Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 24-1030

The Scaling Dynamics and Causal Effects of a District-Operated Tutoring Program

Matthew A. Kraft
Brown University

Danielle Sanderson Edwards
Old Dominion University

Marisa Cannata
Vanderbilt University

Public school systems across the U.S. have made major investments in tutoring to support students' academic recovery in the wake of the COVID-19 pandemic. We evaluate a large urban district's efforts to design, implement, and scale a district-operated, standards-based tutoring program across three years. We draw on extensive interviews and survey data to document the dynamic changes in the program as Metro Nashville Public Schools integrated core operations into its leadership and school structures, expanded tutor supply by pivoting from a volunteer to a teacher-based staffing model, and addressed scheduling constraints by offering tutoring immediately before and after school in addition to during the school day. The district steadily scaled the program across two years, delivering over 125,000 total hours of tutoring to more than 6,800 students while also increasing dosage each semester. Using a collection of experimental and quasi-experimental designs, we find consistent evidence of a small to medium average positive effect on students' reading test scores (0.04 to 0.09 standard deviations), but no average effects on math test scores or course grades in either subject. We discuss four possible explanations for these results, including a limited treatment-control contrast, modest program duration, heterogeneous effects, and miscalibrated expectations of tutoring effects at scale.

VERSION: August 2024

Suggested citation: Kraft, Matthew A., Danielle Sanderson Edwards, and Marisa Cannata. (2024). The Scaling Dynamics and Causal Effects of a District-Operated Tutoring Program. (EdWorkingPaper: 24-1030). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zcw7-4547

The Scaling Dynamics and Causal Effects of a District-Operated Tutoring Program

Matthew A. Kraft
Brown University

Danielle Sanderson Edwards
Old Dominion University

Marisa Cannata
Vanderbilt University

August 2024

Abstract

Public school systems across the U.S. have made major investments in tutoring to support students' academic recovery in the wake of the COVID-19 pandemic. We evaluate a large urban district's efforts to design, implement, and scale a district-operated, standards-based tutoring program across three years. We draw on extensive interviews and survey data to document the dynamic changes in the program as Metro Nashville Public Schools integrated core operations into its leadership and school structures, expanded tutor supply by pivoting from a volunteer to a teacher-based staffing model, and addressed scheduling constraints by offering tutoring immediately before and after school in addition to during the school day. The district steadily scaled the program across two years, delivering over 125,000 total hours of tutoring to more than 6,800 students while also increasing dosage each semester. Using a collection of experimental and quasi-experimental designs, we find consistent evidence of a small to medium average positive effect on students' reading test scores (0.04 to 0.09 standard deviations), but no average effects on math test scores or course grades in either subject. We discuss four possible explanations for these results, including a limited treatment-control contrast, modest program duration, heterogeneous effects, and miscalibrated expectations of tutoring effects at scale.

Corresponding author Matthew Kraft can be reached at mkraft@brown.edu. We thank Metro-Nashville Public Schools for their support. We are especially grateful to Grace Bailey, Sarah Chin, Krista Davis, Andrew Garland, Bradley Gellman, Erin Holser, Samantha Margolis, and Keri Randolph for their partnership and many contributions to this paper. This research was supported by grants from TN SCORE.

Introduction

A large body of evidence from randomized control trials (RCTs) documents the efficacy of tutoring programs for accelerating student learning (Dietrichson et al., 2017; Fryer, 2017; Inns et al., 2019; Nickow et al., 2024; Pellegrini et al., 2021). This research helped catalyze widespread support for investing in tutoring as a response to the disruptions to schooling caused by the COVID-19 pandemic. In January of 2022, U.S. Secretary of Education Miguel Cardona laid out an ambitious vision by, "challeng[ing] all of our district leaders to set a goal of giving every child that fell behind during the pandemic at least 30 minutes per day, three times per week, with a well-trained tutor." President Biden reinforced this priority in his 2024 State of the Union speech calling "to expand high-quality tutoring." During this time, billions of dollars in federal aid and state initiatives have flowed to districts to support their efforts to integrate tutoring into public schooling at scale (National Student Support Accelerator, 2023).

The urgency and challenges of standing-up tutoring programs coupled with requirements to quickly invest federal COVID-relief aid have led many districts to partner with third-party tutoring organizations to provide these services. Two primary and very different models of contracted tutoring have emerged. The first model provides on-demand tutoring services via text, phone, or web (often outside of school hours) and requires students to proactively use the services and bring questions for tutors to help resolve. Evidence to date suggests that these opt-in tutoring programs are rarely used by students (Barshay, 2023) For example, less than 1% of eligible students in New Mexico participated in free online tutoring outside of school after considerable recruitment efforts (Personalized Learning Initiative Research Team, 2024). Students who utilize these on-demand tutoring supports tend to be higher performing than those most in need of academic support (Robinson et al., 2022). A second model provides more structured, often online, tutoring integrated into the school day where students are matched with tutors hired, trained, and supported by third-party organizations. Research suggests that these models hold more promise (Cortes et al., 2024; Fesler et al., 2023; Kraft et al., 2022; Kraft & Lovison, 2024; Ready et al., 2024; Robinson et al., 2024). However, third-party programs are threatened by the looming end to federal aid given that contracted services with outside vendors are often the first line items to be cut when budgets are tight.

We present evidence on a third approach to scaling tutoring, building it from the ground up and integrating it into a district's core infrastructure and operating systems. In the spring of 2021, we began a research-practice partnership with the Metropolitan Nashville Public Schools (MNPS) to study the implementation and effects of their signature tutoring initiative, Accelerating Scholars. That same semester, MNPS launched a pilot tutoring program with 132 students across three schools. Two years later, MNPS had scaled the program to deliver a cumulative total of over 125,000 hours of tutoring to over 6,800 students. During this same period, the district made large gains in student achievement, ranking 3rd and 6th nationally in math and reading achievement growth among 100 large districts.[1]

Our paper is among the first studies of a large, district-designed and operated tutoring program in the post-COVID era. MNPS also provides an important case study of scaling education reforms by adapting the core program features to a specific context rather than faithfully transferring an established design to a new setting (LeMahieu, 2011; Sabelli & Harris, 2015). As we show, the Accelerating Scholars program utilizes many of the recommended design

[1] These findings are from the Education Recovery Scorecard Project (https://educationrecoveryscorecard.org/)

features hypothesized to drive high-impact tutoring (Kraft & Falken, 2021; Robinson et al., 2021), while also allowing for flexibility in other key features.

We first document how MNPS designed, implemented, and scaled Accelerating Scholars. Here we draw on a wide range of qualitative evidence collected in the field and descriptive evidence from surveys of school-based tutor leads, tutors, and students. We show that the Accelerating Scholars program changed in dynamic ways throughout the scaling process. What started as a program that relied on outside consultants and contractors became one that was increasingly absorbed into the core operating functions of new district staff members dedicated to running Accelerating Scholars. The program also evolved from a model where college volunteers tutored students online into a district-wide program staffed primarily by MNPS teachers and delivered in-person before, during, and after school. These changes reflect the limited supply of volunteer college students and community members willing to serve as tutors, the advantages afforded by in-person tutoring by staff who are integrated into the school system, and the scheduling challenges of providing tutoring exclusively during the school day.

Accelerating Scholars achieved major implementation milestones that have been elusive for many other districts operating tutoring programs at scale (Carbonari, DeArmond, et al., 2024; Carbonari, Dewey, et al., 2024). The program delivered a high degree of dosage, increasing from an average of 16 to 24 tutoring sessions per semester over two years as program size more than quadrupled. Survey data document a clear and meaningful increase in program implementation quality and curricular alignment over four semesters. Both student and tutor satisfaction increased steadily across the semesters of large-scale implementation.

We then evaluate the effect of the Accelerating Scholars tutoring program on student test scores and course grades across five semesters using a collection of experimental and quasi-experimental approaches. These analyses combine a student-level randomized trial of the pilot program, a school-level cluster randomized trial in the first semester of scale-up, and difference-in-differences/event-study models that analyze the effects across two scale-up years (2021-22 and 2022-23). We find consistent evidence across research designs that Accelerating Scholars had a small to medium positive effect on students' achievement on reading tests, on average, between 0.04 to 0.09 standard deviations (SDs). These results are likely driven by tutoring in early elementary (1^{st}-3^{rd}) grades which constituted 78% of all students tutored in reading. We find no evidence that Accelerating Scholars increased middle and high school students' academic achievement on math tests, on average, or that tutoring consistently improved teacher-assigned grades in the subjects in which students were tutored. Heterogeneity analysis reveal the program had moderate to large positive effects on test scores in the middle of the achievement distribution in MNPS (as large as 0.22 SD in reading, 0.11 SD in math), but no effects in the lower or upper achievement ranges. Conversely, positive effects on grades were limited to the very lowest performers.

The modest to null average effects raise important questions about the challenges of operating tutoring programs at scale. We explore possible explanations and find evidence that tutoring impacts were potentially attenuated due to both program and study design features including: 1) a more limited treatment-control contrast than is typically found in the research on tutoring because students in the control group worked with computer adaptive learning programs that provide personalized instruction, 2) a focus on a semester of tutoring rather than a full academic year, 3) heterogeneous effects and the misalignment between students who most benefitted from standards-based tutoring and those who participated in the program, and 4) mis-calibrated expectations for effects of tutoring programs at scale.

Our paper makes several contributions to the literature as well as policy and practice. We build on prior studies of large-scale tutoring programs (n treated >1,000) in Norway (Bonesrønning et al., 2022; Kirkebøen et al., 2021), Chile (Cabezas et al., 2011), the United Kingdom (Lloyd et al., 2015; Thurston et al., 2019), and the U.S. (Bhatt et al., 2024; Guryan et al., 2023; May et al., 2015; Robinson et al., 2024; Sirinides et al., 2018) to provide some of the first evidence concerning the implementation dynamics and impacts of a large-scale, district-designed and operated tutoring program. Such evidence is critical given the likely challenges of sustaining funding for tutoring programs that are operated by third-party providers via external contracts. Our study complements work by the Road 2 Recovery research team who evaluate districts' efforts to scale tutoring as well as other academic recovery strategies during the 2021-2022 and 2022-23 school years using a value-added framework and finds few statistically or practically significant effects (Carbonari, DeArmond, et al., 2024; Carbonari, Dewey, et al., 2024).

We also contribute to an interdisciplinary literature on the challenges of scaling promising social and public interventions (Al-Ubaydli et al., 2020; A. Banerjee et al., 2017; Coburn, 2003; Gupta et al., 2021; Honig, 2006; Manna, 2010; Muralidharan & Niehaus, 2017). Kraft et al. (2024) document a clear pattern of attenuated effectiveness, on average, as tutoring programs increase in size that may be explained by the challenges of implementing an existing program in new and diverse contexts (Chambers et al., 2013). Two related papers explore the scaling dynamics of the Match/Saga model of high-dosage tutoring. Davis et al. (2017) find that tutor quality does not decline in relationship to a rank-based measure of effectiveness informed by hiring interviews, while Bhatt et al. (2024) show that tutoring costs and staffing demands can be reduced without compromising effectiveness through the use of computer-assisted learning programs during tutoring sessions. Approaches that shift from replicating an externally developed program to developing an internal initiative that can be aligned to both core improvement ideas and existing systems hold promise for more sustainable improvements at scale (Cohen-Vogel et al., 2016, 2016; Sabelli & Harris, 2015). For example, research shows that tutoring programs delivered via mobile phones in low-income contexts can be successfully scaled by a range of providers across different countries (Angrist et al., 2023).

Our paper also provides a valuable case study that can inform districts' ongoing efforts to scale and sustain tutoring programs. We highlight the ways in which staffing and scheduling constraints become far more salient as programs scale. This suggests that best-practice design features for small-scale tutoring programs may not be feasible or optimal for programs operating on a larger scale. Finally, we examine the tradeoff districts must weigh between tutoring intensity and the number of students a program can serve when choosing design features such as dosage and student-tutor ratios.

Accelerating Scholars Program Design

MNPS is a large urban district encompassing the capital city of Tennessee and serving over 81,000 students across 128 K-12 traditional public schools (our focus) and 31 charter public schools. The district enrolls a diverse population of students. Three-quarters of MNPS students are students of color; 38% are Black, and 34% are Hispanic. Forty percent of MNPS students are considered economically disadvantaged. One out of five students receive English learner services. Students with disabilities make up 14% of the district population.

After piloting the program in the spring of 2021, MNPS launched Accelerating Scholars in the fall of 2021 with the vision of building a sustainable, evidence-based tutoring program.

MNPS Superintendent Dr. Adrienne Battle set an ambitious public goal to scale the program to serve over 7,000 students. The district anchored this signature initiative around a set of core design features while also taking a more flexible portfolio approach to other elements as the program evolved over time.

Fixed Design Features

High Dosage: Accelerating Scholars was designed with the expectation that students would receive a total of 90 minutes of tutoring per week. This initially was organized in the form of three 30-minute tutoring sessions per week but evolved to become two 45-minute sessions per week for some students.

Sustained Duration: Students participating in Accelerating Scholars did so for 8 weeks during the pilot, 10 weeks in 2021-22, and 12 weeks in 2022-23 over the course of an academic semester. In some cases, students received multiple semesters of tutoring given individual needs and tutoring capacity.

Relationships: Students worked with the same tutor over the course of a semester (and possibly beyond) to foster strong relationships and allow for more personalized instruction.

Aligned Curriculum: MNPS developed and refined a detailed set of curriculum guides and student activity workbooks for each grade and subject. These materials were designed to be aligned with Tennessee state standards and the core curricular material used by MNPS teachers. Pacing guides organized each tutoring session around five main activities: an opening conversation to build the relationship, an explanation of the purpose of the session, a warm-up activity, the core instruction, and an end-of-session reflection. See Appendix A for more details about curricular materials.

Delivered at School: All Accelerating Scholars tutoring took place in students' MNPS school buildings.

Tutor Training and Support: All tutors were required to pass a background check and complete four asynchronous modules on program goals and components, academic content, using the Accelerating Scholars online tutoring platform, and child protection and privacy guidelines. Starting in 2022, tutors that did not hold an active teaching licensure were required to also complete a one-week early reading training course to tutor K-3rd grade reading or the TN ALL Corps Tutor Training program for tutoring in 4th through 9th grade math or reading.

Flexible Design Features

Personnel: Accelerating Scholars cast a wide net to maximize tutor supply. Tutors included local undergraduate students across all majors, students enrolled in teacher preparation programs, community volunteers, retired educators, as well as MNPS teachers, paraprofessionals, and other school-based staff members. The program also incorporated paid literacy tutors coordinated by PENCIL, a local non-profit organization, and contracted with the firm Tutored by Teachers during Fall 2022 to provide secondary math tutoring.

Delivery mode: Tutoring occurred online as well as in person. Online tutoring took place on a specialized tutoring platform created and operated by GoSchoolBox. College students and volunteers largely tutored online while MNPS staff primarily tutored in person.

Scheduling: Tutoring occurred at school both during and outside of school hours. Tutoring during the day was mandatory and embedded in students' daily personalized learning time (PLT) classes when many students received individualized supports as part of the district's Multi-Tiered System of Supports (MTSS), Independent Education Plans (IEPs), English learners program, and gifted and talented program. Tutoring outside of school hours took place before and after school. Schools often targeted students whom they knew were already coming to school early or staying late. MNPS approached these families to request their permission for their student to participate in the Accelerating Scholars program outside of school hours.

Student-tutor ratios: Accelerating Scholars capped student-tutor ratios at 4:1 and differentiated ratios by tutor type such that college students and community volunteers worked 1:1 with tutees, paraprofessionals worked with 2:1, and certified teachers worked with 3 or 4 students in one session.

Student Assignment to Tutoring

The process of assigning students to receive tutoring was fluid and evolved over time. In general, MNPS targeted students who scored between the 15^{th} and 60^{th} percentile nationally on diagnostic assessments to participate in Accelerating Scholars as long as tutoring did not conflict with other specialized services offered during PLT such as MTSS Tier II or Tier III interventions, special education, or English language instruction. Students who scored below the 10^{th} and 25^{th} percentiles were eligible for Tier III or Tier II supports, respectively, as part of the district's MTSS program.

Accelerating Scholars staff provided participating schools with lists of eligible students based on prior test scores and asked them to curate the lists by removing any students who were receiving specialized services during PLT. In the early phases of the program, Accelerating Scholars staff entered eligible students from schools' curated lists into an online system in batches, prioritizing students with lower test scores. Participating tutors would select students based solely on their grade, subject, and assigned time for tutoring. As Accelerating Scholars shifted to more in-person tutoring by MNPS personnel, schools assumed a greater degree of independence in selecting students for tutoring and matching them with a tutor based on need and scheduling considerations. Eligible students who were not matched with a school-based tutor were then added to the online matching pool.

We illustrate the nature of selection into tutoring based on student performance on these diagnostic tests in Figure 1. Overall, Accelerating Scholars served students who were largely in the middle of the achievement distribution among MNPS students, but still lower performing relative to national benchmarks. In MNPS, approximately 39% and 49% of students score below the 15th and 25th national percentiles in reading, and 29% and 46% are below these respective thresholds in math. Figure 1 also illustrates that students who scored both above and below these eligibility thresholds also participated in tutoring. This is due to multiple factors, including schools selecting students below the test score threshold for additional tutoring supports outside of school hours, schools that had tutoring capacity but fewer students who fell within the eligibility range, and the use of multiple diagnostic tests and course-based benchmark tests to inform selection. Compared to non-participating students, tutored students were also more likely to be female, Black, and economically disadvantaged (see Appendix Table B1). Non-participating students were more likely to be students with IEPs and English learners, given these students were likely receiving other specialized services.

Tutors and Tutor Leads

MNPS utilized a wide range of tutor types to staff Accelerating Scholars. Using payroll records, we estimate that roughly 73% of unique tutors were MNPS employees (the sum of 50% core classroom teachers, 7% paraprofessionals, and 16% other school-based staff [e.g. instructional coaches, assistant teachers]). Records on the 27% of remaining non-MNPS employees are less reliable, but rough estimates suggest these were composed of 12% university students, 7% community volunteers, 2% retired educators, 1% PENCIL early literacy tutors, 1%

Tutored by Teachers tutors, and the remaining 4% unknown. Among non-MNPS employees who served as tutors, half reported some prior teaching experience on the tutor survey. In terms of demographics, four out of five non-MNPS tutors were female, and nearly two-thirds were White.

MNPS teachers tutored before school, during their prep periods throughout the school day, and after school. MNPS teachers earned a $1,200 stipend per semester for each small group of students that they tutored in 2021-22, which increased to $1,400 in 2022-23 when tutoring expanded from 10 weeks to 12 weeks. Paraprofessionals earned a stipend of $875 a semester plus applicable overtime. MNPS also employed a "tutor lead" to coordinate the program at each school. Most tutor leads were teachers (42%), instructional coaches (26%), administrators (14%), or had other roles (18%) in their schools. Tutor leads were responsible for recruiting teachers and paraprofessionals at their school to tutor, overseeing the selection, scheduling, and matching of students to tutors, providing ongoing support to tutors, communicating with tutors when students were absent from school, and troubleshooting technological problems for students engaging in virtual tutoring. Tutor leads earned between $1,750 to $2,500 per semester, depending on the number of students tutored at their school.

Over 1,130 of the 5,000 educators employed by MNPS chose to tutor for the Accelerating Scholars program. As shown in Table 1, participating MNPS educators shared similar observable characteristics as those who did not tutor with the exception that Black educators and elementary school teachers were more likely to tutor and elective teachers were less likely to participate.

Legislative Context

Developments in state policy played an important role in shaping the evolution of the Accelerating Scholars program. In January 2021, Governor Lee called a special legislative session on education which resulted in the passage of two interrelated laws: the Tennessee

Learning Loss Remediation and Student Acceleration Act and the Tennessee Literacy Success Act. These laws established new funding streams and expectations for tutoring, as well as a greater focus on early-grades literacy. The first Act established that (beginning in Spring 2023) any third-grade student who does not achieve grade-level proficiency on the English Language Arts (ELA) portion of their Tennessee Comprehensive Assessment Program (TCAP) exam would be retained in 3rd grade, unless the student met specific criteria. Students could be promoted to 4th grade if they attended either a summer learning camp or participated in an approved tutoring program in 4th grade (students with particularly low TCAP scores must do both things to be promoted). The law also established TN ALL Corps tutoring, which provided funding and requirements for tutoring programs that allow students to be promoted. The Literacy Success Act codified the state's reading instruction around foundational literacy skills and required schools to use an approved universal screening test for all K-3 students so instruction can be tailored to identified student needs.

The development of TN ALL Corps is particularly relevant for the context of tutoring in MNPS. ALL Corps provided funding and required districts to offer ALL Corps-aligned tutoring as a condition of promoting 3rd graders who did not meet expectations on TCAP. The law stipulated that ALL Corps tutoring must have a 3:1 student-tutor ratio in grades 1-5 and a 4:1 ratio in grades 6-8. It also required that students receive 90 minutes of tutoring a week, through either two 45-minute sessions or three 30-minute sessions, with a focus on either ELA or math for an entire semester. The state provided a guidebook of literacy instructional routines for early literacy that are aligned with the state's focus on foundational literacy skills and required the use of the online learning platform Zearn for math.

Data & Research Design

Administrative Data

Our analyses draw on data maintained by MNPS in their student information system database which includes student demographic data, course enrollment and grades, student performance on diagnostic and state standardized tests, data on participation and activities students engaged in as part of MTSS Tier II and Tier III services, and teacher employment records maintained by the Office of Human Resources. MNPS also provided detailed data on the implementation of Accelerating Scholars including student tutoring attendance records linked to tutor IDs.

Survey Data

We administered surveys to students, tutors, and tutor leads each semester starting in Fall 2021 to learn more about their experiences with the Accelerating Scholars program. Surveys included Likert-scale items for all surveys as well as open-response items for tutors and tutor leads. Response rates ranged between 48% to 72% for elementary students, between 32% to 62% for secondary school students, between 33% and 83% for tutors, and between 60% and 87% for tutor leads (See Appendix Table B2).

Primary Academic Outcomes

We leverage two measures of academic achievement in reading and math as our primary outcomes. We evaluate the effect of Accelerating Scholars on scores from standardized achievement tests administered in the winter and spring. The tests used by MNPS during our study period changed over time and included the Measures of Academic Progress (MAP), the TNReady state standardized test, and the FAST Early Reading and aRead/aMath diagnostics assessments. We provide a description of each test and detail the outcome test scores and measures of prior achievement for each subject, grade, and semester in Appendix C and Appendix Table C1. We standardize scale scores in the full district sample within subject, grade, semester, and assessment.

We also analyze teacher-assigned course grades in ELA and math classes as a broader measure of student academic effort and achievement. Grades are assigned on a traditional percent scale for each quarter. Any MNPS student who has less than a 50% average receives a course grade of 50%.

Treatment-Control Contrast

The treatment-control contrast for students participating in Accelerating Scholars differed based on when students received tutoring. During PLT periods, tutoring took the place of work on computer-adaptive learning programs such as Lexia CORE 5, SPIRE, and iReady. For students who participated in Accelerating Scholars before or after the school day, tutoring extended the school day by effectively adding more instructional time. For some students, this replaced time that would have been spent at home or in extra-curricular activities. For other students, it was a pull-out approach during afterschool enrichment programs that operated out of MNPS schools such as the YMCA Fun Company and other community organizations. Thus, tutoring took the place of non-instructional time or academic enrichment activities that were likely less focused and intensive during after-school settings.

Student-Level Blocked Randomized Control Trial (Spring 2021)

MNPS launched a pilot tutoring program in the spring of 2021 across three schools with the support of BrightPath Tutors, a tutoring program newly founded by two recent Vanderbilt graduates. This pilot program involved one elementary school, one middle school, and one high school. Tutoring focused on reading in 3rd through 5th grade and math in 6th, 7th and 9th grades.

BrightPath Tutors recruited college students to volunteer to tutor MNPS students online three times per week for 45 minutes over the course of eight weeks. The program served 132 students, roughly half of whom were still attending school remotely.

We randomly assigned 278 students to receive tutoring within grades and subjects, stratifying by students' self-reported intention of finishing the school year either in-person or remotely. The timing of the randomization process created some challenges for obtaining accurate student rosters given the fluid enrollment and high degree of student mobility during hybrid learning and COVID-19 surges. Balance tests in Appendix Table B3 show that students randomized to treatment and control were similar on nine of 11 observable characteristics. We find two significant differences where students randomized to treatment were 3.7 percentage points more likely to be Asian and scored 0.12 SD higher on baseline tests. Our models include controls for all observable baseline characteristics to account for potential chance sampling differences.

We pre-registered our analysis plan with the AEA RCT registry (AEARCTR-0008156) to estimate intent to treat (ITT) effects of a semester of tutoring in this pilot program for each subject separately as follows:

$$Y_{igs} = \alpha + \beta(Treat_i) + \gamma X_i + \delta Z_g + \varepsilon_{igs} \quad (1)$$

where Y_{igs} captures student achievement on standardized tests or grades in the tutored subject for student i in grade g in school s. We include a vector of observable student characteristics, X_i, that includes indicators for gender, race, and economic disadvantage, English learner, special education status, a lagged outcome measure as well as indicators for individual grade randomization blocks, Z_g, and estimate robust standard errors. We complement these reduced form estimates with treatment on the treated (TOT) estimates from two-stage least squares (2SLS) models where we use the random assignment to tutoring as an instrument for participating in at least one tutoring session.

Cluster-Level Blocked Randomized Control Trial with Cross-Over (Fall 2021)

MNPS undertook an ambitious effort to scale up the pilot program by launching Accelerating Scholars in Fall 2021. Tutoring focused on reading in 1st through 3rd grades and math in 8th and 9th grades over the course of 10 weeks during a semester. The district selected a total of 90 traditional schools (13 high schools, 25 middle schools, and 52 elementary schools) to participate in the program by ranking schools on a composite score based on test scores and capacity to offer tutoring, prioritizing those with the lowest scores and highest capacity to implement the program. Limited district capacity and tutor supply constraints created the need for a staggered rollout across the fall and spring semesters, providing an opportunity to randomly assign schools to launch the Accelerating Scholars program in the fall. The district selected 16 of the 90 schools with the highest need to start in Fall 2021, effectively removing them from our randomization sample. We removed six additional schools from our RCT sample because they were operating small-scale independent tutoring programs which they integrated into Accelerating Scholars in the spring. We randomly assigned the remaining 68 schools to either begin participation in the tutoring program in Fall 2021 (n treatment=29) or Spring 2022 (n control=39), blocking on school level, with greater probability of assignment to the control group to ensure an overall balance of participating schools across semesters. Balance tests, shown in Appendix Table B4, reveal no substantive or statistically significant differences across treatment and control groups for the 12 different school-level observable characteristics we measure.

We pre-registered our cluster randomized trial analysis plan with the AEA RCT registry (AEARCTR-0008233). The key source of exogenous variation is randomization at the school-level. Each school then made subjective judgements about which students to tutor, taking into consideration students' needs, the other services students were receiving, and scheduling constraints. Given this, we define our analytic sample to include all students in tutored subjects and grades who were eligible to receive tutoring based on their baseline assessments. Here, we prioritize TOT estimates given that approximately 25% of students eligible for tutoring were actually assigned a tutor. Importantly, this low "take-up" was not a product of non-compliance, but rather the limited ability of MNPS to offer tutoring to all eligible students.

We estimate the effect of a semester of Accelerating Scholars on students' performance on the FAST early reading and MAP math and reading formative assessments administered in January of 2022 as well as 2nd quarter grades. We fit the following 2SLS model:

1st stage: $$D_{is} = \alpha_1 + \beta_1(Treat_s) + \gamma X_i + \delta Z_l + \varepsilon_{is} \quad (2)$$

2nd stage: $$Y_{is} = \alpha_2 + \beta_2 \widehat{D_{is}} + \theta X_i + \varphi Z_l + \omega_{is} \quad (3)$$

where in the first stage we leverage the cluster-random assignment of schools to treatment as an instrument for D_{is}, an indicator for whether individual student *i* in school *s* attended at least one tutoring session in Fall 2021. We include the same vector of covariates X_i described above as well as indicators for schooling level randomization blocks, Z_l, and cluster our standard errors at the school level.

Difference-in-Differences Design (Fall 2021 to Spring 2023)

Accelerating Scholars sought to expand its tutoring efforts in 2022-23 by offering tutoring in additional grade levels. All MNPS schools became eligible to offer tutoring in reading in 1st through 8th grades and math in 4th through 9th grades with 101 out of 128 MNPS traditional public schools participating for at least one semester. We estimate the effects of Accelerating Scholars across the four semesters of large-scale implementation by leveraging a difference-in-differences (DiD) design and event study models. For this set of analyses, we define our analytic sample as all MNPS students who were ever in a tutored grade and had the potential to be treated, regardless of students' prior test scores used to assess eligibility for Accelerating Scholars. This allows us to construct a comparison group with stronger common support given that an increasing number of students who scored below the 15th percentile received tutoring in addition to other specialized supports in 2022-23.

We use a semester-by-student panel that begins in Fall 2020 and ends in Spring 2023 to estimate our DiD/event study models. We begin with the following two-way fixed effects (TWFE) model:

$$Y_{it} = \beta D_{it} + \mu_i + \pi_t + \lambda_{sg} + \varepsilon_{it} \quad (4)$$

where we model academic outcomes for student *i* in semester *t* as a function of student fixed effects, μ_i, semester fixed effects, π_t, and D_{it}, an "ever-treated" treatment indicator that takes on a value of 1 in the first semester in which students first participated in tutoring and every semester thereafter. In addition to the requisite fixed effects for units and time periods, we also include school-by-grade fixed effects, λ_{sg}, to ensure any differential selection patterns into tutoring across schools and grades are held constant. We cluster standard errors at the student level.

This TWFE approach addresses the non-random selection into the program by comparing the changes in individual students' outcomes before and after they participated in tutoring to the changes in outcomes of students in the same grade in the same school who did not participate in tutoring over the same time period. The design assumes that in the absence of the tutoring program, participants would have experienced a similar change in their outcomes as students in the comparison groups.

We then fit event study models to explore the dynamic treatment effects of tutoring over time and assess the plausibility of the parallel trends assumption. We modify equation (4) as follows:

$$Y_{it} = \sum_{r=-5}^{3} \beta_r 1(t = t_D^* + r) + \mu_i + \theta_t + \lambda_{sg} + \varepsilon_{it} \qquad (5)$$

where we replace our binary indicator for treatment, D_{it}, with a set of indicators for time relative to treatment where t_D^* captures the semester of first treatment and *r* is the number of years relative (prior to or post) to treatment. Here, t_0 captures the immediate effect of a semester of Accelerating Scholars tutoring, while t_1 reflects a weighted combination of 1) dosage effects for the 42% of students tutored in reading and 33% in math who received a second consecutive semester of tutoring (either fall to spring or spring to fall), and 2) any persistent effects of tutoring from t_0 for those students who did not receive a second consecutive semester of tutoring. For grade outcomes, we replace school-by-grade fixed effects with teacher-by-school-by-course fixed effects to better account for differential grading standards across schools, teachers, and classes. We display results in our event study figures in a window between *r*=[-3,1] given the increasing sample imbalance for estimates in the tails of our event study specification due to the short six-semester panel and the heterogenous timing of treatment.

We also assess the robustness of our event-study estimates to recent concerns raised about bias in TWFE models due to heterogeneous treatment effects (Callaway & Sant'Anna, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). Following Cengiz et al. (2019), we create distinct estimation samples for each semester-specific treated cohort of students and include only never-treated students as the comparison group, stacking these samples and estimating their joint effect simultaneously while allowing all model terms to differ across cohorts.

Interviews & Focus Groups

We collected a range of qualitative data to better understand the nature of program implementation in 2021-22 and 2022-23. This included eight interviews with organizational leaders to provide broad perspectives as well as interviews or focus groups with principals, tutor leads, and external tutors. The interviews were semi-structured to include common questions for all participants which we adapted based on the specific portfolio of tutoring at that school (e.g., all in-person, all virtual, or both). We sought interviews with internal tutors (i.e., teachers and paraprofessionals who were also tutors), but the burden teachers faced in the aftermath of the pandemic meant this was not feasible. In 2021-22, we interviewed 10 principals and 25 tutor leads. We interviewed or included in a focus group 18 external tutors from a total of eight schools. In 2022-23, we interviewed an additional 34 tutor leads. In both years, the distribution of schools selected for principal and tutor lead interviews reflected the grade levels of schools participating in tutoring.

Findings

The Dynamics of Scaling Accelerating Scholars

We first describe how Accelerating Scholars was developed and evolved over time. The interview, survey, and administrative data we collected provide evidence of three major shifts as MNPS took Accelerating Scholars to scale. First, MNPS increasingly integrated the Accelerating Scholars program into its core operations run by district staff. Second, the composition of tutors changed substantially as the program scaled. Third, the district's approach to program integration and alignment with core instruction evolved over time.

Centering Accelerating Scholars in MNPS

The initial development of Accelerating Scholars involved multiple external partners. In Spring 2021, the district appointed a high-level administrator, their Chief Strategy Officer, to oversee the launch of the program. The CSO partnered with three non-profit organizations to design and deliver tutoring to students. These organizations included: PENCIL, a local nonprofit organization with a longstanding relationship with MNPS that builds community and business partnerships with schools to provide resources including supplies, tutoring, and career education opportunities; BrightPath, a new tutoring organization launched by two Vanderbilt graduates; and Tennessee SCORE, an education advocacy organization. The district also began working with TNTP, which provided support for tutor training and curricular materials, and Vanderbilt University, which had developed Tutor Nashville to provide opportunities for Vanderbilt students and faculty to support MNPS during virtual instruction, in Fall 2021.

While MNPS was the final decision-maker, multiple organizations were overseeing key aspects of the work and providing recommendations for what should happen. One organizational leader described a consultant hired by SCORE as the "quarterback" that facilitated the overall work in Fall 2021. With so many partners, there was some confusion about "who does what when and how." One factor that initially limited more robust MNPS oversight was the limited number of MNPS personnel focused on Accelerating Scholars. Organizational leaders noted in Fall 2021 that the bulk of the work was falling on the CSO, who oversaw multiple large-scale initiatives, while other districts engaged in similar work had a full team.

The district began to expand its internal staffing to support Accelerating Scholars in Spring 2022. By the 2022-23 academic year, the day-to-day work of operating Accelerating Scholars became largely centered within a team of MNPS staff. Compared to interviews from the prior year, tutor leads were much more positive about the organizational support they had for their work. Further, some of the partners from 2021-22, such as BrightPath, were phased out and other partners, such as SCORE, shifted to more limited supporting roles.

Administrative data also provide evidence of Accelerating Scholars becoming more centered in MNPS as it grew from a boutique program to a district-wide initiative that shaped core operations. We show the number of students tutored in each grade and subject over time in Figure 2. In the first three semesters of the program, tutoring was limited to early literacy and secondary math with the number of schools and students participating steadily increasing over time. By Spring 2023, 4,000 students in grades 1-9 participated in tutoring across 101 schools.

The increasing centralization and integration of Accelerating Scholars into core district operations came with several advantages. First, the district was able to steadily increase the number of weeks of tutoring offered per semester from 8 to 12 and bolster attendance. This resulted in an increase in the number of average tutoring sessions students attended over time from 16 to 24 as shown in Table 2. The greater centralization within MNPS also came with improved communication and aligned goals. Tutor leads were more satisfied with communication and support in 2022-23. For example, in Spring 2022, a tutor lead asked for "better communication … more reciprocal communication" from the district coordinator that allowed them to ask questions and not just receive information. Another tutor lead who served in this role in both years noted the improvement in district communication. They said, "And as the years have progressed, they've added more people at the district office to meet the needs … Because when all this first started, [coordinator] was handling all 14 bazillion emails. …It was crazy. So now there are people that support us." Overall, tutor leads expressed more positive experiences and the program in 2022-23 and wanted to see it continue to grow.

Tutor satisfaction also increased over time as the program became more centralized. Figure 3 Panel A depicts results from tutor surveys on the likelihood of recommending the program to a friend and satisfaction with support. Both measures increase every semester. Student satisfaction with tutoring remained high over time with approximately 80% of elementary school and 60% of middle school students reporting high levels of enjoyment of their tutoring sessions, as seen in Figure 3 Panel B.

Change in Tutor Composition

The rapid growth of the Accelerating Scholars program required the district to adopt a flexible approach to staffing. In Figure 4, we depict this evolution by displaying the total number of tutored students, by tutor type, over time. The program began working entirely with volunteer tutors, but only 5% of tutors were volunteers by Spring 2023. BrightPath, one of the early key partners, was founded as a way to get volunteers involved in schools, focusing on college students and segments of the population that were experiencing underemployment as a result of COVID-related economic impacts. The involvement with PENCIL, which has a long history of offering ways to volunteer in MNPS schools, also demonstrates the volunteer-focused approach of the early design. However, despite considerable efforts to expand their pool of volunteer tutors including contracting a public relations firm to conduct an intensive local recruitment campaign, the supply of volunteers had hit a plateau as the economy recovered and college students returned to their campuses in person.

In Fall 2021, MNPS began looking to its own staff - teachers, paraprofessionals, and other school-based staff - as a way to meet the ambitious goal of tutoring 7,000 students set by the superintendent. Over time, Accelerating Scholars came to rely almost entirely on MNPS teachers. By Spring 2023, 85% of students in Accelerating Scholars were being tutored by an MNPS teacher.

This shift in tutor composition had several advantages and drawbacks. While tutor leads valued volunteer tutors and said students often appreciated knowing that someone cared about them enough to volunteer their time, they also expressed concerns about the effectiveness of what they could do for students. One tutor lead described a view echoed by others, "Teachers and paraprofessionals…those are the tutors that are more higher impact. They're having better relationships with students. They have better structure. They already know the content better. The paid staff is more efficient and effective than the volunteers." Beyond the instructional expertise of MNPS staff, tutor leads described more interactions with in-person tutors – such as providing new curriculum materials, coordinating around absences, and connecting with classroom teachers – than with virtual tutors. This interaction was facilitated by their co-location in the school as well as existing professional relationships.

MNPS also saw their teachers as an attractive alternative to volunteers because their instructional experience made it more feasible for them to work with small groups of students rather than just 1:1 as volunteers were asked to do, increasing capacity. Thus, as the composition of tutors changed over time, so did the student tutor ratios. Tutoring shifted from 100% 1:1 to 11%, with 3:1 becoming the modal ratio by Spring 2023. This also reflects the requirements of ALLCorps tutoring at a ratio of 3:1 up to 5th grade.

While the instructional expertise of teachers made them good tutors, there were challenges in recruiting enough teachers to meet all students' needs. Principals and tutor leads reported that teachers and paraprofessionals were primarily motivated by the additional compensation they could receive for tutoring. The challenge of recruiting teachers to tutor was partly due to scheduling constraints, given that teachers could only tutor while school was in session during their planning time. Teacher planning time did not always align with when students had a scheduled PLT, especially in middle and high schools. Tutor leads and principals also reported concerns about teacher burnout as teachers had no unstructured time during the day. Especially in schools with higher teacher turnover, participants worried that school capacity to tutor more students was limited.

A second key consequence of this shift in the composition of tutoring was the change from virtual to in-person tutoring, as seen in Table 2. In Spring 2021, the pilot tutoring program was staffed entirely by volunteers who all tutored virtually. In Fall 2021, 43% of tutoring was done in person. By Spring 2023, 93% of tutoring was delivered in person. The initial use of virtual tutoring allowed for greater access to tutors, particularly when schools were closed to visitors to minimize the spread of COVID. Volunteer tutors also appreciated being able to join remotely as they needed less time off work to meet with their students.

Overall, though, participants expressed concerns over virtual tutoring. Tutor leads and tutors suggested that technology problems affected about half of their scheduled sessions, with students often starting a 30-minute session 10 minutes late because they needed to find working headphones or a laptop charger, or because the laptop would automatically reboot to load new software. There were also reports that virtual tutoring was slower because asking students to write something with a virtual drawing tool took longer than a pencil. Finally, virtual tutoring could create loud and distracting environments as students typically joined the tutoring session using a laptop and headphones from their regular classroom or the hallway outside their classrooms. Some schools created designed tutoring spaces in the library or an empty classroom to minimize these distractions.

Despite these challenges with virtual tutoring, MNPS contracted with an external tutoring organization to supplement tutor capacity in secondary math. Far fewer volunteers and MNPS staff members felt comfortable tutoring students in upper-level math compared to early literacy. Tutored by Teachers, which pays credentialed teachers to tutor virtually, worked with 127 middle school and 9th grade students in math in Fall 2022. The contracted tutors offered greater capacity of tutors with more training and support, as well as greater flexibility in scheduling tutors during the school day for students. In 2022-23, tutor leads who worked with external tutors indicated that the contracted organization provided oversight and support for tutors, which meant that tutor leads had to provide less direct support.

Evolving Integration and Alignment of Tutoring into the School's Instruction

From the very start, MNPS was committed to developing a tutoring program that was integrated into their core programming and directly aligned with standards-based instruction happening in students' core classes. The district remained committed to these principles but also had to iteratively redesign and adapt the program to achieve these goals as the program expanded. One core change in their approach to tutoring over time was the gradual expansion of tutoring immediately before and after school hours. Tutoring during the Spring 2021 pilot and initial Fall 2021 scale-up occurred almost entirely during school hours. By Spring of 2023, 53% of all tutoring sessions were before or after school (13% before and 40% after; Table 2).

Multiple factors contributed to this shift towards tutoring outside of regular school hours but still at students' school buildings. One was the availability of tutors. Recruiting enough volunteers who could come during the day was a challenge. Logistically, more teachers who were willing to serve as tutors could be scheduled to tutor before or after school than during their planning periods. Physical space was another scheduling constraint during the school day. For in-person tutoring, the tutor and students typically met in the teacher's classroom, which was usually empty since it was the teacher's planning time. Paraprofessionals had to find an unused space in the building for the small group to meet, although there are few such spaces in most schools. Another benefit of tutoring outside of school hours was that tutor leads were able to provide oversight and support. Putting students in a dedicated tutoring space for virtual tutoring required the tutor lead to be available to monitor that space. Over a third of tutor leads were full-time classroom teachers, which meant they were not available when tutoring was happening during the day. Finally, the option of tutoring outside of the school day made it possible for students receiving specialized support during their PLT period to also participate in tutoring.

Curriculum was another core aspect of tutoring implementation that continued to evolve over time as the district sought to strengthen the program and respond to tutor feedback. MNPS provided tutors with grade-level curriculum materials and workbooks with activities. These materials were developed by MNPS and its partners, with a priority to align them to core instruction. Some tutors reported that these standards-based curriculum materials were too repetitive with what the student did in class (e.g., students reported they just did the same activity that day) and were not always at the level appropriate for the student. One elementary tutor said, “The tutoring the curriculum provided for the tutoring was the classroom curriculum as well. Which is where we found the redundancy happening … So when we shifted and started pulling more supplemental work, supplemental articles. That was like a little bit better.” This type of supplementing the curriculum, however, seemed to work better for teachers as they were already familiar with the materials because they used them for regular instruction with students. External tutors, however, struggled more with the materials, especially in secondary math where tutors may not have had strong content knowledge.

MNPS responded to curricular concerns by revising the ELA curriculum to minimize redundancy and working with the district’s curriculum and instruction office to do a major overhaul on the math curriculum in Spring 2021. The math curriculum was revised again in Fall 2022 to introduce Zearn, as required by ALLCorps. The ELA curriculum also went through further revisions in Fall 2022 to align with the ALLCorps instructional guide that focuses on foundational literacy skills. Despite the revisions, adapting the provided curriculum to students’ specific needs was a continual challenge because tutors did not have access to diagnostic assessments of their students’ knowledge and skills. External tutors expressed a desire to know more about the individual needs of their students. For example, one tutor said, “I'd want to have some regular touchpoint with the teacher, or I'd want to know that the tutor lead really knew what was up with this kid.” The initial tutoring design expected teachers of tutored students to provide reports to tutors, but this ended up not being feasible. Another tutor explained the difficulty of connecting with their student’s teacher by saying, “I get the feeling from the feedback I've gotten from everybody that this is just one more thing on their plate that they have." Communication between tutors and teachers of tutored students was easier with in-person tutoring due to existing professional relationships and the proximity of tutors and teachers.

The Effects of Accelerating Scholars on Student Achievement

Experimental estimates of the effect of a semester of Accelerating Scholars tutoring from our student-level RCT of the initial pilot program and cluster-level RCT of the first semester of scale-up find little evidence that the program achieved the large effects found in many prior RCTs of tutoring programs. ITT estimates from the pilot program shown in Table 3 are small in magnitude, insignificant, and of inconsistent sign for both standardized tests and course grades. However, the small size of the pilot program limits the statistical power we have to rule out small to moderate effects.

ITT estimates from the first semester of scale-up shown in Table 4 are uniformly positive, but again small in magnitude and imprecisely estimated. When we scale estimates for test scores to be TOT estimates we arrive at point estimates of 0.04 SD in reading and 0.03 SD in math with very large confidence intervals. We do find one marginally significant effect of an increase of 1.3 percentage points in students' grades in ELA classes for ITT and 4.2 percentage points for TOT.

Analyses from our DiD models that identify a cumulative average effect of Accelerating Scholars across four semesters of the scale-up suggest the program was successful at improving student test scores in reading but had no effect on students' performance on math tests or grades in either subject. As shown in Table 5, effects on standardized achievement in reading are 0.09 SD in our preferred model with school-by-grade fixed effects. We find precisely estimated null effects on students' test scores in math as well as grades across our preferred models. Test score estimates are quite stable across models that exclude and include school-by-grade fixed effects or teacher-by-school-by-course fixed effects adding further confidence in our estimates. The slightly larger effects on reading test scores from our DiD model compared with our RCT estimates is consistent with evidence of a dosage effect. Treated students in our DiD analytic sample for reading received an average of 1.49 semesters of tutoring, and those in our math sample received a slightly lower average dosage of 1.35 semesters.

Event study estimates shown in Figure 5 are consistent with findings from both our cluster-RCT and DiD models. We estimate that Accelerating Scholars increased student performance on standardized tests in reading by 0.05 SD in the first semester of tutoring and that this effect doubled to 0.10 SD in the semester after (t_1). This semester-specific effect of 0.05 SD in reading is nearly identical to the 0.04 SD effect found in our cluster-RCT. These results are also suggestive of a dosage effect given that 42% of treated students received a second consecutive semester of tutoring in reading. However, we cannot separately identify dosage effects from the temporal dynamics of treatment effects in our event study framework. Event study estimates for test scores in math suggest a potential Ashenfelter dip where students selected for treatment experienced a negative shock to achievement in the semester prior to treatment and then rebounded to their pre-treatment levels after being treated. We find no compelling evidence of dynamic treatment effects on students' grades in either subject.

Robustness

We test the robustness of our DiD and event-study estimates in several ways and find our results are consistent across modeling approaches. One concern is that students selected for tutoring experienced a negative random shock to achievement and were then selected on this measure to be eligible for tutoring. These students might experience a rebound or reversion to the mean in their academic performance on standardized tests in the following year, creating the perception of a positive treatment effect. We test the sensitivity of our DiD estimates to a potential Ashenfelter dip by refitting our models after dropping test scores in the year just prior to the first treated year. As shown in Table 6, our estimates for student achievement are unchanged with an estimated effect in reading of 0.09 SD and no effect in math. Point estimates for course grades in both subjects remain trivial in magnitude, with point estimates that are less than a third of a percentage point.

We also explore an alternative modeling approach for our DiD estimates that reflects the time-varying nature of the treatment. Our preferred models estimate treatment as a persistent effect across all periods after a student first participated in tutoring. Here, we define treatment as a semester-specific measure that is coded as zero in the semesters which a student did not receive tutoring even if they were tutored in a prior semester. Any positive effects of tutoring that persist beyond the semester in which students participate would cause this approach to understate the full effects of tutoring. Using this approach, we find that a semester of Accelerating Scholars tutoring increases test scores in reading by 0.05 SD, but with no effect in math. These estimates shown in Table 7 are consistent with our findings from event study models which find an effect of 0.04 SD on student performance on standardized tests in the first semester a student was tutored.

Finally, we test the robustness of our TWFE event study models to the threat posed by heterogeneous treatment effects using a stacked modeling approach. As shown in Appendix Figure 1 and Appendix Table B5, estimates from this alternative modeling approach are nearly identical to those from our primary model given our short panel and limited number of cohorts.

Heterogeneous Effects Across Students

The large-scale nature of Accelerating Scholars and its somewhat decentralized student selection process resulted in a wide variety of students participating in the program. Here we explore whether tutoring was more effective for supporting students in different parts of the achievement distribution as well as with different student characteristics. We test for heterogeneous effects on achievement by estimating quantile treatment effects from our DiD model using unconditional quantile regression (Rios-Avila, 2020). This approach is advantageous because it avoids any potential confounding due to a mechanical negative relationship between prior achievement and tutoring program effects where students with negative shocks experience the largest regressions to the mean.

We find clear and consistent evidence that tutoring had distinctly heterogeneous effects on students' test scores. As shown in Figure 6, treatment effects in both reading and math are concentrated in the middle to the upper-middle ranges of the relative test score distribution among MNPS students (30th to the 70th percentiles in reading and 50th to 70th percentiles in math). Effects became as large as 0.22 SD in reading for the 50th percentile and 0.11 SD in math for the 60th percentile. Point estimates turn negative in the tails of the test score distribution among the lowest and highest performing students.

Differential effects for grades are also evident but follow a very different pattern. We find positive effects of tutoring on students' grades in both ELA and math in the very lowest range of the grade distribution but no effects, or even negative effects, in the middle and upper ranges of the grade distribution. We estimate positive effects on grades of 0.97 percentage points in ELA and 2.56 percentage points in math at the 10th percentile of the distribution among MNPS students. These findings illustrate that the Accelerating Scholars tutoring program was substantially more effective at improving the skills captured on standardized diagnostic and summative assessments among students in the middle of the district-specific performance distribution but lead to improvements in class performance as judged by teachers among students with the lowest grades.

We test for differential effects based on students' background characteristics by interacting the treatment indicator from our ever-treated DiD model with indicators for special education status, economic disadvantage, English learner status, gender, and race. Results of these analyses in Table 8 suggest that Accelerating Scholars was more effective for English learners. Estimates for English learners are 0.04 SD larger on reading tests and 0.08 SD larger on math tests. We find smaller effects on test scores for female students in both reading (-0.04 SD) and math (-0.07 SD) and conversely, larger effects on course grades for female students in ELA (0.64 percentile points) and math (1.25 percentile points). Few other consistent differences in the effects of tutoring emerge.

Flexible Program Characteristics

A key approach for taking tutoring to scale in MNPS was the decision to build Accelerating Scholars around a core set of design principles and to allow other key features to evolve over time. This within-program variability creates the opportunity to examine the effects of different flexible design features. We view these analyses as exploratory given the non-random nature of student assignment to different tutoring modalities and the bundled nature of these flexible design characteristics. For example, volunteer tutors almost always tutored online with an individual student. MNPS teachers largely tutored in person with small groups of students.

We test for heterogeneous effects across the four flexible program characteristics by interacting our time-varying treatment term DiD model described in the robustness section with indicators for semester-specific tutoring characteristics such as having an MNPS teacher as a tutor, tutoring that occurred during school hours, tutoring that was in-person, and tutoring that used a 1:1 student-tutor ratio. Prior research has identified all of these features as hypothesized characteristics of effective tutoring programs. As shown in Table 9, we find no compelling evidence that tutoring effects varied systematically across these flexible design features. Estimates are generally small and insignificant across outcomes. The few significant differences we detect are one-off findings that are not consistent in sign or magnitude across our two test scores and two grade outcomes. We view the absence of any clear differences across these flexible design features as suggestive evidence that MNPS was able to build a tutoring program around its core design principles that was robust to a range of flexible design features.

Discussion

Contextualizing the magnitude of tutor impacts

MNPS successfully designed, built, and scaled a district-operated tutoring program that delivered over 125,000 hours of tutoring across five semesters to 6,848 unique students – a major accomplishment. We find the program improved student achievement in reading meaningfully, with small to medium effects on average, but had no effects on math achievement or student grades. These results contrast with the large average effects documented in meta-analytic reviews of RCTs evaluating tutor programs based on small- to medium-scale in-person tutoring programs implemented voluntarily in the pre-COVID-19 era (Dietrichson et al., 2017; Fryer, 2017; Inns et al., 2019; Nickow et al., 2024; Pellegrini et al., 2021). Here we discuss four possible explanations for these differences including: 1) a limited treatment-control contrast, 2) modest program duration, 3) heterogeneous effects, and 4) miscalibrated expectations.

Most studies of tutoring programs compare the effect of in-school tutoring to traditional whole-class instruction, and out-of-school tutoring to non-academic afterschool activities. The treatment-control contrast in this setting was markedly different. Roughly 55% of tutored students received tutoring in their PLT classes during the school day. Students who did not participate in tutoring used this time to work with computer-adaptive learning programs or received Tier II or III services such as small-group instruction from a teacher. This counterfactual condition approximates individualized instruction in tutoring to a large degree. The roughly 45% of students who received tutoring before or after the school day did so in large part during afterschool enrichment programs. Although these programs were not explicitly focused on academic acceleration, they did provide time for students to work on their homework and offered academic support. Thus, even these students may have forgone time dedicated to school-related activities when participating in afterschool tutoring.

A second possible explanation is the somewhat limited duration of the program. The majority of students participated in one semester of tutoring with a targeted dosage of 15 total hours in 2021-22 (1.5 hours * 10 weeks) and 18 hours in 2022-23 (1.5 * 12 weeks). While Accelerating Scholars was successful at delivering much of this intended dosage, the focus on one-semester of tutoring during the scale-up period caused the total annual dosage to be notably less than other tutoring programs, on average. For example, Kraft et al. (2024) find an average total intended dosage of 32 hours of tutoring across an average of 16.5 weeks of tutoring in their review of tutoring programs.

Another explanation that is consistent with our findings is that the Accelerating Scholars model of standards-based tutoring was only effective at raising test scores among students in the middle of performance distribution, but it served students across the full performance range. For example, the most sizable effects in reading are concentrated in the 40th to the 60th percentiles of the performance distribution, and all positive effects are in the 50th to the 70th percentile range for math. This concentration of effects on test scores among the middle of the test score distribution may be due to the nature of the Accelerating Scholars tutoring model. MNPS designed the program as a supplemental source of personalized instruction in core grade-level content. This universal approach to tutoring followed a set curricula likely designed for the average student, in contrast to tutoring models that personalize the content of instruction based on diagnostic assessments of students' individual learning needs.

It is possible that Accelerating Scholars might have had a larger impact if the program was able to better target students who most benefit from the standards-based model of tutoring. As shown in Figure 7, only 30% of students tutored in reading and 33% of students tutored in math were in the percentile ranges where effects were most concentrated. It is also possible that these heterogeneous effects reflect differences in the counterfactual condition that varied systematically across the achievement distribution. Many lower performing students were already receiving some type of individualized support which may have attenuated the treatment-control contrast to a greater degree in this range of the achievement distribution.

A final explanation worth noting is the possible need to recalibrate expectations for the effects of tutoring programs at scale. There exists limited evidence on the expected effects of tutoring programs delivered at the scale of Accelerating Scholars. Kraft et al. (2024) find a clear pattern of declining effect sizes when comparing the pooled effects of smaller versus larger tutoring programs. Two recent studies evaluate efforts to implement and scale a range of tutoring models across districts in the post-COVID era and find small and insignificant effects (Carbonari, DeArmond, et al., 2024; Carbonari, Dewey, et al., 2024). Estimates of large-scale programs may also have somewhat smaller effect sizes, mechanically, because of the more heterogeneous population these studies use to estimate the standard deviation for calculating an effect size statistic (Fitzgerald & Tipton, 2024). For example, the standard deviation of baseline tests scores for students randomized to the control group in our pilot RCT is 0.46 SD. Using this sample to calculate our estimated effects would more than double the magnitude of our point estimates relative to using the full sample which has a standard deviation of 1 by construction. We standardized by the full district-wide population of MNPS test takers across all our estimates.

Sustaining tutoring at scale

Two central features serve to limit the growth of tutoring programs: high-costs and limited tutor supply. The MNPS Accelerating Scholars program approached these challenges by designing a district-run and operated tutoring program that evolved over time. Funding for Accelerating Scholars came almost entirely from sources outside the district including investments by the Bloomberg Foundation, federal funds from the Elementary and Secondary School Emergency Relief Fund (ESSER), a Connected Literacy Grant to the non-profit organization PENCIL that partnered with MNPS, and state funding as part of the TN ALL Corps program. All these funding sources are time-limited, which raises important questions about the sustainability of the program without these external funds. We estimate that the program costs about $750 per student per semester or $1,500 per academic year of tutoring. Stipends for MNPS staff who served as tutors were the primary cost driver, accounting for 80% of total costs.

MNPS's approach to primarily staffing its tutoring program with current employees may prove to be a sustainable design given that it draws from a stable pool of labor supply. While teachers earn an hourly wage of about $40 for tutoring, which is higher than market wages paid by most tutoring programs, the district's choice to have teachers work with small groups of students serves to reduce the per-student cost of this staffing decision. An open question is if MNPS staff will choose to tutor on a consistent basis or if they will find the work too demanding on top of their regular teaching responsibilities.

Conclusion

The long-term future of individualized instruction in public education depends critically on the ability of districts to successfully deliver tutoring as part of their regular programming and to sustain the funding and staff necessary to do so. The development and implementation of MNPS's Accelerating Scholars program provides a valuable example of one district's ambitious effort to develop a district-operated tutoring program at scale. Our research partnership that spanned multiple years serves to reveal how the program evolved in important ways to meet the challenges of scaling. This suggests that ongoing efforts to evaluate tutoring programs in the post-COVID-19 era would benefit from sustained partnerships given that point-in-time evaluations likely do not reflect the dynamic evolution of programs in this early phase of innovation and implementation. These findings also illustrate the importance of moving beyond estimates of the average effects of tutoring to exploring heterogeneous effects among participating students. Our results suggest that effects at scale could be increased with an effort to better target programs to those students who benefit the most, and to adapt programs to better meet the needs of those students who benefit the least. Important challenges remain for delivering tutoring at scale that is both impactful and cost-effective. Insights from MNPS's Accelerating Scholars program can contribute to these ongoing efforts as districts work to both share best practices and adapt approaches to fit their local contexts.

References

Al-Ubaydli, O., List, J. A., & Suskind, D. (2020). 2017 Klein Lecture: The Science Of Using Science: Toward An Understanding Of The Threats To Scalability. *International Economic Review*, *61*(4), 1387–1409.

Angrist, N., Ainomugisha, M., Bathena, S. P., Bergman, P., Crossley, C., Cullen, C., Letsomo, T., Matsheng, M., Marlon Panti, R., Sabarwal, S., & Sullivan, T. (2023). *Building resilient education systems: Evidence from large-scale randomized trials in five countries* (Working Paper 31208). National Bureau of Economic Research. https://doi.org/10.3386/w31208

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73–102.

Barshay, J. (2023, October 16). PROOF POINTS: Schools keep buying online drop-in tutoring. The research doesn't support it. *The Hechinger Report*. https://hechingerreport.org/proof-points-schools-keep-buying-online-drop-in-tutoring-the-research-doesnt-support-it/

Bhatt, M., Guryan, J., Khan, S., LaForest, M., & Bhavya Mishra. (2024). Can technology facilitate scale? Evidence from a randomized evaluation of high dosage tutoring. *NBER Working Paper*, *32510*. https://doi.org/10.17605/OSF.IO/UW8EH

Bonesrønning, H., Finseraas, H., Hardoy, I., Iversen, J. M. V., Nyhus, O. H., Opheim, V., Salvanes, K. V., Sandsør, A. M. J., & Schøne, P. (2022). Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment. *Journal of Public Economics*, *216*, 104765. https://doi.org/10.1016/j.jpubeco.2022.104765

Cabezas, V., Cuesta, J. I., & Gallego, F. A. (2011). *Effects of short-term tutoring on cognitive and non-cognitive skills: Evidence from a randomized evaluation in chile* (J-PAL Working Paper). Abdul Latif Jameel Poverty Action Lab. https://www.povertyactionlab.org/sites/default/files/research-paper/493%20-%20short-term%20tutoring%20May2011.pdf

Callaway, B., & Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., & Goldhaber, D. (2024). *Impacts of academic recovery interventions on student achievement in 2022-23* (303–0724; CALDER Working Paper). Center for Analysis of Longitudinal Data in Education Research at the American Institutes for Research. https://caldercenter.org/publications/impacts-academic-recovery-interventions-student-achievement-2022-23

Carbonari, M. V., Dewey, D., Kane, T. J., Muroga, A., DeArmond, M., Dizon-Ross, E., Goldhaber, D., Morton, E., Davison, M., & Hashim, A. K. (2024). *The impact and implementation of academic interventions during Covid: Evidence from the road to recovery project* (275-0624–2; CALDER Working Paper). Center for Analysis of Longitudinal Data in Education Research at the American Institutes for Research. https://caldercenter.org/sites/default/files/CALDER%20WP%20275-0624-2.pdf

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, *134*(3), 1405–1454.

Chambers, J., Brodziak de los Reyes, I., & O'Neil, C. (2013). *How much are districts spending to implement teacher evaluation systems*. RAND.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.

Cohen-Vogel, L., Cannata, M., Rutledge, S. A., & Socol, A. R. (2016). A model of continuous improvement in high schools: A process for research, innovation design, implementation, and scale. *Teachers College Record: The Voice of Scholarship in Education*, *118*(13), 1–26. https://doi.org/10.1177/016146811611801301

Cortes, K., Kortecamp, K., Loeb, S., & Robinson, C. (2024). *A scalable approach to high-impact tutoring for young readers: Results of a randomized controlled trial* (Working Paper 32039). National Bureau of Economic Research. https://doi.org/10.3386/w32039

Davis, J. M. V., Guryan, J., Hallberg, K., & Ludwig, J. (2017). *The economics of scale-up* (w23925). National Bureau of Economic Research. https://doi.org/10.3386/w23925

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, *87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Fesler, L., Gu, A., & Chojnacki, G. (2023). Air tutors' online tutoring: Math knowledge impacts and participant math perceptions. middle years math grantee report series. In *Mathematica*. Mathematica. https://eric.ed.gov/?id=ED628638

Fitzgerald, K. G., & Tipton, E. (2024). Using extant data to improve estimation of the standardized mean difference. *Journal of Educational and Behavioral Statistics*, 10769986241238478. https://doi.org/10.3102/10769986241238478

Fryer, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experimentsa. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of*

Economic Field Experiments (Vol. 2, pp. 95–322). North-Holland. https://doi.org/10.1016/bs.hefe.2016.08.006

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Gupta, S., Supplee, L. H., Suskind, D., & List, J. A. (2021). Failed to Scale: Embracing the Challenge of Scaling in Early Childhood. In *The Scale-Up Effect in Early Childhood and Public Policy* (pp. 1–21). Routledge.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review*, *113*(3), 738–765. https://doi.org/10.1257/aer.20210434

Honig, M. I. (2006). *New directions in education policy implementation: Confronting complexity*. Suny Press.

Inns, A. J., Lake, C., Pellegrini, M., & Slavin, R. (2019). *A quantitative synthesis of research on programs for struggling readers in elementary schools* (Best Evidence Encyclopedia). Center for Research and Reform in Education.

Kirkebøen, L. J., Gunnes, T., Lindenskov, L., & Rønning, M. (2021). *Didactic methods and small-group instruction for low-performing adolescents in mathematics: Results from a randomized controlled trial* (Working Paper 957). Discussion Papers. https://www.econstor.eu/handle/10419/249147

Kraft, M. A., & Falken, G. T. (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open*, *7*, 233285842110428. https://doi.org/10.1177/23328584211042858

Kraft, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. *AEA Papers and Proceedings*, *112*, 614–618. https://doi.org/10.1257/pandp.20221038

Kraft, M. A., & Lovison, V. S. (2024). *The effect of student-tutor ratios: Experimental evidence from a pilot online math tutoring program*. https://doi.org/10.26300/87MA-E949

Kraft, M. A., Schueler, B. E., & Falken, G. T. (2024). *What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability* (EdWorkingPaper 24–1031). Annenberg Institute at Brown University. https://edworkingpapers.com/ai24-1031

LeMahieu, P. (2011, October 11). What we need in education is more integrity (and less fidelity) of implementation. *Carnegie Foundation for the Advancement of Teaching*. https://www.carnegiefoundation.org/blog/what-we-need-in-education-is-more-integrity-and-less-fidelity-of-implementation/

Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., & Ahmed, H. (2015). Paired reading: Evaluation report and executive summary. In *Education Endowment Foundation*. Education Endowment Foundation. https://eric.ed.gov/?id=ED581127

Manna, P. (2010). *Collision course: Federal education policy meets state and local realities*. CQ Press.

May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., Gillespie, J. N., & Tognatta, N. (2015). Year one results from the multisite randomized evaluation of the i3 scale-up of reading recovery. *American Educational Research Journal*, *52*(3), 547–581. https://doi.org/10.3102/0002831214565788

Muralidharan, K., & Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, *31*(4), 103–124.

National Student Support Accelerator. (2023). *A snapshot of state tutoring policies*. https://studentsupportaccelerator.org/briefs/snapshot-state-tutoring-policies

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

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective Programs in Elementary Mathematics: A Meta-Analysis. *AERA Open*, *7*, 2332858420986211. https://doi.org/10.1177/2332858420986211

Personalized Learning Initiative Research Team. (2024). *Realizing the promise of high dosage tutoring at scale*. The University of Chicago Education Lab. https://educationlab.uchicago.edu/resources/realizing-the-promise-of-high-dosage-tutoring-at-scale-preliminary-evidence-for-the-field/

Ready, D. D., McCormick, S. G., & Shmoys, R. J. (2024). *The effects of in-school virtual tutoring on student reading development: Evidence from a short-cycle randomized controlled trial* (24–942; EdWorkingPaper). Annenberg Institute at Brown University. https://edworkingpapers.com/ai24-942

Rios-Avila, F. (2020). *Recentered influence functions (rifs) in stata: Rif regression and rif decomposition*. *20*(1). https://journals.sagepub.com/doi/full/10.1177/1536867X20909690

Robinson, C. D., Bisht, B., & Loeb, S. (2022). The inequity of opt-in educational resources and an intervention to increase equitable access. *Annenberg Institute at Brown University EdWorkingPaper*, *22–654*. https://www.edworkingpapers.org/sites/default/files/ai22-654.pdf

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). *Accelerating student learning with high-dosage tutoring.* (EdResearch for Recovery Design Principles). EdResearch for Recovery Project. https://eric.ed.gov/?id=ED613847

Robinson, C. D., Pollard, C., Novicoff, S., White, S., & Loeb, S. (2024). The effects of virtual tutoring on young readers: Results from a randomized controlled trial. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://edworkingpapers.com/ai24-955

Sabelli, N. H., & Harris, C. J. (2015). The role of innovation in scaling up educational innovations. In C.-K. Looi & L. W. Teh (Eds.), *Scaling Educational Innovations* (pp. 13–30). Springer Singapore. https://doi.org/10.1007/978-981-287-537-2_2

Sirinides, P., Gray, A., & May, H. (2018). The impacts of reading recovery at scale: Results from the 4-year i3 external evaluation. *Educational Evaluation and Policy Analysis*, *40*(3), 316–335. https://doi.org/10.3102/0162373718764828

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Thurston, A., Cockerill, M., & Craig, N. (2019). Using cooperative learning to close the reading attainment gap for students with low literacy levels for Grade 8/Year 9 students. *International Journal of Educational Research*, *94*, 1–10. https://doi.org/10.1016/j.ijer.2019.02.016

Tables

Table 1. Teacher characteristics by ever tutored.

	Full Sample	Ever Tutored	Never Tutored	Difference
Prop. BA	0.32	0.33	0.32	0.01
Prop. MA	0.38	0.38	0.38	0.00
Prop. MA+	0.08	0.09	0.08	0.01*
Prop. Ed.S.	0.03	0.04	0.03	0.01**
Prop. PhD	0.03	0.02	0.03	-0.01**
Avg. Years Experience	10.84	10.64	10.89	-0.25
Avg. Annual Salary	$63,656	$58,282	$64,964	-$6,682
Pct Black	0.28	0.35	0.27	0.08***
Pct White	0.66	0.60	0.67	-0.07***
Prop. Hispanic	0.02	0.01	0.02	-0.01**
Prop. Asian	0.01	0.01	0.01	0.00
Prop. Other Race/Unknown	0.03	0.03	0.03	0.00
Prop. Elementary Teacher (Gr. 1-5)	0.25	0.38	0.21	0.17***
Prop. Math Teacher	0.05	0.05	0.05	0.00
Prop. Exceptional Ed. Teacher	0.10	0.11	0.10	0.01
Prop. ELA/Reading Teacher	0.06	0.04	0.07	-0.03***
Prop. ECE/Kindergarten Teacher	0.13	0.15	0.12	0.02**
Prop. Social Studies Teacher	0.04	0.02	0.05	-0.03***
Prop. Science Teacher	0.04	0.03	0.04	-0.01*
Prop. Elective Teacher	0.14	0.06	0.16	-0.10***
Prop. English Learner Teacher	0.07	0.07	0.07	0.00
Prop. Teacher-Other Subject	0.02	0.02	0.02	-0.00
Prop. Paraprofessional	0.15	0.13	0.15	-0.02*
N Teachers	5,880	1,139	4,741	

Note. *** p<0.01, ** p<0.05, * p<0.1. Employee characteristics from the employee's most recent semester in the data were used. Annual Salary is missing for 86 employees. Years Experience is only available for teachers. ECE=Early Childhood Education.

Table 2. Tutoring characteristics by semester.

	Full Sample	Spring 2021	Fall 2021	Spring 2022	Fall 2022	Spring 2023
Reading	71%	36%	70%	72%	69%	74%
Math	30%	64%	30%	28%	32%	27%
In Person	82%	0%	43%	73%	86%	93%
Virtual	19%	100%	58%	27%	14%	8%
Volunteer	14%	100%	41%	21%	11%	5%
Paid Tutor	7%	0%	0%	0%	10%	10%
MNPS Teacher	57%	0%	35%	61%	58%	62%
MNPS Parapro.	5%	0%	4%	3%	5%	5%
Other MNPS Employee	17%	0%	19%	15%	17%	18%
During School	57%	100%	90%	68%	50%	49%
Before School	11%	0%	3%	5%	13%	13%
After School	34%	0%	8%	29%	39%	40%
1:1 Tutoring	18%	100%	41%	26%	13%	11%
2:1 Tutoring	22%	0%	15%	22%	18%	27%
3:1 Tutoring	47%	0%	20%	38%	53%	54%
4:1+ Tutoring	13%	0%	24%	15%	17%	8%
Avg. Sessions Attended	21.29	Unknown	15.84	19.75	20.70	23.53
N Students	10,039	131	840	1,772	3,179	4,117

Note. Percentages may not add up to 100% because students could have experienced multiple tutoring modes or were tutored in multiple subjects.

Table 3. RCT treatment effects of a semester of pilot program tutoring on standardized tests and course grades.

	Reading Test Score	Math Test Score	ELA Grade	Math Grade
Intent to Treat (ITT)	-0.048	0.048	-0.331	1.480
	(0.123)	(0.076)	(1.803)	(1.837)
1st Stage	0.910***	0.956***	0.909***	0.954***
	(0.046)	(0.027)	(0.045)	(0.028)
Treatment on the Treated (TOT)	-0.053	0.050	-0.364	1.552
	(0.135)	(0.079)	(1.984)	(1.928)
F-Statistic	399.5	1,294	418.3	1,140
Observations	99	147	98	136

Note. Robust standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. F-statistic is from the TOT estimates. Models include grade fixed effects and the following covariates: a student's instructional modality (remote or unknown), race, gender, economically disadvantaged status, English learner status, disability status, and lagged outcome.

Table 4. RCT treatment effects of a semester of Accelerating Scholars tutoring in Fall 2021 on standardized tests and course grades.

	Reading Test Score	Math Test Score	ELA Grade	Math Grade
Intent to Treat (ITT)	0.014	0.006	1.336*	0.641
	(0.032)	(0.033)	(0.680)	(2.283)
1st Stage	0.322***	0.196***	0.320***	0.263***
	(0.038)	(0.044)	(0.038)	(0.063)
Treatment on the Treated (TOT)	0.043	0.032	4.179*	2.440
	(0.098)	(0.169)	(2.212)	(8.658)
F-Statistic	71.43	20.35	71.30	17.18
Observations	3,289	1,427	3,048	744

Note. Robust standard errors clustered at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. F-statistic is from the TOT estimates. Covariates include student race, gender, economically disadvantaged status, English learner status, disability status, and lagged outcome. 9th graders are excluded from the sample for the analyses where course grade is an outcome. The model where math test score is the outcome also includes a grade-level (e.g. elementary, middle, high school) fixed effect because randomization occurred in grade level blocks.

Table 5. Difference-in-differences treatment effect estimates of Accelerating Scholars tutoring on standardized tests and course grades.

	Test Score				Grades			
	Reading		Math		ELA		Math	
Tutored	0.075***	0.090***	0.014	0.007	-0.734***	-0.134	0.004	0.332
	(0.009)	(0.009)	(0.011)	(0.011)	(0.135)	(0.120)	(0.290)	(0.256)
Student FE	X	X	X	X	X	X	X	X
Semester FE	X	X	X	X	X	X	X	X
School by Grade FE		X		X				
Teacher by School by Course FE						X		X
Observations	184,143	184,143	140,585	140,585	181,756	181,756	123,875	123,875
Adj R Squared	0.728	0.735	0.795	0.800	0.504	0.609	0.500	0.592

Note. Robust standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 9th graders are excluded from the sample for the analyses where course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester or has been tutored in a past semester. FE=fixed effect.

Table 6. Difference-in-differences treatment effect estimates in a restricted sample.

	Test Score				Grades			
	Reading		Math		ELA		Math	
Tutored	0.070***	0.086***	-0.008	-0.015	-1.040***	-0.305**	0.013	0.122
	(0.011)	(0.011)	(0.013)	(0.013)	(0.166)	(0.148)	(0.331)	(0.293)
Student FE	X	X	X	X	X	X	X	X
Semester FE	X	X	X	X	X	X	X	X
School by Grade FE		X		X				
Teacher by School by Course FE						X		X
Observations	179,890	179.890	138,574	138,574	177,276	177,270	122,098	122,097
Adj R Squared	0.729	0.736	0.796	0.801	0.503	0.609	0.501	0.593

Note. Robust standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sample excludes observations of treated students the semester prior to first receiving tutoring. 9th graders are excluded from the sample for the analyses where course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester or has been tutored in a past semester. FE=fixed effect.

Table 7. Difference-in-differences treatment effect estimates using a time-varying definition of treatment.

	Test Score				Grades			
	Reading		Math		ELA		Math	
Tutored	0.042***	0.050***	0.015	0.014	-0.496***	0.138	-0.002	0.364
	(0.008)	(0.008)	(0.011)	(0.011)	(0.125)	(0.111)	(0.290)	(0.257)
Student FE	X	X	X	X	X	X	X	X
Semester FE	X	X	X	X	X	X	X	X
School by Grade FE		X		X				
Teacher by School by Course FE						X		X
Observations	184,143	184,143	140,585	140,585	181,756	181,756	123,875	123,875
Adj R Squared	0.727	0.735	0.795	0.800	0.504	0.609	0.500	0.592

Note. Robust standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 9th graders are excluded from the sample for the analyses where course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester. FE=fixed effect.

Table 8. Difference-in-differences treatment effect estimates by baseline student characteristics.
Panel A. Test Scores

	Reading					Math				
Tutored	0.097***	0.091***	0.081***	0.111***	0.075***	0.002	0.012	-0.001	0.046***	0.029
	(0.009)	(0.011)	(0.010)	(0.013)	(0.016)	(0.011)	(0.014)	(0.012)	(0.016)	(0.023)
Tutored*Spec. Educ.	-0.120***					0.065				
	(0.038)					(0.040)				
Tutored*Econ. Dis.		-0.003					-0.012			
		(0.017)					(0.021)			
Tutored*English Learner			0.037**					0.079**		
			(0.018)					(0.036)		
Tutored*Female				-0.041**					-0.073***	
				(0.0162)					(0.0209)	
Tutored*Black					-0.004					-0.033
					(0.021)					(0.027)
Tutored*Hispanic					0.053**					-0.025
					(0.021)					(0.032)
Tutored*Asian					0.022					0.082
					(0.037)					(0.059)
Tutored*Other Race					0.116					-0.023
					(0.106)					(0.133)
Observations	184,143	184,143	184,143	184,143	184,143	140,585	140,585	140,585	140,585	140,585
Adj R Squared	0.735	0.735	0.735	0.735	0.735	0.800	0.800	0.800	0.800	0.800

Panel B. Course Grades

	ELA					Math				
Tutored	-0.0787	-0.296**	-0.214	-0.457***	-0.378*	0.540**	0.545*	0.336	-0.328	-0.946*
	(0.123)	(0.147)	(0.140)	(0.170)	(0.208)	(0.265)	(0.320)	(0.272)	(0.363)	(0.496)
Tutored*Spec. Educ.	-0.779					-2.450***				
	(0.501)					(0.858)				
Tutored*Econ. Dis.		0.390*					-0.461			
		(0.225)					(0.492)			
Tutored*English Learner			0.358					0.0376		
			(0.245)					(0.730)		
Tutored*Female				0.641***					1.249***	
				(0.219)					(0.477)	
Tutored*Black					0.193					1.797***
					(0.275)					(0.596)
Tutored*Hispanic					0.574**					0.992
					(0.287)					(0.726)
Tutored*Asian					0.151					2.503*
					(0.522)					(1.379)
Tutored*Other Race					0.151					2.454
					(1.475)					(2.257)
Observations	179,876	179,876	179,876	179,876	179,876	122,123	122,123	122,123	122,123	122,123
Adj R Squared	0.608	0.608	0.608	0.608	0.608	0.589	0.589	0.589	0.589	0.589

Note. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Models include student and semester fixed effects. Models where test score is the outcome include a school-by-grade fixed effect. Models where course grade is the outcome include a teacher by school by course fixed effect. 9th graders are excluded from the sample for the analyses where course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester or has been tutored in a past semester. Baseline characteristics are from the semester where test score is first observed for the student.

Table 9. Difference-in-differences treatment effect estimates by program characteristics.

Panel A. Test Scores

	Reading				Math			
Tutored	0.065***	0.047***	0.070***	0.051***	0.010	0.008	0.015	0.010
	(0.012)	(0.011)	(0.019)	(0.009)	(0.016)	(0.017)	(0.022)	(0.012)
Tutored by MNPS Teacher	-0.025*				0.007			
	(0.015)				(0.021)			
Tutored During Sch. Hrs.		-0.008				0.009		
		(0.015)				(0.021)		
Tutored in Person			-0.025				-0.002	
			(0.020)				(0.025)	
Tutored 1:1				-0.005				0.017
				(0.020)				(0.019)
Student FE	X	X	X	X	X	X	X	X
Semester FE	X	X	X	X	X	X	X	X
School by Grade FE	X	X	X	X	X	X	X	X
Observations	184,143	184,143	184,143	184,143	140,585	140,585	140,585	140,585
Adj R Squared	0.735	0.735	0.735	0.735	0.800	0.800	0.800	0.800

Panel B. Course Grades

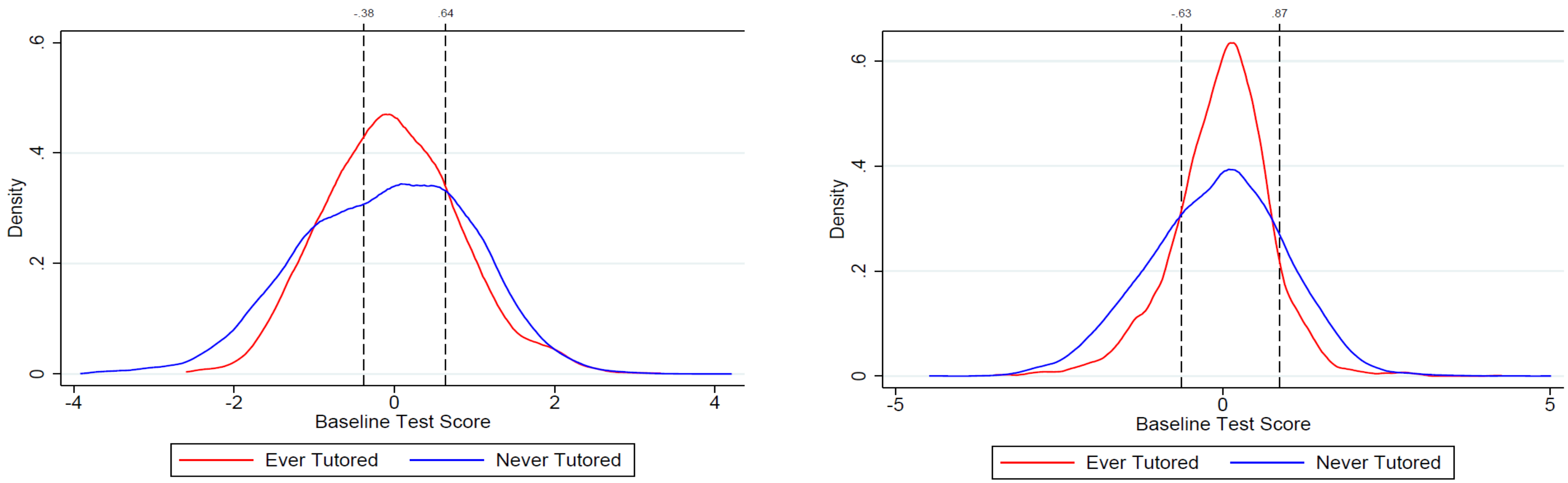
	ELA				Math			
Tutored	0.093	0.070	0.035	0.201*	0.153	-0.239	0.097	0.524*
	(0.161)	(0.150)	(0.252)	(0.118)	(0.422)	(0.380)	(0.762)	(0.268)
Tutored by MNPS Teacher	0.077				0.326			
	(0.201)				(0.512)			
Tutored During Sch. Hrs.		0.136				0.980**		
		(0.205)				(0.480)		
Tutored in Person			0.124				0.307	
			(0.274)				(0.800)	
Tutored 1:1				-0.399				-0.920**
				(0.269)				(0.465)
Student FE	X	X	X	X	X	X	X	X
Semester FE	X	X	X	X	X	X	X	X
Teacher by School by Course FE	X	X	X	X	X	X	X	X
Observations	181,756	181,756	181,756	181,756	123,875	123,875	123,875	123,875
Adj R Squared	0.609	0.609	0.609	0.609	0.592	0.592	0.592	0.592

Note. Robust standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 9th graders are excluded from the sample when course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester. FE=fixed effect.

Figures

Figure 1. Baseline test score distribution by subject tutored.

Panel A. Reading *Sample* *Panel B. Math Sample*



Note. Test scores are measured in standard deviation units. N. Students ELA Sample=39,043. N. Students Math Sample=29,875. A student's baseline test score is from the semester the student in first observed in our difference-in-differences analytic sample. Vertical lines represent test scores standardized within MNPS at the 15th and 60th percentile nationally.

Figure 2. Number of students tutored by grade, subject, and semester.

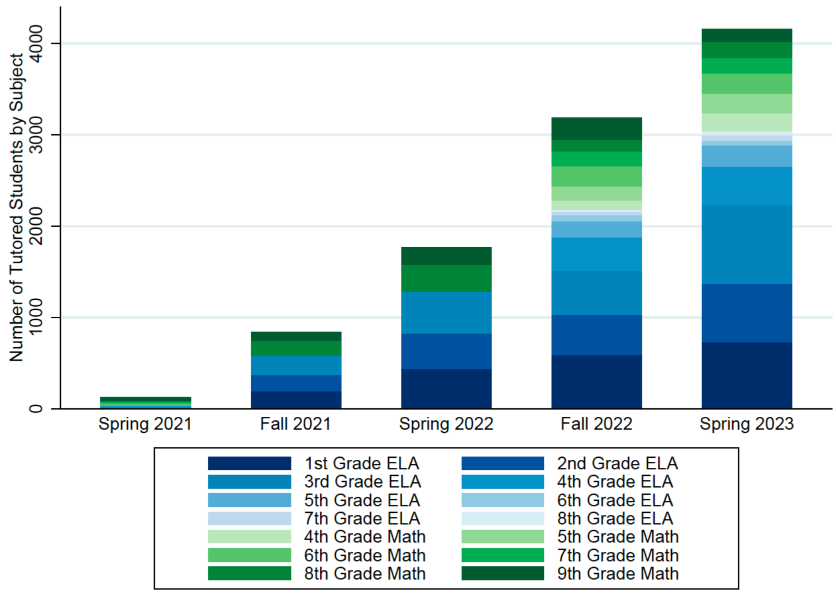
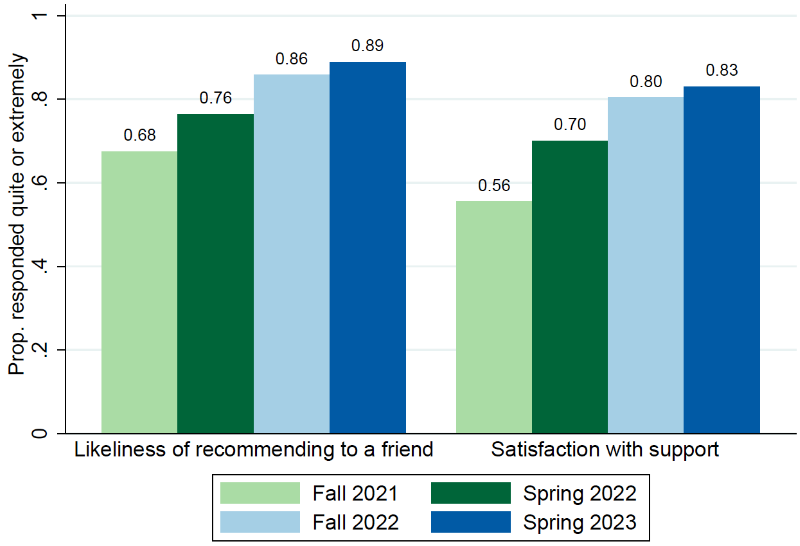
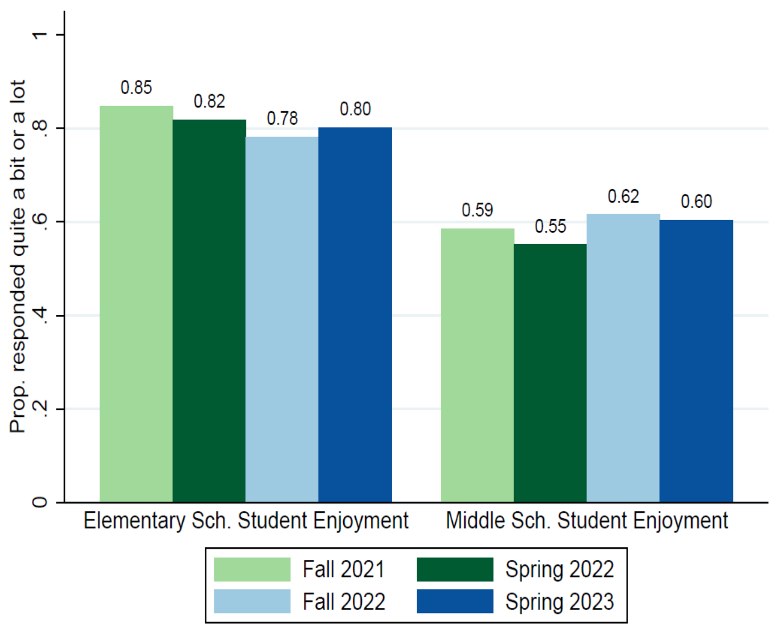


Figure 3. Reported tutor and student satisfaction by semester.

Panel A. Tutors



Panel B. Students



Note. Tutors were asked: How likely are you to recommend participating in the MNPS Accelerating Scholars tutoring program to a friend? Overall, how satisfied were you with the amount of tutoring support you received? Elementary students were asked: How much do you enjoy your tutoring sessions? Middle school (6th-9th grade) students were asked: How much did you enjoy learning from your tutor? Elementary students were not given the quite a bit option.

Figure 4. Number of tutors by type and semester.

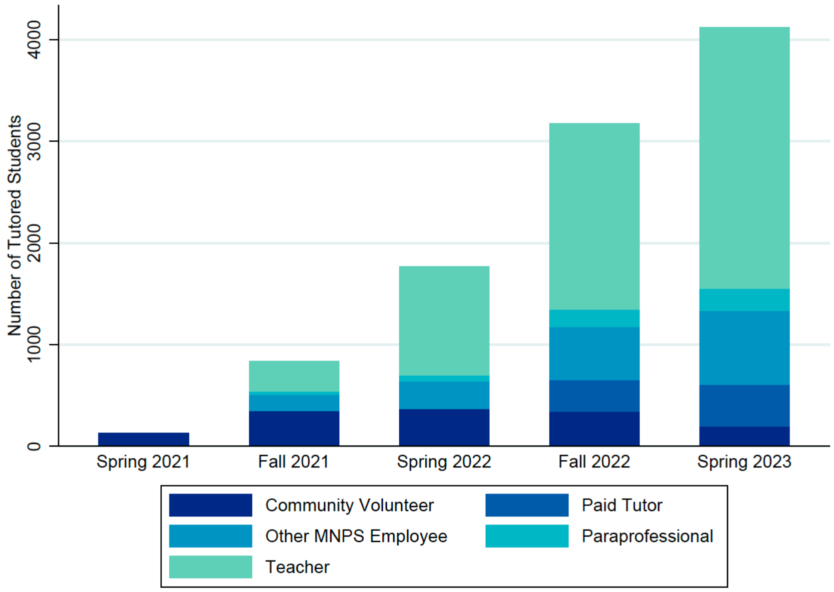
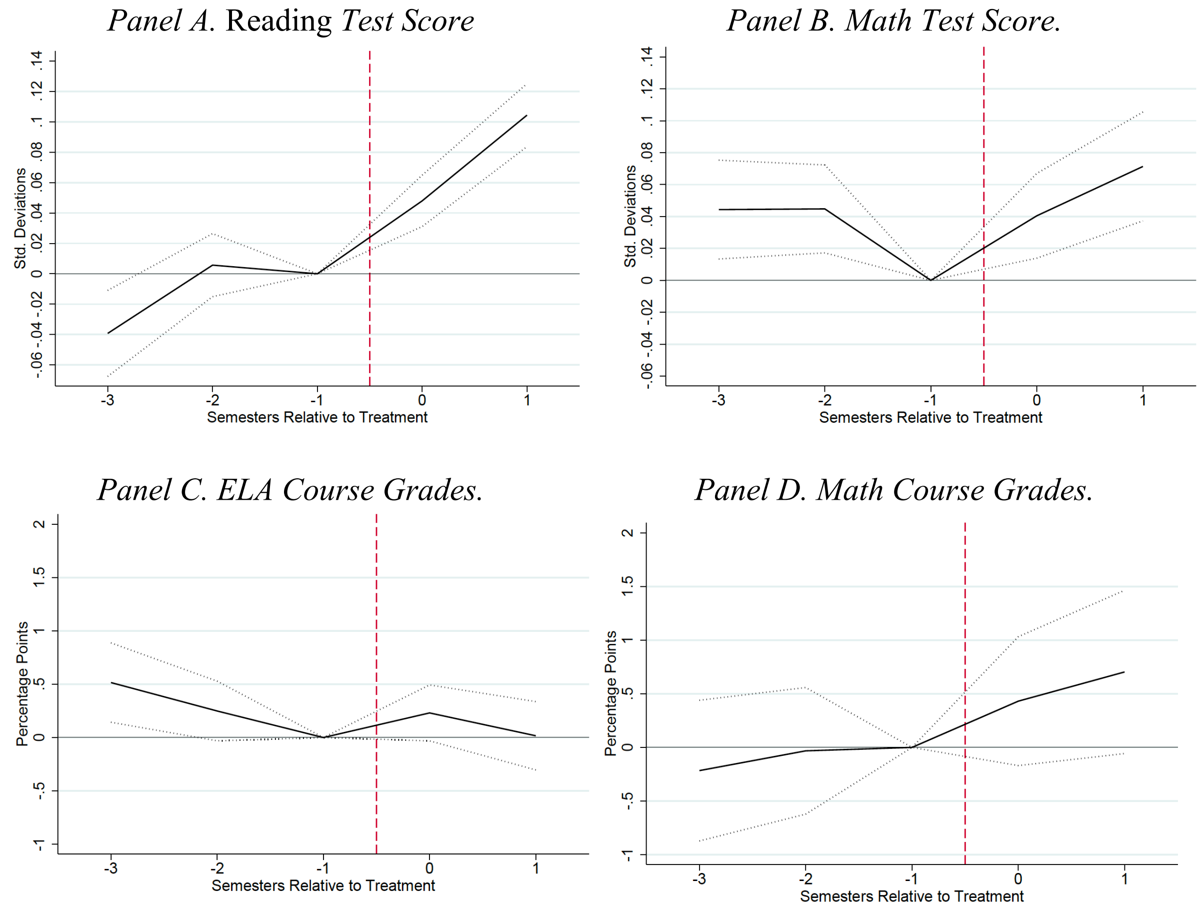
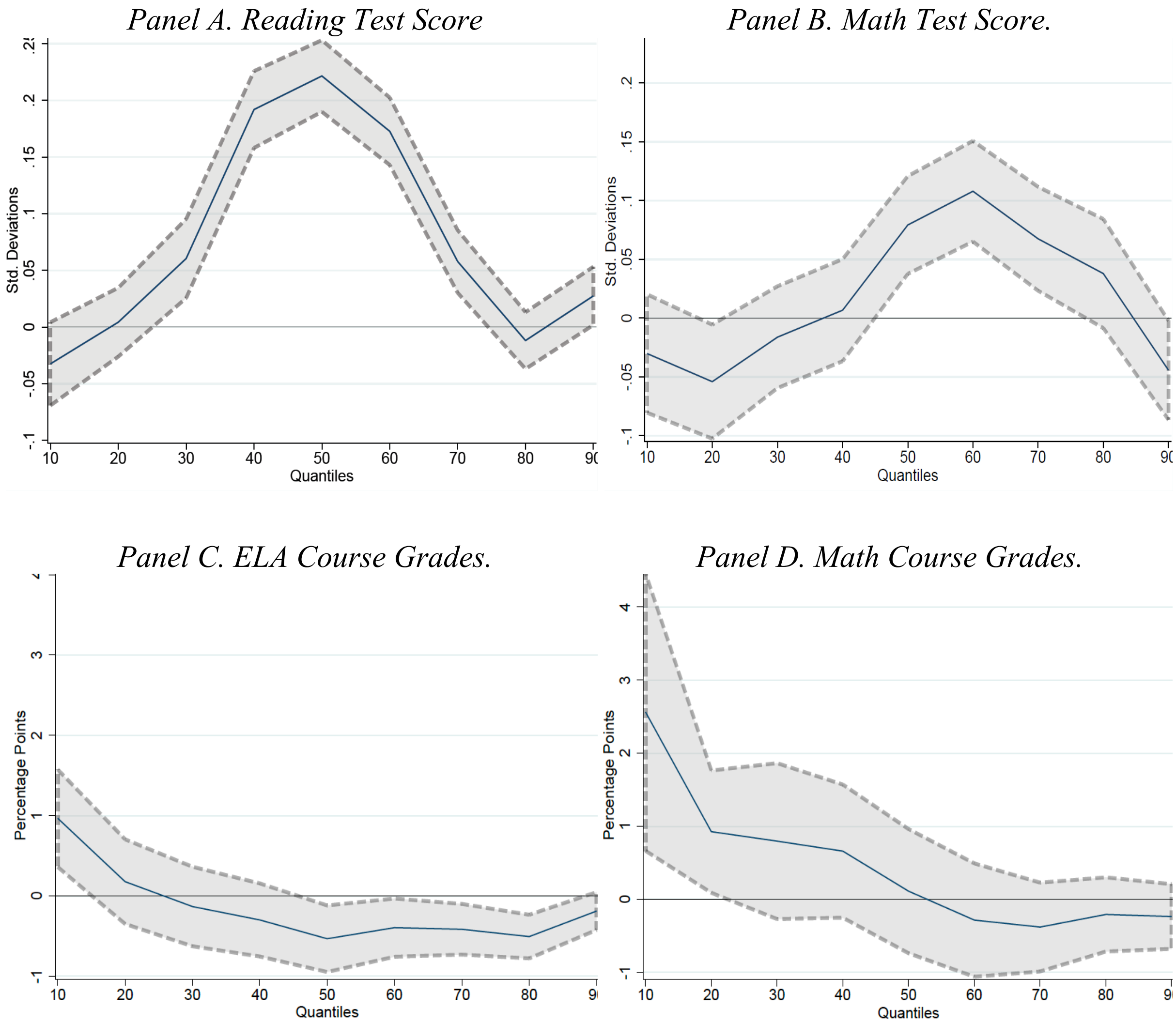


Figure 5. Event study treatment effect estimates.



Note. 95% confidence intervals. Robust standard errors clustered at the student level. Panel A N. Observations=184,143. Panel B N. Observations=140,585. Panel C N. Observations=181,756. Panel D. N. Observations=123,875. Models include student and semester fixed effects. Models where test score is the outcome include a school-by-grade fixed effect. Models where course grade is the outcome include a teacher by school by course fixed effect. 9th graders are excluded from the sample for the analyses where course grade is an outcome.

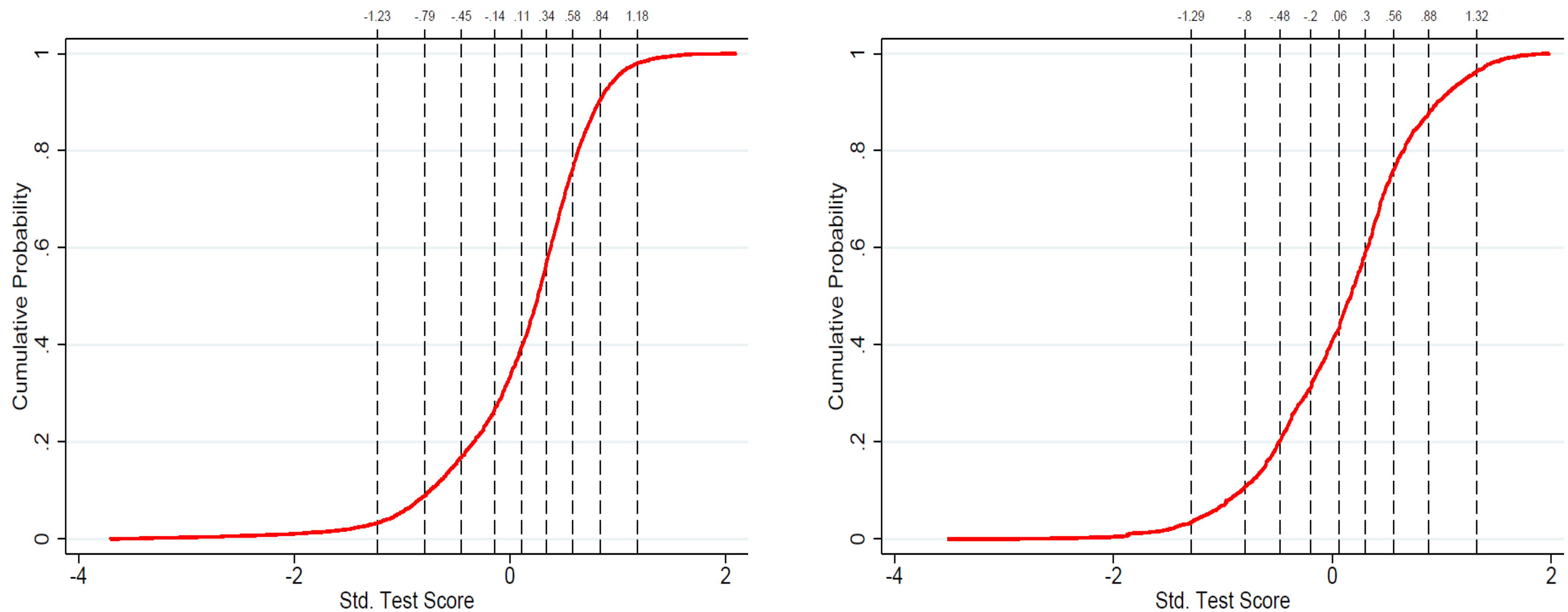
Figure 6. Difference-in-differences quantile treatment effect estimates



Note. 95% confidence intervals. Robust standard errors clustered at the student level. Panel A N. Observations=184,143. Panel B N. Observations=140,585. Panel C N. Observations=181,756. Panel D. N. Observations=123,875. Models include student and semester fixed effects. Models where test score is the outcome include a school-by-grade fixed effect. Models where course grade is the outcome include a teacher by school by course fixed effect. 9th graders are excluded from the sample for the analyses where course grade is an outcome.

Figure 7. Cumulative density function of standardized test scores of tutored students

Panel A. Reading *Tutored Students* *Panel B. Math Tutored Students*



Note. Panels A and B include 9,109 and 3,193 student-year observations of tutored students after initial tutoring respectively. Dashed lines represent test scores at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of the analytic sample.

Appendix A

Accelerating Scholars partnered with TNTP to develop set of instructional materials that were adapted from or aligned to core instruction. At the start of the program, reading tutors used the Tennessee Foundational Skills Curriculum Supplement (TNFSCS) in 1st and 2nd grades and Core Knowledge Language Arts (CKLA) in 3rd grade. Mat tutors used *Illustrative Math*. As the program evolved, it worked with TNTP to develop pacing guides and workbooks for the TNFSCS and CKLA curriculum. The district also developed materials using *Newsela* aligned to the *myPerspectives* core content in 6th-8th grade English Language Arts. The district switched to using the Zearn math Curriculum in Spring 2022 across 4th through 8th grade given this new TN ALL Corps requirement. In 9th grade, Accelerating Scholars used *Math by Example* and *Algebra by Example* to provide aligned tutoring content for 9th grade Integrated Math I.

Appendix B: Appendix Tables and Figures

Appendix Table B1. Student characteristics by ever tutored.

	Reading			Math		
	Tutored	Not Tutored	Difference	Tutored	Not Tutored	Difference
Prop. Female	0.517	0.481	0.036***	0.531	0.486	0.045***
Prop. Black	0.442	0.352	0.090***	0.539	0.383	0.156***
Prop. Hispanic	0.277	0.323	-0.046***	0.222	0.315	-0.093***
Prop. White	0.237	0.276	-0.039***	0.203	0.256	-0.053***
Prop. Asian	0.039	0.045	-0.006*	0.028	0.042	-0.014***
Prop. Other Race	0.005	0.005	0.000	0.008	0.004	-0.004***
Prop. Econ. Dis.	0.458	0.422	0.036***	0.480	0.418	0.062***
Prop. Special Educ.	0.062	0.114	-0.052***	0.078	0.123	-0.045***
Prop. English Learners	0.241	0.292	-0.051***	0.107	0.241	-0.134***
Avg. Baseline Ach.	-0.029	-0.103	0.074***	-0.017	-0.091	0.074***
Avg. Baseline Grade	85.036	82.809	2.23***	81.518	81.132	-0.385
N Students	4,653	34,390		2,037	27,837	

Note. *** p<0.01, ** p<0.05, * p<0.1. Student characteristics are from the semester the student is first observed in our difference-in-differences analytic samples. Baseline achievement and course grade are for the tutored subject. Baseline course grades are missing for 14% of the ELA sample and 23% of the math sample.

Appendix Table B2. Survey response rates.

Survey	Statistic	Fall 2021	Spring 2022	Fall 2022	Spring 2023
Elementary Student	Pct. Responded	48%	60%	54%	72%
	N. Tutored Students	586	1284	2302	2362
Middle & High School Student	Pct. Responded	51%	33%	32%	62%
	N. Tutored Students	254	488	282	524
Tutor	Pct. Responded	33%	55%	60%	83%
	N. Tutors	500	816	997	1215
Tutor Lead	Pct. Responded	79%	60%	77%	87%
	N. Tutor Leads	47	86	93	99

Appendix Table B3. Spring 2021 student-level RCT balance tests.

	Treat	Control	Difference
Prop. Female	0.566	0.580	-0.014
Prop. Black	0.434	0.406	0.028
Prop. Asian	0.052	0.015	0.037*
Prop. Hispanic	0.169	0.196	-0.027
Prop. White	0.316	0.370	-0.053
Prop. Two or More Races	0.029	0.015	0.010
Prop. Econ. Dis.	0.309	0.326	-0.017
Prop. English Learner	0.007	0.036	-0.029
Prop. Special Education	0.015	0.036	-0.022
Avg. Prior Test Score	0.329	0.209	0.120**
Avg. Prior Course Grade	86.960	87.030	-0.074
N Students	136	138	274

Note. *** p<0.01, ** p<0.05, * p<0.1. 278 students were randomized into treatment. 4 (1 control, 3 treatment) do not have any demographic or prior test score information.

Appendix Table B4. Fall 2021 school-level cluster RCT balance tests.

	Treat	Control	Difference
Avg. Prop. Female	0.486	0.493	-0.007
Avg. Prop. Black	0.429	0.426	0.003
Avg. Prop. Hispanic	0.318	0.301	0.017
Avg. Prop. White	0.211	0.232	-0.021
Avg. Prop. Asian	0.038	0.036	0.002
Avg. Prop. Other Race	0.004	0.005	-0.001
Avg. Prop. Economically Disadvantaged	0.487	0.471	0.017
Avg. Prop. Special Education	0.155	0.152	0.003
Avg. Prop. EL	0.267	0.243	0.024
Avg. Std. Math Score	-0.142	-0.068	-0.074
Avg. Std. Reading Score	-0.153	-0.069	-0.085
Avg. Total Students	583.8	564.4	19.43
N Schools	29	39	68

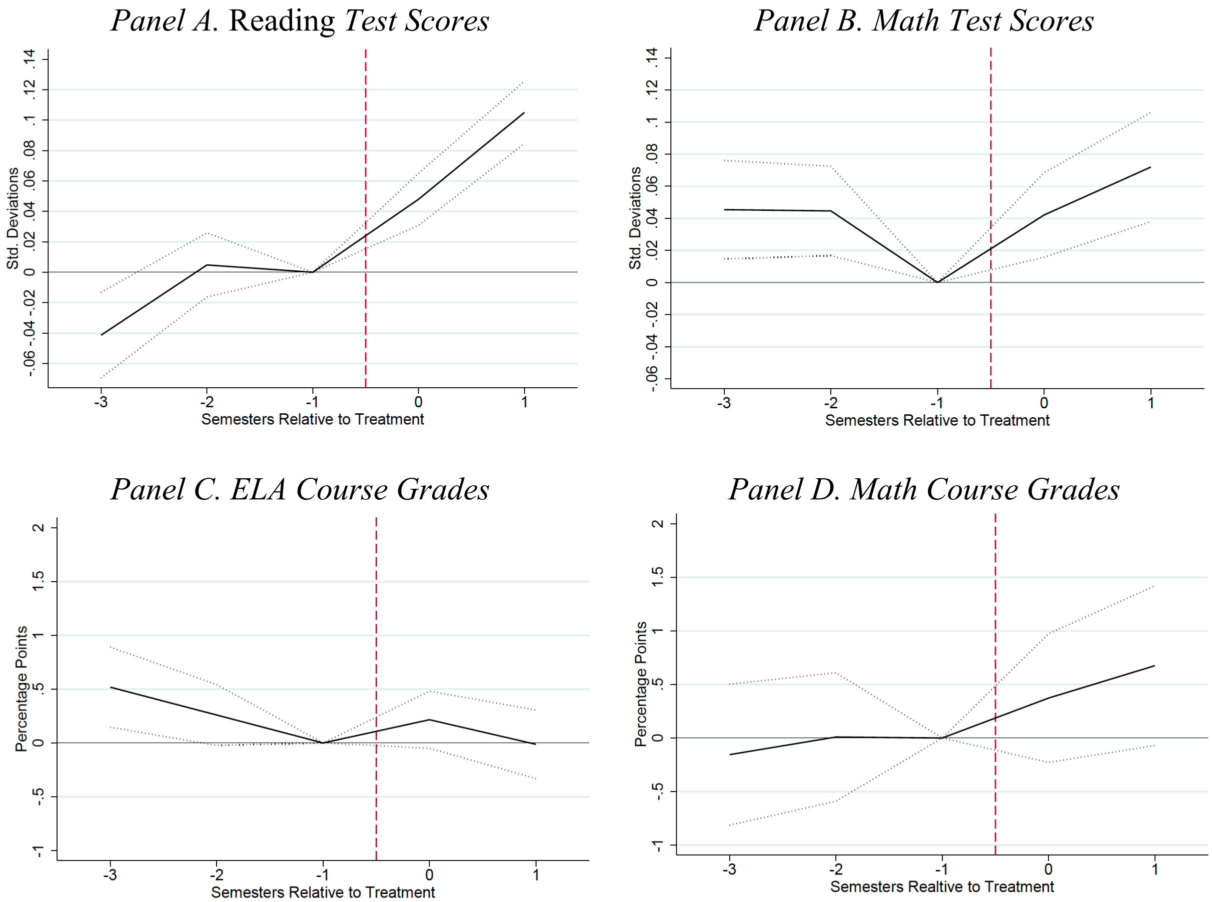
Note. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table B5. Main and stacked event study treatment estimates.

	Reading Test Score		Math Test Score		ELA Course Grade		Math Course Grade	
	Main	Stacked	Main	Stacked	Main	Stacked	Main	Stacked
Pre								
t=-5	0.019	0.015	0.141***	0.146***	1.661***	1.695***	2.784***	2.769***
	(0.033)	(0.033)	(0.035)	(0.036)	(0.399)	(0.395)	(0.682)	(0.679)
t=-4	-0.026	-0.031	0.108***	0.110***	0.511*	0.526*	1.122***	1.147**
	(0.019)	(0.019)	(0.021)	(0.021)	(0.265)	(0.265)	(0.418)	(0.415)
t=-3	-0.039***	-0.041***	0.044***	0.045**	0.515***	0.518**	-0.216	-0.155
	(0.014)	(0.014)	(0.016)	(0.016)	(0.190)	(0.190)	(0.334)	(0.336)
t=-2	0.006	0.005	0.045***	0.045**	0.249*	0.259	-0.0318	0.009
	(0.011)	(0.011)	(0.014)	(0.014)	(0.143)	(0.144)	(0.301)	(0.306)
Post								
t=0	0.048***	0.048***	0.041***	0.042**	0.231*	0.216	0.431	0.373
	(0.009)	(0.009)	(0.014)	(0.013)	(0.134)	(0.135)	(0.306)	(0.306)
t=1	0.104***	0.105***	0.071***	0.072***	0.017	-0.012	0.703*	0.675
	(0.011)	(0.011)	(0.017)	(0.017)	(0.163)	(0.162)	(0.388)	(0.380)
t=2	0.184***	0.191***	0.000	-0.002	-0.515**	-0.562*	-2.408	-2.398
	(0.015)	(0.014)	(0.026)	(0.026)	(0.239)	(0.232)	(2.158)	(1.885)
t=3	0.193***	0.199***	-0.057	-0.060	-0.654*	-0.684	3.093	3.126
	(0.023)	(0.022)	(0.065)	(0.065)	(0.364)	(0.361)	(3.646)	(3.493)
Observations	184,143	668,745	140,585	532,586	181,756	658,570	123,875	470,624
Adj R Squared	0.735	0.747	0.800	0.805	0.609	0.621	0.592	0.605

Note. Robust standard errors clustered at the student level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Models include student and semester fixed effects. Models where test score is the outcome include a school-by-grade fixed effect. Models where course grade is the outcome include a teacher by school by course fixed effect. 9th graders are excluded from the sample for the analyses where course grade is an outcome. Tutored equals one if a student if a student is being tutored this semester or has been tutored in a past semester.

Appendix Figure B1. Stacked Event Study Treatment Estimates



Note. 95% confidence intervals. Robust standard errors clustered at the student level. Panel A N. Observations=668,745. Panel B N. Observations=532,586. Panel C N. Observations=658,570. Panel D. N. Observations=470,624. Models include student and semester fixed effects. Models where test score is the outcome include a school-by-grade fixed effect. Models where course grade is the outcome include a teacher by school by course fixed effect. 9th graders are excluded from the sample for the analyses where course grade is an outcome.

Appendix C: Test Score Outcomes

During our study period, the 2020-21 to 2022-23 school years, students in grades 3-8, as well as students enrolled in high school courses with end-of-course exams, took the state standardized test, Tennessee Comprehensive Assessment Program (TCAP) assessments, each school year in the late spring. Also, MNPS administered formative assessments multiple times a year to its kindergarten-9th grade students. The formative assessment used and how often it was administered varied over time. In 2020-21, MNPS administered the Measures of Academic Progress (MAP) assessment in August 2020, January 2021, and May 2021 to all K-9 students. In 2021-22, MAP was administered to 2-8 grade students in August 2021, February 2022, and May 2022. However, the May MAP test administration was optional, and most schools did not participate in it. Instead of MAP, Kindergarten and first grade students were required to take the Fastbridge (FAST) Early Reading and Math assessments in August, January, and May of both the 2021-22 and 2022-23 school years. During the 2022-23 school year, MNPS students in grades 2-9, took the FASTtrack Math and Reading assessments in the fall, winter, and spring.

We describe each of these assessments below.

TCAP- assesses mastery on Tennessee Academic Standards for that subject and grade level at the end of the school year. Assessments last multiple hours and are taken over the course of multiple days. In grades 3-8, students take state standardized assessments in English Language Arts, Mathematics, and Science. In grades 6-8, students also take a social studies assessment. While high school students also take end-of-course exams as part of TCAP, we do not use test scores from these assessments because they depend on the courses students are enrolled in. For students in grades 3-8, we only use the composite score (which we standardize within grade, subject, and year across all MNPS students) on the math and reading assessments because those are the tutored subjects in our study.

MAP- measures students' reading and math knowledge. It is a computer-adaptive test that takes 45-60 minutes to complete. MNPS used MAP tests as a screening tool for academic intervention. We use students' math and reading RIT scores standardized within subject, grade, and test administration in our study.

FAST Early Math/Reading- assesses early reading and developing math skills typically used in kindergarten and 1st grade. Skills measured include (but are not limited to) phonemic awareness, fluency, alphabetic principles, numeral identification, counting, addition, and subtraction. During each test administration, teachers orally administer 3-4 subtests for each subject that last less than five minutes each individually to each student. Subtests vary from administration to administration. MNPS uses this assessment as a screening tool for academic intervention. We use the composite score across each subject's subtests, which we standardize within grade and test score administration.

FASTtrack Math/Reading- are screening assessments that can predict future performance and indicate instruction needed. For each subject. there are two assessment components to FASTtrack, a curriculum-based measure that takes less than five minutes to complete and a 30-60 question computer-adaptive test (aRead or aMath) that take 10-30 minutes to complete. We only use the composite scores from the aRead and aMath assessments in our study. We standardize these within subject, grade, and test administration across all MNP students.

For a list of outcomes used by semester, grade, and subject, please see Appendix Table C1.

Appendix Table C1. Test score outcome by grade and semester

	Fall 2020	Spr. 2021	Fall 2021	Spr. 2022	Fall 2022	Spr. 2023
Kindergarten	Winter 2021 MAP Read.	Spr. 2021 MAP Read.	Winter 2022 FAST Early Read.	Spr. 2022 FAST Early Read.		
1st Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 FAST Early Read.	Spr. 2022 FAST Early Read.	Winter 2023 FAST Early Read.	Spr. 2023 FAST Early Read.
2nd Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.		Winter 2023 FAST aRead	Spr. 2023 FAST aRead
3rd Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.	Spr. 2022 TNReady Math/ELA	Winter 2023 FAST aRead	Spr. 2023 FAST aRead
4th Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.	Spr. 2022 TNReady Math/ELA	Winter 2023 FAST aRead/aMath	Spr. 2023 FAST aRead/aMath
5th Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.	Spr. 2022 TNReady Math/ELA	Winter 2023 FAST aRead/aMath	Spr. 2023 FAST aRead/aMath
6th Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.	Spr. 2022 TNReady Math/ELA	Winter 2023 FAST aRead/aMath	Spr. 2023 FAST aRead/aMath
7th Grade	Winter 2021 MAP Math/Read.	Spr. 2021 MAP Math/Read.	Winter 2022 MAP Math/Read.	Spr. 2022 TNReady Math/ELA	Winter 2023 FAST aRead/aMath	Spr. 2023 FAST aRead/aMath
8th Grade	Winter 2021 MAP Math	Spr. 2021 MAP Math	Winter 2022 MAP Math	Spr. 2022 TNReady Math	Winter 2023 FAST aRead/aMath	Spr. 2023 FAST aRead/aMath
9th Grade	Winter 2021 MAP Math*	Spr. 2021 MAP Math*	Winter 2022 MAP Math		Winter 2023 FAST aMath	Spr. 2023 FAST aMath

Note. 9th grade outcomes in Fall 2020 and Spr. 2021 school year are only used for the pilot RCT. For the Fall 2021 cluster RCT, the lagged outcome is Fall 2021 FAST Early Read. in grade 1 and MAP in the other tutored grades.