs. But, in some cases, students were encouraged to participate based on their academic performance. These programs also tended to serve subsets of schools and/or students. For example, two of the districts offered a program to all students in Grades K–8 at a subset of schools, three districts had a program open only to a subset of grades, and three districts targeted students with certain characteristics (e.g., students with disabilities, students with test scores below a certain threshold).

Districts also varied in the amount of after-school/out-of-school programming they offered targeted students. Most programming began in the fall, but four districts did not begin at least one of their programs until the spring semester. After-school and before-school programs were offered 1–5 days per week and ranged from 30 minutes to 1.5 hours of academic support per session. In total, the amount of additional instructional time offered through these programs ranged from roughly 7.5 to 100 hours per school year. Saturday programming was mostly offered in the spring and typically ran for 4.5 hours each day. Because the participation data for Saturday programming was inconsistent, the total amount of time that enrolled students spent at Saturday programming is unknown.

Virtual Learning Tools

Virtual learning tools (e.g., iReady, ALEKS, Dreambox) were used in four of the 12 districts to add academic time to students' days beyond core instruction. Although many districts used virtual learning tools during core instruction, the four districts opted for these tools explicitly to add instructional time in the relevant subjects beyond the time allocated during core instruction. Districts typically targeted students who were performing below grade level or whom a teacher had recommended for participation. One district required all students in a subset of schools to use the program. Districts assigned students to use the platform both during and outside of school hours. Two districts required targeted students to use the program during an intervention period during school, whereas two other districts expected students to use the program outside of school. Students could access the programming outside of school in all districts if they desired. The amount of time students were expected to use the program ranged from 30 minutes to 3 hours per week (approximately 18 to 108 hours per year).

Extended School Calendars

Finally, two of the 12 districts extended the school year in a subset of schools to give students additional days of instruction throughout the year. One district implemented three different models of an extended school year across participating schools, while the other district used one model. In two of the models, the additional days were not distinguishable from regular school days. In the other two models, the additional days provided slightly reduced math and reading instructional time, and more time was allocated for enrichment and social-emotional learning (SEL) activities. All extended school days were offered in person, staffed by teachers from the participating school sites, and had student–teacher ratios consistent with the schools' typical classroom ratios. The two districts varied in the schools and students they targeted, with one district targeting the lowest performing schools, and the other selecting schools based on school and family interest. Three of the programs were designed so that all students at the participating schools received additional days of instruction, whereas the remaining program specifically targeted students who were most in need of additional academic support.

The four models also varied in the total number of extra days and additional instructional time they provided to participating students relative to the traditional school calendar. One model provided students with 3 additional, typical days of instruction. Two other models gave students 22 additional days of instruction, but one of the models had half-days for each additional day. The full-day model consistently provided students with an additional 2 hours of reading instruction and 1.5 hours of math instruction per day, totaling an extra 44 hours of reading and 33 hours of math per year relative to the traditional calendar. The half-day model was less consistent in its instructional time across days but offered up to 1.5 hours of instruction in math and reading per day, totaling a maximum possible 33 additional hours in math and reading over the year. The remaining model provided students with up to 18 additional days of school. For this model, each day included 1.5 hours of reading instruction and 1 hour of math instruction, totaling a maximum of 27 additional hours of reading and 18 additional hours of math over the course of the year.

RQ 3: What were the initial results of these interventions?

Data

We used the information collected about the interventions during the district interviews to determine the relevant variables for capturing students' eligibility, assignment, and participation in the interventions, as well as variation in the intervention between and among students (e.g., a remote vs. in-person session, provider type). Districts also provided data on student demographics, enrollment, and daily attendance for all K–8 students.

We requested data from all 12 districts for each intervention available to students in Grades K–8 but received sufficient data on a timeline necessary to estimate intervention impacts for this report in only four.[15] These four

[15] These four districts were selected for impact analysis because they were the only districts that provided data before the requested deadline. Note that at least four of the other eight districts that did not provide data before the deadline also did not implement academic interventions at a sufficient scale and/or did not collect data that would enable the research team to conduct a meaningful analysis of the intervention.

districts implemented a variety of interventions that provided additional instruction in math and/or reading and targeted students in a range of grade levels, from kindergarteners through eighth graders.

Data

As with our analysis of achievement loss and rebound, we use MAP Growth assessment scores as the outcome measure of student academic performance. We estimate the impact of each recovery intervention using a value-added framework that controls for observable pretreatment student characteristics, as well as pretreatment test scores. This approach has been used to understand the impact of schools on student outcomes in general (e.g., McEachin et al., 2016), as well as to evaluate the impact of educational programs and policies on students' achievement (Barry & Sass, 2022).

Value-added methods can provide unbiased estimates of program impacts if students' assignment to treatment is as good as random after conditioning for observable pretreatment characteristics. However, in several of the participating districts, we saw evidence that student participation in academic interventions was based on information on academic progress that became available during the school year. Specifically, second semester participation was related to how a student performed on the winter 2021-22 MAP Growth assessment—even after controlling for the prior spring 2021 and fall 2021 test scores. If students who were making slower progress during the school year were more likely to be assigned to treatment, then our estimate of the impact of the treatment would be negatively biased (i.e., receiving treatment would be a signal of midyear troubles, not just a function of prior achievement).[16]

To account for midyear treatment assignment, we estimated the following model, measuring achievement growth semester by semester when program participation data were available separately by semester:

$$MAP_{igjts} = \alpha_0 + \alpha_1 Treatment_{igts} + \alpha_2 Eligible_{igts} + priorMAP_{igts}\gamma_{gts} + X_{igt}\theta_{gt} + \delta_{jgts} + \epsilon_{igts}$$

Here, MAP_{igjts} is the standardized end-of-term MAP Growth score for student *i* in grade g at school *j* in semester *t* and subject *s*. The $Treatment_{igts}$ measure is either a binary indicator of treatment receipt or a continuous measure of the number of hours of treatment received; we also included binary or continuous measures of students' participation in other recovery programs available in the district. For some recovery programs, students are eligible to participate if they scored below a certain level on a previous MAP Growth assessment or other standardized test. In those cases, $Eligible_{igts}$ is a binary indicator for whether student i met the program eligibility requirements, interacted with grade level (these variables are not included if eligibility for a program is not based on prior test scores). $priorMAP_{igts}$ is a matrix with a cubic function for each of two previous MAP Growth scores (e.g., fall 2021 and winter 2021–22, if term *t* is equal to the spring semester) in the same

[16] Note that our estimate would still be biased if program participation were associated with other unobservable student characteristics, such as socio-emotional challenges, home situations, or course grades.

subject, interacted with grade level and the term in which the treatment occurred. X_{igt} is a vector of baseline student characteristics, including race and ethnicity, gender, special education status, disability status, FRPL eligibility, and ELL status, as well as the start-of-term MAP Growth score in the other tested subject and the instructional week in which the end-of-term MAP Growth assessment was taken. δ_{jgts} contains school-grade-semester-subject-level fixed effects.

In one district, MAP Growth testing rates were notably low in spring 2022, with roughly 50 percent of tested grades not taking the assessment in that final term. Additionally, for several of their interventions, the available data did not specify how much of the total spring programming occurred before the spring MAP Growth assessment. As a result, in that district, we estimated the impact of first semester treatment participation only, using fall 2021 scores as the baseline achievement measure and winter 2022 results as the outcome.

Generally, the analytic sample for each district is limited to those students who have MAP Growth assessment scores from the start and end of the term in which treatment took place (e.g., fall 2021 and winter 2021-2022 for first semester recovery programs), as well as from two terms prior (e.g., spring 2021).[17] See Appendix A for more detail on model specifications and the placebo tests we conducted.

Results

Table 7 shows the estimated impacts of treatment on math achievement for each of a series of math interventions in 4 of the 12 districts. We report estimates for 10 district/intervention combinations across different grade ranges; for District D, which had a particularly large number of interventions, we report estimates for categories of interventions (classroom, small group, supervised, and other) as well as for all their interventions combined.[18] In column (1), we report the coefficient on the indicator of whether a student received any treatment with math achievement as the outcome. In seven of 10 cases, the confidence interval for the estimated impact includes zero, implying that we could not reject the null hypothesis of no impact. In the remaining three cases, the estimates were marginally significant impacts: District A Tutoring, District B Assigned Software and District D All Interventions.

Column (2) shows coefficients from a corresponding placebo test, in which we replicate the original model but exchange the outcome to be achievement in another subject that was not the focus of the intervention (in this case, reading.) We use the placebo test as an indicator of selection bias—that is, that the students who

[17] For some districts, we are also able to include state standardized test scores in the value-added model, enabling us to include students with missing MAP Growth scores if they have non-missing state test scores from the same term. See Appendix for more detail.
[18] Given the large number of interventions implemented in District D, we categorized their interventions into three overall categories: supervised learning (e.g. virtual instruction with a teacher in the classroom); small group direct instruction (less than approximately 10 students); classroom instruction (more than approximately 10 students); or other. The "All Interventions" category for District D is made up of the combination of all these categories.

participated in the treatment differed in ways not captured by the covariates. In only one of the three cases in which we found small positive coefficients on participation did the intervention also pass the placebo test: District A Tutoring. Columns 3 and 4 show the estimated treatment effect per hour of treatment, along with its corresponding placebo test. Here, we see significant positive effects that successfully pass the placebo test for a handful of interventions across multiple districts: Districts A Tutoring, District C Tutoring, and District D Small Group.

For context, we also report in columns 5 and 6 the average dosage that treated students received of each intervention per semester, as well as the estimated impact we would have expected to see if the programs had the same impact per hour as found in the pre-pandemic research on high-quality tutoring (Nickow et al, 2020). Because of the limited programming dosage provided during 2021-22, those impacts are also quite small, ranging from .01 to .05. In most but not all cases, both the hourly expected impact and total expected impact exceed the actual treatment effects of the district interventions.

Figure 3 displays these results graphically, both for the estimated effect of any treatment and for the estimated effect of an hour of treatment. In seven out of 10 cases, the effect is not statistically discernible from zero. However, in many cases, the confidence intervals also include the impact we might have expected for tutoring based on the pre-pandemic research. In other words, while we cannot reject that most interventions failed to improve math achievement, we also cannot reject the possibility that several interventions were just as effective (on a per hour basis) as high-quality pre-pandemic tutoring interventions.

Table 8 and Figure 4 show comparable results for 10 district/intervention combinations targeted at reading achievement. In only one case (District C Tutoring #1) was the estimate for the effect of any participation statistically different from zero—but the point estimate was negative.[19] For the estimated effects of an hour of treatment, District A Tutoring #2 and District C Tutoring #2 were statistically different from zero, though only the former was positive.

Because of the small, negative and/or null effects estimated for each intervention, we did not estimate interaction effects of interventions for students who participated in multiple programs. Nevertheless, a small proportion of students received multiple ELA interventions in two of the four districts and math interventions in three of the four districts. The percentage of students receiving multiple interventions in a subject in these districts ranged from 5 to 22 percent. A higher percentage of students were receiving at least one intervention in both math and ELA, ranging from 14 to 74 percent across the four districts.

When we consider the specifics of participation in these interventions, the estimated impacts shown in Figures 3 and 4 are unsurprising. Often, both the number of students served and the amount of instruction provided

[19] We think this is likely to be caused by selection bias—according to conversations with administrators in that district, teachers frequently assigned students to participate in the program throughout the semester if that student was having difficulties with reading.

were lower than planned. For example, except for one very small-scale program (which targeted and served about 5% of 1st graders in the district), districts' tutoring programs intended to serve between 22 and 35 percent of students in the targeted schools and grades. However, over the course of the school year, the data indicate that these programs reached only 20 to 30 percent of their intended enrollment, totaling 5 to 10 percent of all students in the targeted schools and grades.

The dose of programming students received also typically fell short of districts' plans. We found districts that had planned on offering students between 30 and 60 hours of mathematics tutoring per year ended up, on average, providing students between 12 and 14 hours of math tutoring. For students who did participate, the number of sessions and the length of sessions were also often less than originally planned. In one district that had planned to offer students 90 sessions of tutoring over the course of the school year, students attended 13 sessions on average. In another district, math tutoring sessions were supposed to provide 100 minutes of instruction during the week over five sessions; in practice, the average student attended 28 minutes of tutoring per week.

RQ 4: How might the implementation of the interventions help explain their initial results?

Data

In addition to the interviews described previously, we conducted a third wave of in-depth interviews with district intervention leaders in charge of seven interventions from three of the four districts in the summer of 2022. We conducted this third wave of interviews only with districts and interventions for which we had completed a credible impact analysis and could engage district intervention leaders in a conversation about both the implementation and impact results of their interventions. Table 9 outlines the number of district administrators interviewed for each district and the intervention programs discussed.

Method

We used a semi-structured interview protocol that included a presentation of the results of the intervention to its district leader. The results shared included descriptive statistics on the composition and number of students who received programming, descriptive statistics on the amount of programming students received, and a summary of any impact estimates. Following our sharing of the results, we asked questions about how student assignment, participation, staffing, delivery, and content varied (or not) from the original plans for the program. We also asked about the barriers district leaders faced in trying to carry out their original plans. Additional details and the limitations of our interview approach and qualitative analysis are described in Appendix A.

Results

The interviews revealed major implementation challenges that impeded districts' academic recovery intervention efforts, including challenges in (a) reaching the targeted students consistently and equitably across schools; (b) program staffing and staff capacity, (c) scheduling and delivering intervention services; (d) engaging families as partners in intervention delivery; (e) adapting intervention programs to accommodate existing federal, state, and district policies; and (f) building central office capacity and internal systems for scaling intervention programs Importantly, each of these challenges was situated in and often exacerbated by the challenging context of COVID recovery.

Issues Associated with Reaching Targeted Students

Each of the seven interventions for which we conducted interviews were planned to target students who were behind academically. Targeting was typically based on one or more test performance thresholds (e.g., students who had scored below the 20th percentile on the MAP Growth test). But schools sometimes incorporated other eligibility criteria, such as low attendance rates, low course grades, or teacher recommendations when assigning students to programs.

Our interviews indicated that educators often had to balance district-mandated eligibility criteria against the professional judgments of school leaders and teachers about which students had the greatest needs and/or would be best served by the intervention. To strike this balance, district leaders often decentralized decisions about student eligibility and participation in intervention programs at the school and classroom levels. While a decentralized approach generated buy-in from school actors for implementing recovery efforts and, at times, resulted in schools engaging in better outreach, recruitment, and matching of students with intervention services, it also created inequities in student access to intervention services across schools.

For example, in one district, teachers recommended students who scored above the threshold for an intervention because the teachers' experience suggested to them that the students' test scores were inflated. For a different intervention in the same district, leaders insisted on district-level eligibility criteria (e.g., test score thresholds) for the first wave of students who accessed the intervention and then allowed schools to use their own criteria to identify a second wave of students to access the same intervention. The district leaders felt this approach improved local buy-in and allowed schools to expand access to intervention students to all students in need, although it also increased variation in who participated in the intervention across school sites.

Relatedly, while district leaders observed some benefits from school-level management of student recruitment for intervention services, variation in school efforts to invite and recruit students affected which and how many students participated across schools. One intervention leader noted that schools that relied on district-provided messaging and recruitment had lower participation rates than schools that customized their recruitment messaging and timelines to ensure that families were aware of the program, knew how to access it, and felt comfortable accessing it.

In some cases, local school-level eligibility decisions may have helped deliver support to the students most in (i.e., teachers know best). But at other times, local decision-making appears to have directed services away from the intended population to students with fewer academic needs. In one school district, for example, a reading program intended to help students who performed at or below the 15th percentile of a school's test score distribution was used in some schools to provide support to "bubble" students on the cusp of proficiency. In another district, 31% of the students who participated in a math intervention intended to serve students at or below the 20th percentile on the fall 2021 math MAP Growth test had scored above the 40th percentile. In two of the districts, leaders reported that schools occasionally ended up using tutoring to help students who were performing at grade level but struggling with a specific topic. One leader shared, "I think it [tutoring] is happening with the wrong set of kids."

Another reason schools may not have adhered to the original targeting criteria for an intervention was the realization that interventions were sometimes mismatched with student needs. For example, a leader of a math intervention in one district said that some schools found that the students originally targeted for the intervention lacked the foundational skills needed to benefit from it. In response, the district expanded eligibility for the program from the lowest 25 percent of math performers to the lowest 30–35 percent of performers and gave the teacher discretion to identify the students in this group who matched intervention qualifications.

In short, although the guidelines and process for assigning students to interventions may appear routine on paper, the districts struggled to apply eligibility criteria consistently across students and schools, which could had made it hard to ensure that the students most in need received the intervention.

Issues Associated with Hiring and Deploying Intervention Staff and Staff Capacity

The district leaders we interviewed encountered a range of staffing issues during implementation. Most notably, districts had to hire and manage many new providers to staff intervention programs, which presented challenges for ensuring staff quality and delayed the rollout of intervention services into schools. Broader staffing challenges stemming from the pandemic (e.g. COVID outbreaks amongst staff, influx of new or temporary staff members due to shortages etc.), limited school familiarity with intervention programs, and multiple competing district initiatives further constrained staff capacity for implementing intervention programs.

District leaders shared different approaches for hiring staff for intervention programs. Some districts contracted with vendors or hired intervention specialists to work in schools. Others hired graduate assistants, retired and current teachers, or undergraduate and high school students. When possible, districts leveraged existing staff and existing relationships with vendors, individual volunteers, and community-based organizations to find intervention staff.

Each approach to hiring new intervention providers came with its own challenges to ensuring provider quality. Districts that contracted with vendors to hire intervention providers had less insight into and control over the providers' qualifications and selection process; in these cases, district leaders said they found it difficult to ensure staff quality and consistency throughout the year. Alternatively, hiring intervention specialists and tutors individually required districts to invest substantial time and resources into recruiting, onboarding, and training. In the case of one tutoring program that relied on community providers, the intervention leader felt they did not have the luxury to do anything beyond basic background checks because of a tight labor market.

Different hiring approaches had consequences for when interventions started during the year and the speed at which they ramped up services to students. District leaders reported that leveraging existing staff and relationships with providers was helpful for getting interventions started earlier in the year, but none of the districts we interviewed were able to staff all their interventions without hiring additional providers. For example, one district spent the first 5 months of the school year negotiating contracts with tutoring vendors to ensure that they were federally compliant and could be paid for using ESSER funding. As a result, the district's tutoring programs did not launch in schools until February and March of 2022, as contracts at individual schools were resolved on a rolling basis. In another district, a small team of centralized district staff was responsible for hiring, onboarding, and training tutor providers. The leader of this team felt that the team's limited capacity created a bottleneck that delayed tutors' placement in schools. Moreover, once placed in schools, tutors had to work collaboratively with teachers to identify student needs and content for tutoring. Such collaboration did not occur consistently in schools and further delayed the initiation of tutoring services. In certain schools, teacher–tutor relationships had to be restarted because of persistent teacher turnover.

Once providers were onboarded and delivering interventions to students, broader staffing challenges stemming from the pandemic spilled over into staffing challenges for intervention programs. One intervention leader reported that schools sometimes redeployed intervention specialists to cover regular classrooms and cope with COVID-related teacher absences during the Delta and Omicron surges. The leader of a reading intervention in another district concurred, explaining how the Delta surge affected staffing in one school:

> *At the start of the year at one of our schools they had something like 24 teachers out, they all had COVID, that was two weeks where interventionists were pulled from what they would regularly do. There's no way around it...you need a body in the classroom.*

Reallocating resources in this way not only delayed the intervention; some of the interventionists themselves also ended up being out during key times because of illness. "Usually, it was a domino effect," the leader said. Again, this delayed program launch, in some cases, for up to 4 weeks. In the same district, teachers sometimes used interventionists at the beginning of the year to help get small groups going rather than delivering the intervention; as one leader put it, the interventionists "have an eye on what the school needs," beyond their specific responsibilities.

The lack of familiarity with intervention programs at schools further limited staff capacity to implement intervention programs. District leaders observed more successful implementation of programs when schools were already implementing or piloting these programs before the pandemic, had existing relationships between interventionists and teachers, and when interventions were clearly aligned with core curricula. However, these conditions were not consistently present across schools. One district leader shared that some of the district's schools had tutoring programs in place before the pandemic; these schools were better positioned to grow these programs with the onset of district funding for tutors. While district leaders could have chosen intervention programs that mapped onto core curricula to support teacher uptake of programs, this was not always the case. In cases where intervention programs were closely aligned with core curricula, district leaders noticed better uptake and use by teachers.

Multiple and competing district initiatives also strained educator capacity for implementing intervention programs. Examples of concurrent initiatives and tools that districts adopted during the 2020-21 school year include new core curricula in reading and math, new training and professional development for teachers, COVID quarantine and testing policies and procedures, other digital tools for assessing and remediating student learning, new social-emotional and mental health supports, and other districtwide intervention programs. One district leader asked rhetorically, "How much capacity do people have? It [the multiple initiatives] is so much," implying that educators were overburdened and exhausted by the new policies and programs adopted by the district. Another district leader shared that, because schools were still learning how to implement other intervention programs that served the same student population as their tutoring program, it made it harder to ensure consistent scheduling for students and tutors. As noted earlier, tutors and intervention providers were also sometimes asked to support other intervention programs or fill in for classroom teachers in schools that faced staffing shortages.

Issues Associated with Scheduling and Delivering Interventions

Scheduling challenges offer another explanation for why students often received lower doses of interventions than originally planned. "It comes down to access," said one program leader, "How easy is it to pull a student [from class] and bring them back?" Across all three districts, intervention leaders reported that delivering programs that involved pulling students from class during the school day could be challenging. This was due, in part, to instructional time being fully planned out during the regular school day. Responding to data showing low intervention uptake and dosage, one district leader shared, "All of our literacy minutes were already being used for other things, so the data do not shock me." Some teachers resisted having students pulled from their classes because they did not want them to miss grade-level core instruction. In other cases, students who would have been eligible for a school-day intervention based on their test scores could not receive the service because it conflicted with other, higher priority (or state-mandated) scheduled supports (e.g., ELL/Individualized Education Program services).

In many cases, intervention providers had to navigate schedules with individual teachers to work with their students. This process meant that the same intervention could occur at different times in different buildings, so what students missed during their intervention (i.e., the untreated counterfactual) varied widely across students and schools. One tutoring program director likened scheduling to a complex puzzle, a "game of figuring out where each person goes and fits in [so that]…kids get hours but also we want tutors to get their hours."

Local complexity and discretion sometimes meant that “schools did their own thing [when it came to scheduling] and that is hard for us [the district] to control,” according to one district leader.

Arranging intervention times was not the only scheduling challenge that schools faced. In some cases, schools did not have adequate space for interventionists to work with students in small groups, which further complicated program delivery. A district leader of a math intervention in one district, for example, said:

> *Location was often an issue. Classrooms were not physically designed to have a group pulled in the back in many schools. So, their [students’] time was less because they lost minutes coming and going to the group.*

By contrast, in cases where intervention providers had space to work and could easily bring all their materials into the classroom, this leader reported that schools were better able to provide the planned dose of the intervention.

In addition to school-level issues, district-level schedules could make accessing interventions easier or harder. In the same district that ran into space issues, the district mandated extra intervention minutes for reading in all elementary school schedules but not for math. As a result, reading intervention providers (a position that predated the pandemic) were more likely to find time to work with students compared to math intervention providers (a new position).

In each of these cases, ease of scheduling was in part a function of who was responsible for scheduling and the extent to which intervention times aligned with existing school schedules. When intervention time was accounted for in school schedules and building administrators helped prioritize and coordinate scheduling, program leaders reported fewer scheduling issues. When schools worked directly with external contractors to schedule interventions outside of school hours, district leaders reported fewer issues and constraints with scheduling. However, some leaders shared that, even if scheduling intervention sessions after the school day was easier from a logistical standpoint, this approach limited access for students who wanted to participate in extracurricular activities or did not have access to transportation after school.

As noted earlier, schools also sometimes struggled to provide interventions as planned because of teacher absences during the Delta and Omicron surges. These surges also led to chronic student absences in some schools that reduced the planned-for frequency and dosage. Fear of COVID also affected the interventions; early in the school year, leaders said that some teachers were reluctant to send students to pull-out groups because they feared it would increase everyone’s risk of infection. As students moved in and out of school and experienced stress and pressure related to the pandemic, some interventionists also reported challenges with student behavior that made it harder to deliver the planned dose of academic support. Commenting on the amount of time spent in intervention sessions to manage student behavior, one district leader shared, “If behavior is the thing that students need to get going [in school], maybe behavior should be the intervention.”

Limited Family and Community Engagement

Our conversations with district leaders suggest that families and community-based organizations, who could have been valuable partners in supporting the implementation of intervention programs, were not actively engaged by districts. Moreover, districts may not have given consistent guidance to schools on how to engage families in academic recovery efforts. One district administrator shared that some families and community partners were eager to use the virtual learning tools provided through one intervention to practice with students at home or after school. Yet the district did not offer additional support to families or community partners to increase program uptake. Given that schools struggled to find enough time within the school day to support intervention programming, one solution may be to share resources that families and community members can use with students outside of school. As one district leader put it, "Being more consistent at the district with how we support families and not leaving that to schools would be a true opportunity [and] growth point because kids could have done [the intervention] outside of school." As noted earlier, schools employed different strategies to invite families to participate in academic recovery programs, with some more successful than others, suggesting the need for districts to provide consistent guidance and support to schools for family outreach.

Challenges With Adapting Intervention Programs to Accommodate Existing Federal, State, and District Policies

District leaders also had the challenging task of embedding interventions in an existing system of federal, state, and local policies. At times, this required adapting interventions to accommodate existing rules and procedures which, in turn, delayed the rollout of services or diminished their quality. As noted earlier, to use ESSER funding for a tutoring intervention, leaders in one district had to revise their vendor contracts to meet federal contracting requirements, which delayed the program's rollout. Another district leader discussed having to comply with a state mandate requiring the use of tutoring to deliver a remediation curriculum, even though the leader believed it was more appropriate to use tutoring for grade-level content. As noted earlier, districts were implementing concurrent interventions that could conflict with the academic recovery intervention in ways confusing to teachers. To prevent confusion and frustration, district leaders prioritized aligning the features of the interventions and occasionally had to depart from evidence-based practices. For example, one district administrator discussed increasing tutoring group sizes to more than what is considered best practice to align with the small-group sizes prescribed by the district's recently adopted, multi-tiered system of supports (MTSS) program.

Limited Central Office Capacity and Internal Systems to Scale Interventions

District central offices lacked sufficient bandwidth to oversee and coordinate the implementation of these programs. Many of the representatives we spoke to worked in small teams, consisting of two or three total staff members, who were suddenly in charge of hiring intervention providers, coordinating school schedules, and overseeing implementation of an intervention program for their entire district. Therefore, district leaders had limited time and capacity to manage these processes while also fulfilling other professional roles and responsibilities in the district. In reflecting on the past year, one district administrator shared that they could have provided better professional development to interventionists had it not been for the hours of new literacy training required by state law that they also had to provide for teachers.

District representatives also described working with internal systems that were not designed to handle the demands of intervention programs at such a large scale. As noted earlier, one district's process for hiring, onboarding, and training tutors was time consuming and delayed student placement with tutors. Another district leader shared that compliance management of diverse tutoring providers was cumbersome, primarily because the district did not have internal data systems to uniformly track tutoring hours and attendance across different providers. These remarks suggest that, when implementing intervention programs at scale, districts need to invest in central office staffing and internal systems for overseeing these programs.

In summary, COVID-recovery interventions were often not implemented at the frequency or dosage originally planned in part because schools faced challenges related to reaching the targeted students, staffing, scheduling interventions, and engaging families as partners in academic recovery. Of course, these tasks were challenging in large part because schools were attempting to help students recover from COVID while the pandemic was still happening. In addition, district leaders had limited capacity and systems from within the central office to rapidly take these programs to scale, and sometimes had to adapt intervention programs to accommodate existing policies in ways that delayed services or reduced the quality of services offered to students.

The findings from our interviews underscore the challenging reality of the districts' implementation contexts, which were often full of competing and sometimes chaotic demands. While our findings illuminate how these challenges hindered implementation in the 2021-2022 school year, many of the districts have already developed plans to address some of these persistent challenges in the 2022-2023 school year. Our interviews with district leaders suggest that implementation of recovery interventions is an iterative process that will require continual adjustments to external (e.g., responses to pandemic factors, continued government funding, etc.) and internal (e.g., staffing shortages, family engagement, etc.) factors. Given the unique and challenging implementation contexts of many of these programs, it is promising that students are no longer falling behind and that districts are developing responses to improve on their efforts from 2021-2022.

DISCUSSION AND CONCLUSION

Consistent with other recent evidence (e.g., Jacobson, 2022; Kuhfeld & Lewis, 2022), our analysis of district recovery trajectories suggests that students resumed learning at a pre-pandemic pace during the 2021–22 school year, particularly in math. Unfortunately, we also found that students' test scores remain far below those of similar students during the pre-pandemic school years. To catch up, student learning will need to move at a faster pace than it did pre-pandemic. To be sure, the story is not the same in every district we studied. One district, for example, had small initial declines and is now on pace to have test scores equal to or exceeding their scores from fall 2019. But overall, our analyses suggest that other districts will not reach pre-pandemic levels without significant acceleration.

To accelerate learning and help students catch up, districts are implementing a range of academic interventions. Most of these interventions target subsets of academically struggling students. So far, our analysis finds little evidence that any of these interventions had large positive effects on student test achievement during the 2021-2022 school year. Of course, in theory, the wide range of catch-up efforts in districts could be raising achievement for all students, making it hard to detect treatment effects from the interventions. But as we reported in Figure 2, there is little evidence of broad improvements, except possibly in grades 3-5 math.

To better understand our findings on intervention impacts, we interviewed a subset of district leaders about implementing interventions. Those interviews revealed that districts recovery efforts were often plagued by staffing and scheduling problems and that schools faced challenges reaching the intended students. Partly as a result, many interventions served fewer students than originally intended—and often students who were not in the targeted groups. In some cases, academic interventions displaced regular classroom instruction, reducing the contrast between the intervention "treatment" and business as usual, again making it difficult to detect treatment effects.

The implementation challenges district leaders recounted suggest that the simple-sounding logic of academic intervention—identify students in need and provide them extra support—belies a host of complex design and implementation decisions. Under the existing capacity constraints of district staff and the time constraints of typical school days and years, there are no easy solutions to districts' challenges with staffing and scheduling interventions for students.

Providing sufficient intervention for all students in need is going to require historic action, beginning with a renewed effort to engage families. States and districts can help by providing transparent and accessible measures of students' academic progress and recovery to families and students. Recent surveys indicate that parents currently underestimate the extent to which their own students are behind (Anderson et al., 2022; Hubbard & Burns, 2022). Districts and states may need to do more to inform families and communities about how students are doing now, whether they are on track for recovery, and what can be done if recovery does not look like it is happening at an adequate pace. There is evidence, outside of the pandemic context, that better alignment between grades and measured test scores results in better student achievement (Gershenson et al., 2022). While there is little systematic evidence about how grades have changed because of the pandemic, it may be important for school districts to make sure there is good alignment between grades and other student outcomes (e.g., math and reading assessments), since grades are arguably the most direct means for schools to communicate with parents about student learning.

More generally, our analysis suggests that district leaders could be more intentional about engaging families as partners in academic recovery efforts. Two-way communication and engagement among districts, schools, and families could inform better design and uptake of academic recovery initiatives. Since many catch-up strategies require some level of family involvement in the form of picking students up from after-school programs or accommodating an extended school calendar, it will be important for districts and schools to engage families on their willingness and capacity to support such activities and to proactively address barriers that might hinder students' participation.

Relatedly, many schools are implementing voluntary interventions which require school systems to clearly articulate the extent to which students need supplemental (outside of the regular school day) services and to nudge families to use the intervention(s) to get even moderate student take-up (Robinson et al., 2022). Our interviews with district leaders suggest in some cases families and community-based organizations were eager to use these voluntary interventions but that districts did not offer additional guidance and support. As districts and schools grapple with finding enough time for academic recovery, partnering with families and community-based organizations to offer interventions outside of school could be a promising strategy.

Successfully increasing the scale of interventions in districts will, in some cases, also require more resources (e.g., staff and staff compensation). We show elsewhere that learning losses varied across districts (Goldhaber et al., 2022a) and that the ESSER funds received are likely to be sufficient for recovery in districts that serve high proportions of low-income students but did not spend a large share of time in the 2020-21 school year with remote learning. But ESSER funds are unlikely to be sufficient for the larger share of districts that spent more time in remote status or serve low percentages of high-poverty students (so received less ESSER dollars) (Goldhaber et al., 2022b; Shores & Steinberg, 2022). In addition to funding, our findings suggest that districts may need to invest in central office capacity and internal administrative systems (e.g., data systems, hiring procedures) to implement academic recovery interventions at scale.

Given that a tight labor market negatively impacted the ability of schools to implement some recovery initiatives, districts may also need to cast a broader net to recruit adults to provide interventions in schools and seek out new or expanded partnerships with external organizations. Our interviews indicate that some districts managed to supplement their academic interventions with external partnerships. They tapped local community centers, educator preparation programs, college students, parents, and local community members to provide academic help. Given the scale of the need, these types of external partnerships are viewed as a key resource for expanding their recovery efforts in the 2022-23 school year. Not every district we studied, however, was able to leverage external partnerships to support academic recovery.

Finally, districts will need help to expand their interventions to be commensurate with their students' losses. In most cases, expansion will mean expanding student participation and dosage in existing programs, as well as layering interventions (e.g., high dosage tutoring and an extended school year) for targeted students. We use estimates of the effects of high-dosage tutoring—a strategy used across eight of the 12 districts in the R2R consortium—to illustrate this point. Based on the impact of the pandemic in our consortium of districts, as well as the pace of recovery, they will need to spend between 40 and more than 100 hours on tutoring for the average student to recover in math, and slightly lower estimates for reading.[20]

This level of intervention would be a significant step up in intensity from what was implemented during the 2021–22 school year.

While there are many challenges on the road to academic recovery, districts are working hard, looking ahead, learning from prior implementation, and developing plans to improve in the 2022-2023 school year. Districts do not have to tackle this problem alone. States and other civic leaders can help districts mobilize communities by providing information, political cover (for example, on extending learning time), and investing in the capacity of districts, schools, and communities to support and advocate for recovery. Complete academic recovery—and, ideally, academic acceleration—is as urgent as it is challenging. Especially in the places hit hardest by the pandemic, academic recovery from COVID-19 is likely to require an all-hands-on-deck response for the next several years.

[20] As mentioned previously, these estimates are likely lower bounds on the hours of tutoring that will be needed for full recovery if districts' tutoring programs continue to face implementation challenges that reduce the quality or quantity of tutoring students receive relative to the programs cited in the evidence-base on high dosage tutoring (Nickow et al., 2020).

REFERENCES

Anderson, M., Faverio, M., & McClain, C. (2022, June 2). How teens navigate school during COVID-19. Pew Research Center. https://www.pewresearch.org/internet/2022/06/02/how-teens-navigate-school-during-covid-19/#teens-and-parents-express-their-views-about-virtual-learning-and-the-pandemic-s-impact-on-educational-achievement.

Barry, S. S., & Sass, T. R. (2022). The impact of a 2021 summer school program on student achievement. Georgia Policy Labs Report. https://gpl.gsu.edu/publications/impact-of-a-2021-summer-school-program-on-student-achievement/.

Camera, L. (2022, January 18). Country's biggest school districts resist going remote as closures spread nationwide. U.S. News & World Report. https://www.usnews.com/news/education-news/articles/2022-01-18/countrys-biggest-school-districts-resist-going-remote-as-closures-spread-nationwide

DeArmond, M., Hill, P., Destler, K., & Campbell, C. Whack-a-mole: School systems respond to disrupted learning in 2021. Center on Reinventing Public Education. https://crpe.org/whack-a-mole-school-systems-respond-to-disrupted-learning-in-2021/.

Diliberti, M. K., & Schwartz, H. L. (2022). Districts continue to struggle with staffing, political polarization, and unfinished instruction. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA956-13.html

Domash, A., & Summers, L. H. (2022). How tight are U.S. labor markets? (Working Paper No. 29739). National Bureau of Economic Research. https://www.nber.org/papers/w29739.

Dorn, E., Hancock, B., & Sarakatsannis, J. (2021, July 27). COVID-19 and education: The lingering effects of unfinished learning. McKinsey & Company. https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning.

Education Policy Innovation Collaborative (EPIC). (2021). K–8 Student achievement and achievement gaps on Michigan's 2020–21 benchmark and summative assessments. https://epicedpolicy.org/wp-content/uploads/2022/01/EPIC_BenchmarkII_Rptv1_Dec2021.pdf.

Gershenson, S., Holt, S., & Tyner, A. (2022). Making the Grade: The Effect of Teacher Grading Standards on Student Outcomes. (IZA Discussion Paper No. 15556). https://ssrn.com/abstract=4226363.

Goldhaber, D., Kane, T. J., McEachin, A., & Morton, E. (2022a). A comprehensive picture of achievement across the COVID–19 pandemic years: Examining variation in test levels and growth across districts, schools, grades, and students. (CALDER Working Paper No. 266–0522). https://caldercenter.org/sites/default/files/CALDER%20Working%20Paper%20266-0522_0.pdf.

Goldhaber, D., Kane, T.J., McEachin, A., Morton, E., Patterson, T., & Staiger, D.O. (2022b). The Consequences of Remote and Hybrid Instruction During the Pandemic. (Working Paper No. 30010). National Bureau of Economic Research. https://www.nber.org/papers/w30010.

Hubbard, B., & Burns, A. (2022, June). Hidden in plain sight: A way forward for equity-centered family engagement. Learning Heroes. https://learningheroes.wpenginepowered.com/wp-content/uploads/2022/06/Parents22-Research-Deck-1.pdf.

Jacobson, L. (2022, October). Exclusive literacy data: Small gains since last fall, but no reading rebound. The74. https://www.the74million.org/article/exclusive-literacy-data-small-gains-since-last-fall-but-no-reading-rebound/.

Jordan, P. W., & DiMarco, B. (2022). Educators and ESSER: How pandemic spending is reshaping the teaching profession. FutureEd at Georgetown University's McCourt School of Public Policy. https://www.future-ed.org/educators-and-esser-how-pandemic-spending-in-reshaping-the-teaching-profession/.

Kane, T., Doty E., Patterson, T., & Staiger, D. (2022, October). What do changes in state test scores imply for later life outcomes? Center for Education Policy Research, Harvard University. https://cepr.harvard.edu/files/cepr/files/long_term_outcomes.pdf?m=1666806941.

Kingsbury, G. G., Nesterak, M., & Freeman, E. (2014, March). The potential of adaptive assessment. Education Leadership, 71(6). https://www.nwea.org/research/publication/the-potential-of-adaptive-assessment/
Kuhfeld, M., & Lewis, K. (2022). Student achievement in 2021–22: Cause for hope and continued urgency. NWEA. https://www.nwea.org/research/publication/student-achievement-in-2021-22-cause-for-hope-and-continued-urgency.

Kuhfeld, M., Soland, J., Lewis, K., & Morton, E. (2022, March 3). The pandemic has had devastating impacts on learning. What will it take to help students catch up? Brookings. https://www.brookings.edu/blog/brown-center-chalkboard/2022/03/03/the-pandemic-has-had-devastating-impacts-on-learning-what-will-it-take-to-help-students-catch-up/.

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

Lynch, K., An, L., & Mancenido, Z. (2022, July). The impact of summer programs on student mathematics achievement: A meta-analysis. (Ed Working Paper No. 21-379). Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai21-379.

McEachin, A., Augustine, C. H., & McCombs, J. (2018). Effective summer programming: What educators and policymakers should know. American Educator, 42(1), 10. https://eric.ed.gov/?id=EJ1173313.

McEachin, A., Welsh, R., & Brewer, D. J. (2016). Student achievement within a portfolio management model: Early results from New Orleans. Educational Evaluation and Policy Analysis, 38(4), 669–691.

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on preK–12 learning: A systematic review and meta-analysis of the experimental evidence. (Working Paper No. 27476). National Bureau of Economic Research. https://www.nber.org/papers/w27476.

Polikoff, M., & Houston, D. (2022, September). Experts Say Kids Are Far Behind After COVID; Parents Shrug. Why the Disconnect?. The74m. https://www.the74million.org/article/experts-say-kids-are-far-behind-after-covid-parents-shrug-why-the-disconnect/.

Robinson, C. D., Bisht, B., & Loeb, S. (2022). The inequity of opt-in educational resources and an intervention to increase equitable access. (Ed Working Paper No. 22-654). Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai22-654.

Schwartz, H. (2022, March 14). What is really polarizing schools right now? Education Week. https://www.edweek.org/leadership/opinion-what-is-really-polarizing-schools-right-now/2022/03.

Shores, K., & Steinberg, M. P. (2022). Fiscal Federalism and K–12 Education Funding: Policy Lessons from Two Educational Crises. Educational Researcher, 0(0). https://doi.org/10.3102/0013189X221125764.

Thum, Y. M., & Kuhfeld, M. (2020, April). NWEA 2020 MAP growth: Achievement status and growth norms—Tables for students and schools. NWEA. https://teach.mapnwea.org/impl/NormsTables.pdf.

U.S. Department of Education. 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/.

TABLES AND FIGURES

Figure 1. We Use Data From A Nested Sample of 12 Districts to Examine Recovery Initiatives, Pandemic Losses and Recovery Trajectories, Intervention Impacts, and Implementation

We collected descriptions of recovery interventions in 12 districts...

...we estimated pandemic losses and recovery trajectories in 5 of those 12 districts...

...in 4 of those 5 districts, we examined intervention impacts on student achievement...

...in 3 of those 4 we conducted follow-up interviews to learn more about intervention implementation.

Figure 2. Standardized Achievement Loss and Recovery

A. All NWEA Districts, Nationwide

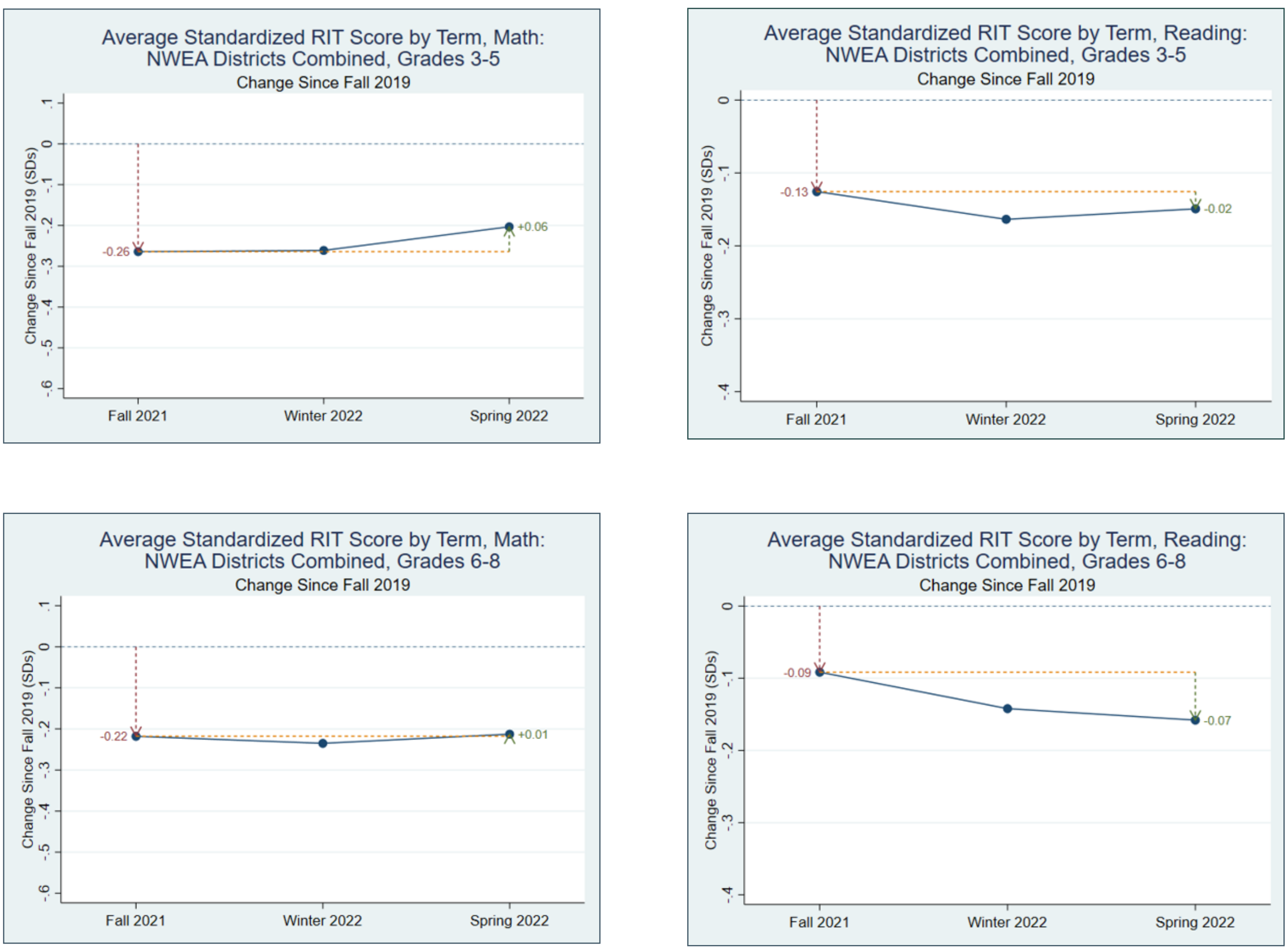
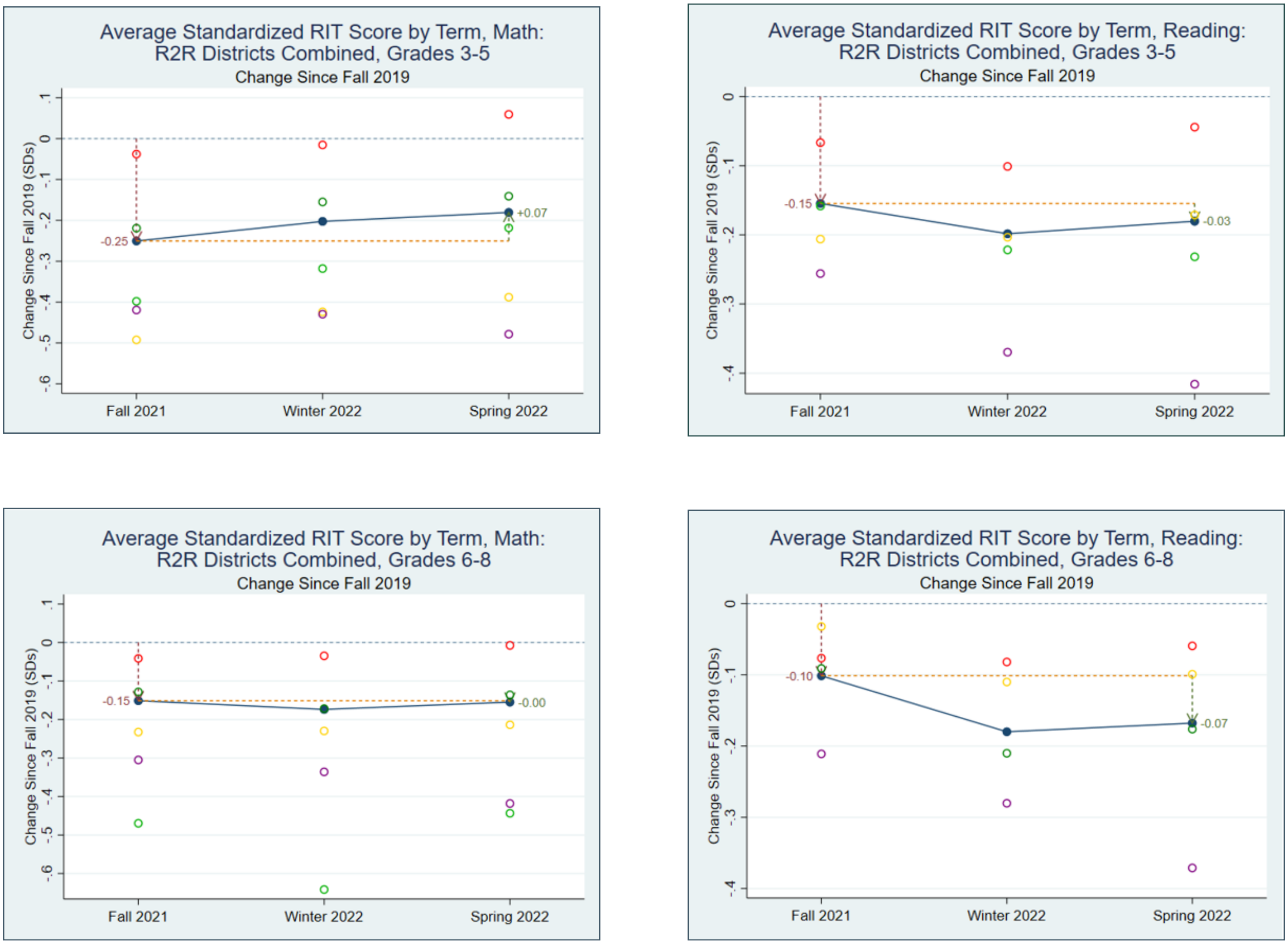


Figure 2. Standardized Achievement Loss and Recovery Cont.

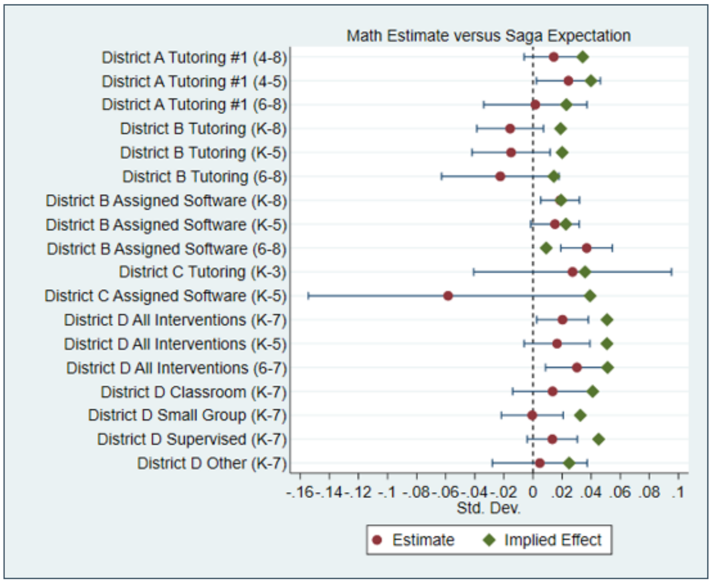
B. R2R Districts Combined



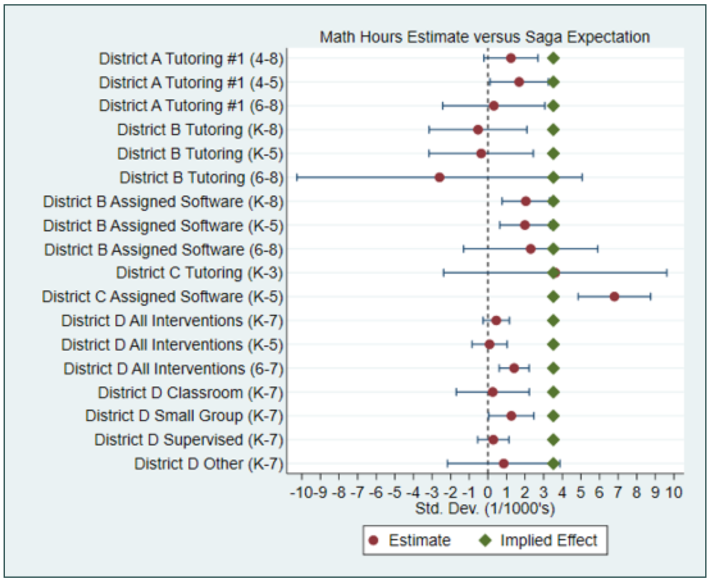
Note: Panel A reports the mean score by grade and subject for the full NWEA sample, standardized by testing period using pre-pandemic norms. The results imply that students erased .06 s.d. of the fall 2021 loss in grade 3-5 math, but fell somewhat further behind in reading. In Panel B, the hollow dots represent standardized score means in each of the five R2R districts for which we estimate achievement loss and recovery; the solid connected dots represent the five districts combined, weighted by student enrollment. When combined, the results are similar to the full NWEA sample, although the magnitude of the losses and the pace of the recovery varied by district.

Figure 3. Estimated Treatment Effect of Math Interventions

A. Impact estimates for binary measure of treatment



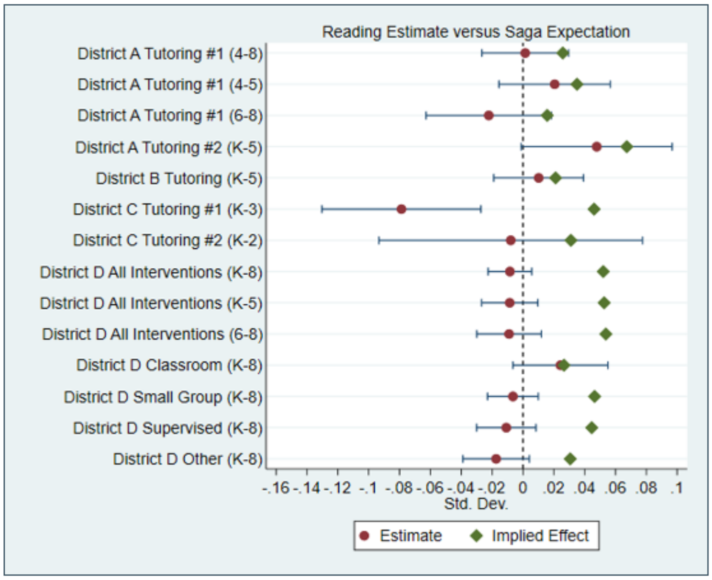
B. Impact estimates for hourly measure of treatment



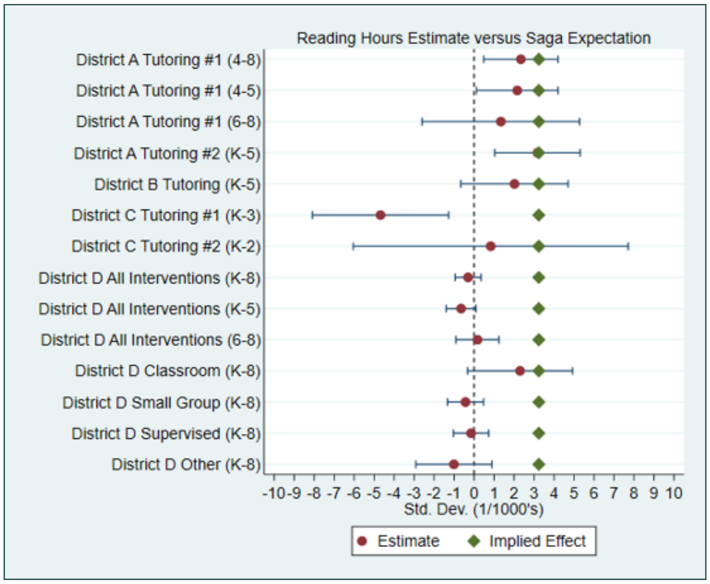
Note: Point estimates (red dot) show the average effect of receiving any amount of math intervention (panel A) or one hour of math intervention (panel B) in a given term on math MAP Growth scores at the end of that term. For all districts aside from District C the model used is a stacked model with a fall and spring term for each student; models for District C include a fall term only. Covariates in the model include participation indicators for other math interventions and reading interventions, prior MAP and state testing (when available) in both math and reading, student demographics, indicators for the calendar week that testing took place for baseline and outcome MAP Growth tests, and school-grade-term fixed effects. When applicable, models also include indicators for a student scoring below a certain MAP Growth threshold at baseline for interventions where eligibility is based on a MAP Growth score cutoff. Blue lines indicate 95% confidence intervals. The implied effect (green diamond) in panel A is calculated as the average dosage received by treated students in hours, multiplied by the estimated average effect of high dosage tutoring in math (0.38 standard deviations per 108 hours of tutoring) according to the meta-analysis by Nickow et al. (2020). The implied effect in panel B is 0.38 standard deviations in accordance with the same literature. For District D, All Interventions refers to the combination of all interventions included in the Classroom, Small Group, Supervised, and Other categories.

Figure 4. Estimated Treatment Effect of Reading Interventions

A. Impact estimates for binary measure of treatment



B. Impact estimates for hourly measure of treatment



Note: Point estimates (red dot) show the average effect of receiving any amount of reading intervention (panel A) or one hour of reading intervention (panel B) in a given term on reading MAP Growth scores at the end of that term. For all districts aside from District C the model used is a stacked model with a fall and spring term for each student; models for District C include a fall term only. Covariates in the model include participation indicators for other reading interventions and math interventions, prior MAP and state testing (when available) in both math and reading, student demographics, indicators for the calendar week that testing took place for baseline and outcome MAP Growth tests, and school-grade-term fixed effects. When applicable, models also include indicators for a student scoring below a certain MAP Growth threshold at baseline for interventions where eligibility is based on a MAP Growth score cutoff. Blue lines indicate 95% confidence intervals. The implied effect (green diamond) is calculated as the average dosage received by treated students in hours, multiplied by the estimated average effect of high dosage tutoring in reading (0.35 standard deviations per 108 hours of tutoring) according to the meta-analysis by Nickow et al. (2020). The implied effect in panel B is 0.35 standard deviations in accordance with the same literature. For District D, All Interventions refers to the combination of all interventions included in the Classroom, Small Group, Supervised, and Other categories.

Table 1. Sample Demographics

	R2R Districts	Nationwide NWEA Districts	U.S. Public School
Average school enrollment	614	467	472
% FRPL	66%	54%	55%
% Asian	3%	4%	4%
% Hispanic	39%	21%	25%
% Black	24%	16%	15%
% White	28%	52%	49%
% City	83%	29%	28%
% Suburb	13%	28%	28%
% Town	0%	11%	12%
% Rural	4%	31%	32%

Note: FRPL=free or reduced priced lunch. The source of the variables is the Common Core of Data (CCD) collected by the National Center for Education Statistics during the 2019-2020 school year.

Table 2. Math Recovery Analysis Sample Demographics by Test Term

		Percent Black					Percent Hispanic					Percent White					Percent male				
		Enr.	Tested				Enr.	Tested				Enr.	Tested				Enr.	Tested			
Sample	Grade	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22
NWEA Districts Nationwide	3-8	19%	16%	16%	16%	16%	23%	19%	19%	19%	20%	58%	50%	49%	49%	49%	57%	51%	51%	51%	51%
NWEA Districts Nationwide	3-5	18%	16%	15%	16%	16%	22%	19%	19%	19%	19%	57%	49%	49%	49%	49%	56%	51%	51%	51%	51%
NWEA Districts Nationwide	6-8	20%	16%	16%	16%	16%	23%	18%	19%	19%	20%	59%	50%	49%	49%	49%	58%	51%	51%	51%	51%
R2R Recovery Districts	3-8	20%	19%	18%	18%	18%	59%	60%	59%	60%	60%	16%	16%	16%	16%	16%	51%	51%	51%	51%	51%
R2R Recovery Districts	3-5	20%	19%	17%	18%	18%	59%	59%	58%	59%	60%	16%	16%	17%	17%	17%	51%	51%	51%	51%	51%
R2R Recovery Districts	6-8	20%	19%	18%	19%	19%	59%	60%	60%	61%	61%	16%	16%	15%	15%	16%	51%	51%	51%	51%	51%
R2R Districts Minimum	3-8	0%	0%	0%	0%	0%	11%	0%	0%	4%	0%	1%	1%	1%	1%	1%	49%	48%	47%	47%	47%
R2R Districts Minimum	3-5	0%	0%	0%	0%	0%	11%	0%	0%	6%	0%	1%	1%	1%	1%	1%	48%	47%	47%	46%	46%
R2R Districts Minimum	6-8	0%	0%	0%	0%	0%	12%	0%	0%	0%	0%	1%	1%	1%	1%	1%	50%	48%	48%	48%	49%
R2R Districts Maximum	3-8	36%	35%	31%	31%	31%	96%	96%	96%	96%	96%	48%	65%	60%	57%	60%	51%	51%	51%	51%	51%
R2R Districts Maximum	3-5	35%	34%	29%	29%	29%	96%	96%	96%	96%	96%	48%	63%	61%	56%	60%	52%	51%	51%	51%	52%
R2R Districts Maximum	6-8	37%	36%	33%	32%	32%	96%	96%	96%	96%	96%	51%	68%	60%	59%	59%	52%	51%	51%	51%	51%

Note: R2R = Road to Recovery. Enr=Enrolled. F19 = fall 2019, F21 = fall 2021, W22 = winter 2022, and S22 = spring 2022. The minimums and maximums for the R2R Recovery districts are presented in lieu of the demographics for each district to preserve district anonymity

Table 3. Reading Recovery Analysis Sample Demographics by Test Term

		Percent Black					Percent Hispanic					Percent White					Percent male				
		Enr.	Tested				Enr.	Tested				Enr.	Tested				Enr.	Tested			
Sample	Grade	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22	F19	F19	F21	W22	S22
NWEA Districts Nationwide	3-8	19%	16%	16%	16%	16%	22%	18%	18%	18%	18%	59%	51%	50%	50%	49%	57%	51%	51%	51%	51%
NWEA Districts Nationwide	3-5	18%	16%	16%	16%	16%	21%	17%	17%	17%	18%	59%	51%	50%	50%	50%	56%	51%	51%	51%	51%
NWEA Districts Nationwide	6-8	20%	16%	16%	16%	16%	23%	18%	19%	19%	19%	60%	50%	49%	49%	49%	58%	51%	51%	51%	51%
R2R Recovery Districts	3-8	20%	19%	18%	18%	18%	55%	56%	56%	56%	56%	19%	20%	19%	20%	20%	51%	51%	51%	51%	51%
R2R Recovery Districts	3-5	19%	19%	17%	17%	17%	47%	46%	46%	46%	46%	26%	27%	28%	28%	28%	51%	51%	51%	51%	51%
R2R Recovery Districts	6-8	20%	18%	18%	19%	19%	60%	61%	61%	61%	61%	15%	15%	15%	15%	15%	51%	50%	51%	51%	51%
R2R Districts Minimum	3-8	0%	0%	0%	0%	0%	11%	0%	0%	6%	0%	1%	1%	1%	1%	1%	48%	47%	47%	46%	46%
R2R Districts Minimum	3-5	0%	0%	0%	0%	0%	11%	0%	0%	0%	0%	1%	1%	1%	1%	1%	48%	48%	47%	47%	47%
R2R Districts Minimum	6-8	0%	0%	0%	0%	0%	12%	11%	13%	13%	13%	1%	1%	1%	1%	1%	50%	50%	51%	51%	51%
R2R Districts Maximum	3-8	36%	35%	31%	30%	31%	96%	96%	96%	96%	96%	48%	63%	61%	56%	60%	52%	51%	51%	51%	51%
R2R Districts Maximum	3-5	35%	34%	29%	29%	29%	96%	96%	96%	96%	96%	48%	68%	60%	59%	60%	52%	51%	51%	51%	52%
R2R Districts Maximum	6-8	37%	37%	33%	32%	32%	96%	96%	96%	96%	96%	47%	48%	50%	49%	50%	52%	51%	51%	51%	51%

Note: R2R = Road to Recovery. Enr=Enrolled. F19 = fall 2019, F21 = fall 2021, W22 = winter 2022, and S22 = spring 2022. The minimums and maximums for the R2R Recovery districts are presented in lieu of the demographics for each district to preserve district anonymity.

Table 4. Estimated Achievement Loss and Recovery, Math, Grades 3–8

District	Grade	Mean normed MAP score: Fall 2019	Mean normed MAP score: Fall 2021	Mean normed MAP score: Winter 2022	Mean normed MAP score: Spring 2022	Changes over time: Fall 2019 to Fall 2021	Changes over time: Fall 2021 to Win 2022	Changes over time: Fall 2021 to Spr 2022	Hours to eliminate the loss: As of Fall 2021	Hours to eliminate the loss: As of Spring 2022
R2R districts combined	3-8	-0.20	-0.40	-0.39	-0.37	-0.20	0.01	0.03	57.70	48.06
	3-5	-0.18	-0.43	-0.38	-0.36	-0.25	0.05	0.07	71.15	51.35
	6-8	-0.22	-0.37	-0.40	-0.38	-0.15	-0.02	0.00	42.94	43.96
District 1	3-8	0.11	0.07	0.09	0.14	-0.04	0.01	0.07	11.29	-7.44
	3-5	0.14	0.11	0.13	0.20	-0.04	0.02	0.10	10.73	-16.88
	6-8	0.08	0.04	0.05	0.08	-0.04	0.01	0.03	11.70	2.07
District 2	3-8	**-0.20**	-0.38	-0.37	-0.35	-0.18	0.01	0.04	51.50	41.03
	3-5	-0.16	-0.38	-0.32	-0.30	-0.22	0.06	0.08	62.25	40.02
	6-8	-0.26	-0.38	-0.43	-0.39	-0.13	-0.04	-0.01	36.44	38.62
District 3	3-8	-0.31	-0.67	-0.64	-0.61	-0.36	0.03	0.06	101.74	85.35
	3-5	-0.28	-0.77	-0.70	-0.67	-0.49	0.07	0.10	139.94	110.26
	6-8	-0.34	-0.58	-0.57	**-0.56**	-0.23	0.00	0.02	66.07	60.68
District 4	3-8	0.31	-0.13	-0.15	-0.01	-0.44	-0.02	0.12	123.95	90.71
	3-5	0.23	-0.17	-0.09	0.01	-0.40	0.08	0.18	113.10	62.06
	6-8	0.41	**-0.06**	**-0.23**	-0.03	-0.47	-0.17	0.03	133.41	126.02
District 5	3-8	-0.54	-0.90	-0.93	-0.99	-0.37	-0.02	-0.08	104.80	128.50
	3-5	-0.60	-1.02	-1.03	-1.08	-0.42	-0.01	-0.06	119.15	135.97
	6-8	-0.46	**-0.76**	**-0.80**	**-0.88**	-0.30	-0.03	-0.11	86.66	118.80

Note: Mean normed MAP scores in **bold** indicate cells where the percent of students tested included between 60 and 80 percent of enrolled students; all other cells include at least 80 percent of enrolled students.

Table 5. Estimated Achievement Loss and Recovery, Reading, Grades 3–8

District	Grade	Fall 2019	Fall 2021	Winter 2022	Spring 2022	Changes over time: Fall 2019 to Fall 2021	Changes over time: Fall 2021 to Win 2022	Changes over time: Fall 2021 to Spr 2022	Hours to eliminate the loss: As of Fall 2021	Hours to eliminate the loss: As of Spring 2022
		Mean normed MAP score								
R2R districts combined	3-8	-0.23	-0.35	-0.42	-0.40	-0.12	-0.07	-0.05	38.22	54.03
	3-5	-0.14	-0.29	-0.34	-0.32	-0.15	-0.04	-0.03	47.68	55.61
	6-8	-0.28	-0.38	-0.46	-0.45	-0.10	-0.08	-0.07	31.27	51.76
District 1	3-8	0.17	0.10	0.08	0.12	-0.07	-0.02	0.02	22.18	16.05
	3-5	0.20	0.14	0.10	0.16	-0.07	-0.03	0.02	20.53	13.61
	6-8	0.13	0.06	0.05	0.07	-0.08	-0.01	0.02	23.62	18.31
District 2	3-8	**-0.40**	-0.49	-0.61	-0.58	-0.09	-0.12	-0.09	28.09	54.38
	6-8	**-0.40**	-0.49	-0.61	-0.58	-0.09	-0.12	-0.09	28.09	54.38
District 3	3-8	-0.36	-0.47	-0.51	**-0.49**	-0.12	-0.04	-0.02	35.66	40.80
	3-5	-0.31	-0.51	-0.51	-0.48	-0.21	0.00	0.03	63.59	52.81
	6-8	-0.40	-0.44	**-0.51**	**-0.50**	-0.03	-0.08	-0.07	9.89	30.59
District 4	3-8	0.26	0.10	0.04	0.03	-0.16	-0.06	-0.07	48.84	71.47
	3-5	0.26	0.10	0.04	0.03	-0.16	-0.06	-0.07	48.84	71.47
District 5	3-8	-0.40	-0.64	-0.73	-0.80	-0.24	-0.09	-0.16	72.94	122.66
	3-5	-0.44	-0.70	-0.81	-0.86	-0.26	-0.11	-0.16	78.96	128.33
	6-8	-0.35	**-0.56**	**-0.63**	**-0.72**	-0.21	-0.07	-0.16	65.10	114.49

Note: Mean normed MAP scores in bold indicate cells where the percent of students tested included between 60 and 80 percent of enrolled students; all other cells include at least 80 percent of enrolled students.

Table 6. Program Usage Across Sample Districts

	Tutoring	Push-in/Pull-out	Out-of-School Time	Virtual Tools	Extended Calendar
District A	X		X		
District B	X	X			
District C	X	X			X
District D			X		
District E	X	X	X	X	
District F	X				
District G				X	
District H	X			X	
District I		X			
District J	X	X			
District K		X	X		
District L	X		X	X	X

Table 7. Estimated Treatment Effects of Math Interventions

Panel A: Districts A and B

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Participation		Hourly			
Intervention (Grades)	Point Estimate (SE)	Placebo Estimate (SE)	Estimated Impact (SE)	Placebo Estimate (SE)	Avg Dosage (Hours)	Implied Effect from Tutoring Research
District A Tutoring #1 (4-8)	0.0143	0.0022	0.00124	-0.00059	9.72	0.0342
	(0.0104)	(0.0139)	(0.00074)	(0.00109)		
District A Tutoring #1 (4-5)	0.0244*	0.0138	0.00168*	-0.00036	11.32	0.0398
	(0.0112)	(0.0191)	(0.00080)	(0.00119)		
District A Tutoring #1 (6-8)	0.0016	-0.0054	0.00032	-0.00052	6.53	0.0230
	(0.0181)	(0.0194)	(0.00140)	(0.00230)		
District B Tutoring (K-8)	-0.0157	-0.0034	-0.00053	-0.00125	5.40	0.0190
	(0.0117)	(0.0142)	(0.00134)	(0.00131)		
District B Tutoring (K-5)	-0.0151	-0.0048	-0.00036	-0.00179	5.70	0.0201
	(0.0137)	(0.0169)	(0.00143)	(0.00130)		
District B Tutoring (6-8)	-0.0225	-0.0066	-0.00260	0.00129	4.06	0.0143
	(0.0206)	(0.0242)	(0.00391)	(0.00379)		
District B Assigned Software (K-8)	0.0186**	0.0159	0.00204**	0.00216**	5.49	0.0193
	(0.0068)	(0.0084)	(0.00065)	(0.00078)		
District B Assigned Software (K-5)	0.0151	0.0032	0.00199**	0.00223*	6.42	0.0226
	(0.0085)	(0.0106)	(0.00069)	(0.00094)		
District B Assigned Software (6-8)	0.0369**	0.037**	0.00230	0.00162*	2.58	0.0091
	(0.0090)	(0.0123)	(0.00184)	(0.00078)		

* p<0.05 ** p<0.01

Panel B: Districts C and D

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Participation		Hourly			
Intervention (Grades)	Point Estimate (SE)	Placebo Estimate (SE)	Estimated Impact (SE)	Placebo Estimate (SE)	Avg Dosage (Hours)	Implied Effect from Tutoring Research
District C Tutoring (K-3)	0.0272	0.0328	0.00362	0.00269	10.18	0.0358
	(0.0347)	(0.0299)	(0.0031)	(0.0025)		
District C Assigned Software (K-5)	-0.0584	-0.19**	0.0068**	-0.000296	11.16	0.0393
	(0.0490)	(0.0520)	(0.0010)	(0.0009)		
District D All Interventions (K-7)	0.0203*	0.0324**	0.00045	0.00098*	14.48	0.0509
	(0.0091)	(0.0096)	(0.00036)	(0.00041)		
District D All Interventions (K-5)	0.0165	0.0252*	0.00009	0.0012*	14.45	0.0508
	(0.0116)	(0.0114)	(0.00048)	(0.00051)		
District D All Interventions (6-7)	0.0302**	0.0543**	0.00141**	0.00064	14.57	0.0513
	(0.0110)	(0.0168)	(0.00041)	(0.00069)		
District D Classroom (K-7)	0.0135	0.0243	0.00026	0.00162	9.26	0.0326
	(0.0140)	(0.0152)	(0.00100)	(0.00097)		
District D Small Group (K-7)	-0.0005	0.0089	0.00126*	0.00069	11.67	0.0411
	(0.0109)	(0.0136)	(0.00062)	(0.00073)		
District D Supervised (K-7)	0.0133	0.02*	0.00029	0.00081	12.83	0.0451
	(0.0088)	(0.0099)	(0.00043)	(0.00055)		
District D Other (K-7)	0.0047	0.0223	0.00085	0.00222	7.07	0.0249
	(0.0166)	(0.0172)	(0.00154)	(0.00161)		

* p<0.05 ** p<0.01

Note: Point estimates show the average effect of receiving any amount of math intervention in a given term on math MAP Growth scores at the end of that term, and the average effect of receiving one hour of math intervention. For all districts aside from District C the model used is a stacked model with a fall and spring term for each student; models for District C include a fall term only. Covariates in the model include participation indicators for other math interventions and reading interventions, prior MAP and state testing (when available) in both math and reading, student demographics, indicators for the calendar week that testing took place for baseline and outcome MAP Growth tests, and school-grade-term fixed effects. When applicable, models also include indicators for a student scoring below a certain MAP Growth threshold at baseline for interventions where eligibility is based on a MAP Growth score cutoff. Placebo estimates show the effect of the any amount of math intervention on MAP Growth reading scores, using the same model specifications. Average dosage indicates the average number of hours treated students received the intervention for each term. Implied effect from tutoring research is calculated as the average dosage in hours multiplied by the estimated average effect of high dosage tutoring in math (0.38 standard deviations per 108 hours of tutoring) according to the meta-analysis by Nickow et al. (2020). For District D, All Interventions refers to the combination of all interventions included in the Classroom, Small Group, Supervised, and Other categories.

Panel A: Districts A, B, and C

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Participation		Hourly			
Intervention (Grades)	Point Estimate (SE)	Placebo Estimate (SE)	Estimated Impact (SE)	Placebo Estimate (SE)	Avg Dosage (Hours)	Implied Effect from Tutoring Research
District A Tutoring #1 (4-8)	0.0015	0.0203*	0.00234*	0.00227**	7.99	0.0259
	(0.0143)	(0.0093)	(0.00095)	(0.00073)		
District A Tutoring #1 (4-5)	0.0206	0.0168	0.00216*	0.00141	10.82	0.0351
	(0.0184)	(0.0144)	(0.00104)	(0.00088)		
District A Tutoring #1 (6-8)	-0.0222	0.024*	0.00134	0.00497**	4.86	0.0157
	(0.0207)	(0.0119)	(0.00201)	(0.00103)		
District A Tutoring #2 (K-5)	0.0478	-0.0281	0.003169**	-0.00088	20.80	0.0674
	(0.0250)	(0.0234)	(0.00109)	(0.00097)		
District B Tutoring (K-5)	0.0102	0.0008	0.00202	-0.00018	6.52	0.0211
	(0.0148)	(0.0136)	(0.00137)	(0.00148)		
District C Tutoring #1 (K-3)	-0.0788**	-0.023	-0.00468**	-0.00148	14.23	0.0461
	(0.0263)	(0.0231)	(0.0017)	(0.0015)		
District C Tutoring #2 (K-2)	-0.00791	-0.000704	0.000834	0.00236	9.60	0.0311
	(0.0436)	(0.0361)	(0.0035)	(0.0032)		

* p<0.05 ** p<0.01

Panel B: District D

Intervention (Grades)	(1) Any Participation: Point Estimate (SE)	(2) Any Participation: Placebo Estimate (SE)	(3) Hourly: Estimated Impact (SE)	(4) Hourly: Placebo Estimate (SE)	(5) Avg Dosage (Hours)	(6) Implied Effect from Tutoring Research
District D All Interventions (K-8)	-0.0085	-0.0007	-0.00030	-0.00004	16.07	0.0521
	(0.0072)	(0.0074)	(0.00033)	(0.00025)		
District D All Interventions (K-5)	-0.0086	-0.0032	-0.00065	-0.00028	16.26	0.0527
	(0.0093)	(0.0091)	(0.00038)	(0.00033)		
District D All Interventions (6-8)	-0.0090	0.0061	0.00017	0.00039	16.60	0.0538
	(0.0107)	(0.0125)	(0.00055)	(0.00035)		
District D Classroom (K-8)	0.0243	-0.0036	0.00230	0.00017	8.23	0.0267
	(0.0157)	(0.0112)	(0.00134)	(0.00098)		
District D Small Group (K-8)	-0.0065	-0.0105	-0.00043	-0.00024	14.35	0.0465
	(0.0084)	(0.0078)	(0.00046)	(0.00037)		
District D Supervised (K-8)	-0.0108	0.0128	-0.00015	0.00052	13.78	0.0446
	(0.0098)	(0.0089)	(0.00045)	(0.00036)		
District D Other (K-8)	-0.0174	-0.0197	-0.00101	-0.00124	9.48	0.0307
	(0.0110)	(0.0118)	(0.00097)	(0.00088)		

* p<0.05 ** p<0.01

Note: Point estimates show the average effect of receiving any amount of reading intervention in a given term on reading MAP Growth scores at the end of that term, and the average effect of receiving one hour of reading intervention. For all districts aside from District C the model used is a stacked model with a fall and spring term for each student; models for District C include a fall term only. Covariates in the model include participation indicators for other reading interventions and math interventions, prior MAP and state testing (when available) in both math and reading, student demographics, indicators for the calendar week that testing took place for baseline and outcome MAP Growth tests, and school-grade-term fixed effects. When applicable, models also include indicators for a student scoring below a certain MAP Growth threshold at baseline for interventions where eligibility is based on a MAP Growth score cutoff. Placebo estimates show the effect of the any amount of reading intervention on MAP Growth math scores, using the same model specifications. Average dosage indicates the average number of hours treated students received the intervention for each term. Implied effect from tutoring research is calculated as the average dosage in hours multiplied by the estimated average effect of high dosage tutoring in reading (0.35 standard deviations per 108 hours of tutoring) according to the meta-analysis by Nickow et al. (2020). For District D, All Interventions refers to the combination of all interventions included in the Classroom, Small Group, Supervised, and Other categories.

Table 9. Supplemental Implementation Interviews and Intervention Programs

	Number of Participants	Intervention Programs
District 1	3	• Tutoring (reading and math) • Small group pull-out intervention (reading) • Extended year
District 2	3	• Tutoring (reading and math) • Small group pull-out intervention (reading)
District 3	3	• Tutoring (reading and math) • Virtual learning program intervention (reading and math)

APPENDIX A. Methods

Intervention Impacts—Variations in Estimation Models

For some districts, program participation was related not only to baseline characteristics and prior scores on the MAP Growth assessments, but also to other measures of prior achievement, such as state standardized test scores. In these cases, we estimated the following value-added model on a semester-by-semester basis, treating each semester, t, as a separate observation:

$$MAP_{igjts} = \alpha_0 + \alpha_1 Treatment_{igts} + \alpha_2 Eligible_{igts} + \ priorMAP_{igts}\gamma_{gts}$$

$$+ priorStatetest_{igts}\beta_{gts} + X_{igt}\theta_{gt} + \delta_{jgts} + \epsilon_{igts}$$

This model differs from our general model described in the main text in that it includes $priorStatetest_{igts}$, a matrix with a cubic function of previous state standardized test scores in the same subject, interacted with grade level. Moreover, the availability of an additional measure of prior student performance enabled us to slightly loosen the sample restrictions for students who could be included in the estimation. Specifically, for observations of second semester program participation (i.e., interventions that took place in spring 2022), we included students so long as they had non-missing winter and spring MAP Growth test scores from 2021–22 and at least one from either a fall 2021 MAP Growth score or spring 2021 state standardized test. In models of this specification, we imputed missing prior test scores using the district-level mean and included a dummy indicating imputation, interacted with all test score variables.

Placebo Tests

The validity of value-added models hinges on successfully accounting for all relevant factors that differ among treated and untreated students and that could also be associated with student outcomes. One standard way to test for bias in our models is to model a treatment's impact on alternative student outcomes that should not be affected by treatment participation. Thus, for each of the math interventions we examined, we also examined impacts on reading scores. Similarly, for reading interventions, we examined impacts on math performance. In most cases, we could not reject the hypothesis that the impact on the untargeted subject was equal to zero, as would be expected in a well- specified model. However, in some cases, such as the combined effect of participating in any math-focused interventions in District D (i.e. "All Interventions"), we estimated a positive impact on math achievement, as well as a positive impact on reading achievement. Such a pattern implies selection into the math interventions based on unobserved factors for which we could not control.

Regression Discontinuity Design

In three of the four districts in which we examined treatment effects of individual interventions, student assignment to one or more interventions was at least partially based on receiving a MAP Growth score below a certain threshold in a previous term. Assignment to treatment based on a cutoff in a continuous variable (here, MAP Growth assessment scores) often provides the opportunity to estimate treatment effects using a regression discontinuity design. We explored this methodology in all three districts, but ultimately found very weak first stages in all of them. These findings were corroborated by our district interviews, from which we learned that students' assignment to particular interventions was frequently influenced by teacher discretion, scheduling issues, and instructional capacity. We plan to continue exploring the viability of using regression discontinuity designs in future analysis, as we expect that R2R districts' program implementation will become more systematic over time and intervention assignment may adhere more to their intended design.

Interviews

Interviews lasted for 60 to 70 minutes and included up to two administrators in a session. (The limited number of respondents at each site may have led us to miss some important contextual factors.) We followed a semi structured protocol in which we asked district administrators to discuss how the program was carried out during the year in terms of student eligibility, dosage, staffing, delivery, and content. We also asked them to describe how actual implementation differed from their plans and the challenges they faced. During the interviews, we shared data on program participation, dosage, and outcomes to inform them of trends we were seeing in the outcomes analysis and to prompt them to think about explanations for these results.

After each interview, we completed an interview outtake form that summarized what we had learned for each descriptive category of our interview protocol (e.g., program dosage plans, program dosage implementation, challenges faced, reflections for next year). We then wrote case memos for each district to document emerging findings from the outtake forms from interview participants across each district, pulling in quotes from cleaned interview transcripts to substantiate findings and establish a chain of evidence to support our claims. These memos focused primarily on explaining results observed in the outcomes analysis of this report and to consider how the district's approach to program eligibility, dosage, providers, delivery, and content might explain their results. Next, we met as a research team to review the case memos for each district and document cross-case themes across districts and interventions.

Our findings from these interviews have important limitations. Most notably, we spoke to a limited number of district administrators from three districts participating in the Road to Recovery network. As such, our implementation findings do not capture the full range of conditions faced by districts, especially for those that were unable to start interventions in the 2021–22 school year. Moreover, even among districts where we conducted interviews, we did not capture the perspective of multiple stakeholders involved in the implementation of interventions, especially those who worked in schools, such as building leaders and teachers.

Appendix B. Description of Programs by Program Type

Tutoring

Across the consortium of participating districts, eight districts confirmed that they had implemented a "tutoring program" during the 2021–22 school year. The initiatives identified across these eight districts as tutoring programs varied on all seven of the following dimensions: which students the program targeted, when the tutoring happened, whether it was virtual or in person, the qualifications and backgrounds of the tutors, the student–tutor ratio, and the frequency and duration of tutoring sessions. Most tutoring programs were centered on math and/or reading but occasionally also supported other school subjects.

Targeting. Tutoring programs varied substantially in the students they targeted. Programs ranged from targeting a subset of students in particular grades at particular schools to being available on demand to all students in the district. For programs that targeted subsets of students, districts set criteria for identifying the students that each program intended to serve. These criteria typically included one or multiple test scores (e.g., scoring below the 25th percentile on the MAP Growth test) and, in many cases, also included other measures (e.g., attendance, course performance, teacher recommendations) to identify students whom the district considered to be most in need of academic support. Districts using the Multi-Tiered System of Supports (MTSS) framework to provide targeted support for struggling students also often used a student's "tier" classification to assign them to interventions in each subject.

Scheduling. When the sessions happened and what they were replacing for students varied across and, in some cases, within programs. Four districts designed programs that would happen primarily or only after school and/or on Saturdays, two districts' programs primarily or exclusively occurred during the school day, and two districts had programs that regularly provided sessions during the school day and outside of the school day. For sessions that occurred during the school day, some districts explicitly mentioned that these sessions were planned for when the student would have had an elective class or would not be not missing core instruction, but that was not always the case.

Instructional modality. We also found differences in the instructional modality of tutoring sessions across and, in some cases, within programs. Whereas three districts offered sessions primarily or only in person, three districts' programs were entirely virtual (typically conducted using a video conferencing software like Zoom), and two districts had a mix of virtual and in-person tutoring. Among districts that offered a mix, some had both a virtual program and an in-person program, while others had one program that matched students to either a virtual or in-person tutor.

Providers. The providers of tutoring ranged widely across and within districts. Five districts contracted with external vendors, typically national companies (e.g., Paper, FEV Tutor) or nonprofits (e.g., Saga Education) or local nonprofit organizations, who screened, hired, and trained the tutors. Alternatively, three districts managed the hiring and training process internally. These districts all approached this process differently: one hired for new full-time tutoring positions in the district; one recruited (and paid) their own staff to provide tutoring in addition to their existing responsibilities; and one recruited people with a variety of backgrounds to be tutors, such as their own staff, local college students, parents, and/or their own high school students. Student–provider ratio. Sessions also varied in their tutor-to-student ratios within and across districts. Although some programs provided 1:1 instruction, the majority aimed for tutors to work with three to five students per session. In one district, however, the maximum tutoring group size was 10 students. One-on-one sessions were particularly common among virtual tutoring programs. Student absences also frequently resulted in sessions that included fewer students than the intended number or maximum allowed.

Dosage. Among programs that were not on demand, the intended dosage for students enrolled in tutoring ranged from two to five sessions per week, totaling 1–4 hours per week. Districts' programs started at different points of the year; some did not start until March 2022, so the total intended tutoring time over the year varied widely, from approximately 15–100 hours. The length of tutoring sessions also varied among districts, schools, and students, ranging from 15 minutes or less for on-demand tutoring services to more than 60 minutes for more targeted tutoring programs. In some cases, student participation in tutoring sessions was optional, which meant that there was also variation in the amount of tutoring that each student received in every district.

Small-Group Push-In and Pull-Out Interventions

Six districts also described implementing small-group interventions, which they often called “intervention time” or “intervention rooms,” designed to provide students with push-in or pull-out reading and math support during school hours. Districts did not consider these programs to be “tutoring” programs, but many of their implementation features were similar to tutoring programs. The main distinction was that these programs were provided only in person; during school hours; and delivered exclusively by certified teachers, specialists, and/or support staff trained to deliver the interventions and employed by the district. The other implementation features of these interventions varied across districts and sometimes between schools within districts.

Targeting. Targeting policies for small-group push-in and pull-out interventions were structured similarly to those for tutoring programs. The small-group interventions varied in whether they were targeted to only a subset of grades and whether they were offered only at a subset of schools. Two districts chose to have their intervention(s) serve the schools with the highest percentages of Black students. Like tutoring programs, most intervention programs relied on a test score thresholds and occasionally also used other indicators like attendance to target programming for students they thought were most in need of academic support. One district, however, targeted students who were not performing on grade level, but who were also above a certain performance threshold. Moreover, in some cases, students who received other specialized services during school hours (i.e., special education services or English learner services) were ineligible for interventions even if they met the other eligibility criteria because their other services conflicted with the timing of the interventions.

Scheduling. All but one of these interventions were “pull-out” interventions, and students were meant to be pulled out of classes that were not providing instruction in the same subject in five of six districts. The other district preferred that students not be pulled out of core instruction but lamented that this was not always possible because of scheduling challenges. In addition to their pull-out intervention, one district had a “push-in” intervention for reading that enabled the interventionist to work with a small group of students in the classroom while other students did “reading stations” with the classroom teacher. Student–provider ratio. Perhaps the greatest source of variation in the implementation of these interventions was the student–interventionist ratio. Ratios were typically smaller for elementary school students, ranging from 1:1 to 1:7. For programs that targeted middle school students, the student–interventionist ratios were larger and more variable, ranging from 1:7 to 1:25.

Dosage. Districts also varied in when they launched these interventions during the school year and how much of the interventions they provided to students. Most interventions began at the beginning of the school year, but one started in October, and one did not start until February. Students received the interventions 4–5 days per week for 25–50 minutes/day. Two districts capped students’ total participation in the intervention—one at approximately 30 hours and one at roughly 50 hours—to enable additional students to participate. The maximum additional instruction provided by one of these interventions was about 140 hours. In some cases, student participation in these interventions was dictated by teacher approval (e.g., teachers feeling comfortable releasing a student from class at that time), resulting in variation regarding how much intervention time students received within the districts.

After School/Out-of-School Time Programs

We also learned about five districts' after-school, before-school, and Saturday school programs. To be included in this study, these programs had to include time specifically focused on academic instruction or support (e.g., homework help). Most after-school and before-school programs provided students with additional instructional time or homework help in math and reading, but students could opt to work on other subjects (e.g., science, Spanish) or focus on enrichment activities (e.g., art, technology) during that time. Alternatively, Saturday programs typically consisted of math and reading instruction, with little or no time allocated specifically for enrichment activities.

All these programs were offered outside of traditional school hours, optional for students, and offered in-person, but they varied within and across districts in their other implementation features. In most cases, the district leaders indicated that their programs were highly decentralized; much of their implementation was determined by the individual school site. Indeed, district administrators often did not have complete information on differences in implementation across school sites. Moreover, for most of these programs, student-level eligibility and attendance data were not tracked consistently by the district or school sites. We note that some of these programs historically had participation fees for students, but they were subsidized by ESSER and free of cost to families this year.

Targeting. The out-of-school time programs varied widely in their targeting strategies. Whereas three districts offered programming to students in grades K–8 at a subset of schools, three districts had after-school opportunities only for a subset of grades. For all programs, students had to opt in to participate. Four districts had at least one program open to all students in the specified schools and grades. Alternatively, three districts had programs that were available only to certain groups of students (e.g., students with disabilities, students with test scores below a certain threshold, homeless students, students invited based on teacher recommendations) within the schools offering them. Two districts had waitlists for their after-school programs.

Scheduling. Four districts had out-of-school time academic programs scheduled for after school, and three districts also or alternatively provided before-school programming at select school sites. Three districts had an out-of-school-time program on Saturdays. Apart from one smaller program that took place at a school "hub site," all out-of-school time programs occurred at students' school sites. Two districts provided regular transportation to and/or from school for students to access the programs.

Providers. Whereas credentialed teachers from the district typically led Saturday programs, the after-school and before-school programs varied on whether they were led by teachers and staff from the district, university teacher prep program students, and/or community members hired by local vendors.

Student–provider ratio. The student-provider ratios in out-of-school-time programs ranged from small-group instruction (1:6) to classroom-based instruction (1:20). There were no consistent differences in student–provider ratios among after-school, before-school, or Saturday programs.

Dosage. Districts also varied in the amount of programming they offered targeted students. Most programming began in the fall, but four districts did not begin at least one of their programs in some or all schools until the spring semester. In some cases, districts did not start a program in the fall because they were still planning it during the fall, but in the case of some of the Saturday programming, the programs were intentionally run only in the spring to prepare students for upcoming state testing. Once programs were running, they also varied in intensity across districts. After-school and before-school programs were offered from 1–5 days per week, for 30 minutes to 3.5 hours per day. The time spent focused on academics each day ranged from 30 minutes to 1.5 hours. Over the course of the year, these programs ranged in the amount of academic support or instructional time they provided to students from about 7.5 hours to about 100 hours. Alternatively, Saturday programming was more likely to be offered only in the spring and had sessions with variable durations across school sites, many of which the district administration team was unaware of. In one instance, the district team knew their program typically ran for 4.5 hours, from 8:30 a.m. to 1 p.m., but they generally did not start before the spring. Because of the inconsistency in programming and the limited duration of Saturday programming over the year, the total range of time enrolled students spent at Saturday programming is unknown.

Virtual Learning Programs

Four districts indicated that they were using virtual learning programs as a COVID recovery strategy that added academic time to students' days beyond what they received through core instruction. Although many districts used virtual learning programs as part of their core instruction, only these four districts confirmed that they had one or more virtual learning programs that added learning time for a subset of students. These districts used a variety of programs to that end, including ALEKS, Dreambox, Zearn, Achieve3000, and others. These virtual learning programs typically provided grade-level and/or below-grade-level support in math or reading. Districts varied in how they targeted students to use these programs, when students accessed them, and how much time students were expected to use them.

Targeting. Districts varied in the grades and students they targeted for virtual learning programs. Three of the four districts targeted at least one program at a subset of grades (e.g., 6–8 or 3–5), and two districts provided at least one program across Grades K–8. All but one of the programs targeted students in need of below-grade-level support. The districts identified these students on the basis of their MTSS tier (informed by test scores, grades, course performance, etc.) or teacher recommendation. Alternatively, one district required all students in a subset of schools to use their program.

Scheduling. Students accessed virtual learning programs both within and outside of school across districts. Two districts formally required their program(s) to be used by students during an intervention period in their schedule, whereas the other two districts expected students to use the program for extra practice outside of school. Across all districts, students could choose to use the program for extra practice outside of school beyond what was expected, if they so desired.

Dosage. The amount of time that students were expected to spend on the virtual learning programs varied by program. For one district's program, teachers had discretion over the amount of extra practice they assigned to students, but the district recommended about 2 hours per week. In other districts, students were expected to use the virtual learning program multiple times a week, totaling between 30 minutes and 3 hours of extra practice per week (approximately 18 to 108 hours per year).

Extended School Calendars

Two districts extended the school year at a subset of schools to provide students with additional days of instruction throughout the year. One of the districts implemented three different models of an extended school year across its participating schools. In two of the four total models, the additional days were not distinguishable from regular school days. The additional days provided through the other two models had slightly reduced instructional time in math and reading because they had a shorter school day or more time allocated for enrichment and social-emotional learning activities. In one district, the extra days were optional for students to attend, whereas they were not in the other district. All of the models provided instruction only in person, were staffed by teachers from the participating school sites, and had student–teacher ratios consistent with the schools' typical classroom ratios. The four models varied in the schools and students they targeted, when in the year the extra days occurred, and the number of additional days and amount of additional instructional time they provided to students.

Targeting. Both districts implemented these extended calendars at a subset of schools. One district selected some of the lowest performing schools (about 25% of their elementary and middle schools) for this intervention, whereas the other district selected schools for each model based on their principal's, staff's and parents' interest in an extended year calendar. This district had 20% of their elementary and middle schools on one model and 2% on the other two models. Three of the four models were designed so all students at the participating schools would receive the extra days of instruction. Alternatively, the other model invited students to attend who were most in need of academic support based on multiple measures (including test scores, attendance, course performance, etc.) and allowed additional students to opt in as space allowed. Ultimately, they were able to serve all interested students.

Scheduling. The models varied substantially regarding when the extra days occurred during the school year. One model had students begin school 5 days earlier and end school 4 days after the traditional start and end dates for other schools in their district. Six additional days for professional development were included throughout the year at the extended year schools so that the students received 3 total extra days of school over the course of the year. Two other models started school 2 weeks before and ended school 4 weeks after the traditional school calendar. They integrated the extra days into the calendar throughout the year, defining certain Wednesdays as the extra days. The last model started school 1 week before and ended school 4 weeks after the traditional school calendar. This model offered the extra days during 5 different weeks throughout the year.

Dosage. The four models also varied in the total number of extra days and amount of instructional time they gave participating students relative to the traditional school calendar. As previously mentioned, one model meant that students had 3 additional, typical days of instruction. Two other models provided students with 22 additional days of instruction, although one of the models had half-days for each of the additional days. The full-day model consistently provided students with an additional 2 hours of reading instruction and 1.5 hours of math instruction per day, totaling an extra 44 hours of reading and 33 hours of math per year relative to the traditional calendar. The half-day model was less consistent in its instructional time across days but offered up to 1.5 hours of instruction in math and reading per day, totaling a maximum possible 33 additional hours in math and reading during the year. A fourth model provided students with up to 18 additional days of school. Each day had 1.5 hours of reading instruction and 1 hour of math instruction per day, totaling a maximum of 27 additional hours of reading and 18 additional hours of math over the course of the year.