The effects of a university-led high impact tutoring program on low-achieving high school students: A three-year randomized controlled trial

Abstract

This study examines a university-led high impact tutoring program at seven high schools by using a randomized controlled trial design. The treatment group (n = 525) participated in high impact tutoring (i.e., groups of 2:1 or 3:1) while the control group (n = 438) attended a remedial mathematics course. The treatment group showed a difference of nearly a half-year of learning (0.13 *SD*) compared to the control group. We also found no evidence that 2:1 student-tutor groups were more effective than 3:1 groups. Although the program produced strong effects, it was delivered at a high cost. Future work is needed to investigate strategies for reducing the cost of high impact tutoring while maintaining effectiveness.

Keywords. High impact tutoring; high dosage tutoring; remedial instruction; instructional time; randomized controlled trial

Introduction

Low performance in mathematics during high school can limit students' academic and life opportunities (Dougherty & Fleming, 2012; Joensen & Nielsen, 2009). Students exhibiting below-average achievement in mathematics when entering high school are less likely to enroll in courses that prepare them for college and high-earning university majors (Aughinbaugh, 2012; Cha, 2015; Long et al., 2009; Trusty & Niles, 2003). These early setbacks appear to persist over the long run (Cohodes & Walters, 2019; Rose & Betts, 2004). High school mathematics achievement has been linked to income, employment, and incarceration in adulthood (Chetty et al., 2014; Duncan et al., 2007; Heckman et al., 2006). Despite substantial evidence underscoring the significance of mathematics achievement, a large percentage of US students reach high school scoring well below proficiency benchmarks (National Assessment for Educational Progress [NAEP], 2024). Disruptions to learning caused by the COVID-19 pandemic have exacerbated this problem. On the NAEP, often referred to as the Nation's Report Card, post-pandemic results for high school-bound eighth-grade students have shown mathematics scores plummeting to levels not observed since 1990 (NAEP, 2024). These declines have exacerbated socioeconomic achievement gaps, being particularly severe among low-achieving and low-income students (NAEP, 2022; 2024).

Poor mathematics achievement is a concerning problem with social and economic ramifications. If learning declines are not remediated to pre-pandemic levels, Hanushek (2023) estimates the post-pandemic cost of a lower-skilled workforce will lead to an aggregate loss of $28 trillion for the US economy. Such striking forecasts indicate that aggressive measures may be needed to raise the achievement of students before they exit the K-12 education system. Responding to this problem is challenging though. Students in need of academic support in high school typically have less access to quality instructional experiences that might help to remediate academic skills (Flores, 2007). The current levels of low mathematics achievement may diminish the potential for the education system to serve as an equalizer of economic and social opportunities since it is low-income students who disproportionately fall behind their peers (Kotok, 2017; NAEP, 2024; Tyson et al., 2007).

Under the Coronavirus Aid, Relief, and Economic Security (CARES) Act, the federal government allocated Elementary and Secondary School Emergency Relief (ESSER) funds for post-pandemic academic recovery efforts (Office of State and Grantee Relations, 2024). Many districts used these funds for “high leverage” instructional programs that increased instructional time – approximately 40% of school districts used their ESSER funds to deliver variations of high impact tutoring models (Jordan et al., 2022). High impact tutoring programs (also referred to as high dosage tutoring) have emerged as one of the most promising strategies for remediating the outcomes of low-achieving students (Guryan et al, 2021). These programs are characterized by intensive small group instruction during the school day in which one trained tutor works with one to four students over at least a ten-week period (Robinson et al., 2021). To be considered a “high impact” tutoring program, sessions must occur at least three times per week and last for 30-60 minutes each session (Kraft et al., 2024). Rigorous evaluations of high impact tutoring programs have exhibited strong gains in mathematics and reading, student attendance, and social and life outcomes among academically underperforming students. In a meta-analysis of 96 randomized controlled trials, researchers reported a pooled effect size of 0.37 SD on academic achievement for high impact tutoring programs (Nickow et al., 2020). Yet, bringing quality high impact tutoring programs to scale may be difficult. Recent meta-analytic research aggregating the results of 282 randomized controlled trials demonstrated that even though high impact tutoring is effective overall, the observed academic benefits of large tutoring programs are not as strong as those of small programs (Kraft et al., 2024)

Key limitations in the literature also remain. At the high school level, randomized experimental research is lacking despite some credible RCT evidence pointing to substantial academic gains for high school students who participate in high impact tutoring programs (de Ree et al., 2021; Guryan et al., 2021). Additionally, few studies disentangle the effects of high impact tutoring from the added instructional time that it provides (de Ree et al., 2021), leaving questions about potential opportunity costs. These questions are important because research has found that supplementary instructional time in core subjects can be beneficial for low achieving high school students (Angrist et al., 2016; Cattaneo et al., 2017; Cohodes & Parham, 2021; Lester & Naff, 2016). High impact tutoring also demands considerable human and financial resources. If gains from high impact tutoring are mostly attributable to providing students with increased instructional time, high schools might deploy resources for academic remediation differently.

The purpose of this study is to investigate the effects of a university-led model of high impact tutoring for high school students by using a randomized controlled design for three separate cohorts of ninth-grade students at seven high schools in Oklahoma. The analyses in this study address the following two questions:

Research Question 1. *Does a university-led high impact tutoring program produce stronger academic outcomes than a remedial mathematics course for students who enter ninth grade achieving below grade level in mathematics?*

Research Question 2. *Is there a difference in academic outcomes for tutored students based on student-tutor group size (i.e.,2:1 versus 3:1), hours tutored, and student background characteristics?*

In the pooled sample, students (n = 525) in the treatment group participated in high impact tutoring (i.e., groups of 2:1 or 3:1) three class periods a week for an entire academic year. In the control group, students (n = 438) attended a remediation mathematics course. This study's design builds on previous experimental studies of high impact tutoring by using a remedial mathematics class for the control condition. This control condition allows us to distinguish the effects of high impact tutoring from the added instructional time that it offers. This design might also create a higher threshold for observing program effects than in studies using business-as-usual control conditions that do not provide instructional time in mathematics. The analyses also test for differences in outcomes between 2:1 and 3:1 student-tutor group sizes, providing evidence on whether larger tutor-student group sizes can be used to expand the reach of high impact tutoring.

This study has key research and policy implications given the expansion of high impact tutoring models as well as ongoing post-pandemic academic recovery initiatives occurring in schools across the United States.

Self-determination Theory and High Impact Tutoring

The principles of Self-determination Theory offer compelling support for why high impact tutoring can accelerate student learning (Coyle et al., 2017; Howard et al., 2021; Taylor et al., 2014). Self-determination Theory contends that human motivation stems from innate psychological needs for autonomy, competence, and relatedness (Deci & Ryan, 2012). Autonomy refers to one's need to feel in control over the actions one takes while competence describes the need to feel effective by mastering tasks and learning new skills (Deci & Ryan, 2012). When individuals feel competent, they are expected to persist in their learning. Along with autonomy and competence, the notion of relatedness emphasizes an inherent need to have a connection with others and to feel a sense of belonging (Deci & Ryan, 2012). When individuals feel a strong sense of relatedness, they will, in theory, feel supported and understood.

Applying Self-determination Theory to the context of high impact tutoring, it is conceivable that tutors can respond to individual interests, concerns, and questions to a greater degree than what is possible in whole classroom settings. As a result, the work of tutors can help to increase students' sense of purpose, ownership, and overall autonomy, potentially heightening motivation to learn (Adams & Khojasteh, 2018; Gershenson, 2016; Simons et al., 2010). By expanding instructional time, high impact tutoring offers students opportunities to become competent in mathematics by practicing problems and mastering new skills that they are personally ready to learn (Cerasoli & Ford, 2014). High impact tutoring also has a relational component. Students who study with a tutor for a sustained period may be able to form a personal bond with their tutor. In programs that employ tutors who are near peers, tutors may be able to relate to tutees, building strong relationships with them over time (Colvin, 2007; Duran, 2017; Roscoe & Chi, 2000). Such interpersonal connections can hypothetically fulfill developmental needs pertaining to social attachment and relatedness with others, which can then be converted into other assets, such as greater motivation to learn (Adams et al., 2018; Lohmeier & Lee, 2011).

Empirical Research on High Impact Tutoring

In the empirical literature, high-impact tutoring has shown robust evidence of being a valuable personalized instructional intervention (Kraft et al., 2024). In a comprehensive review, rigorous experimental studies of high impact tutoring show large positive effects on student achievement in mathematics and reading (Kraft et al., 2024; Nickow et al., 2020). Kraft et al. (2024) combined the results of 282 randomized controlled trials of high impact tutoring programs, finding an overall pooled effect size of 0.42 SD. These results were driven by large effects from literacy tutoring programs during elementary school. Results from this meta-analytic work found that large programs had smaller effects, ranging from 0.21 SD for tutoring programs serving 400-499 students and 0.16 SD for programs serving 1,000 students or more. Among effective program characteristics, high impact tutoring seems to be most beneficial when delivered in-person in "doses" consisting of three or more sessions per week (Robinson et al., 2021). Many highly effective high impact tutoring programs operate for an entire school year (Guryan et al., 2021) and use 3:1 student-tutor ratios or less (Kraft et al., 2024**).** Along with these characteristics, the use of high-quality instructional materials that are aligned with course content seems to create the conditions for success (Robinson et al., 2021).

While outcomes from high impact tutoring are encouraging, few US-based randomized studies have investigated the effects of high impact tutoring on mathematics achievement during high school. Remediating mathematics' skills in high school has proven challenging (Heinrich et al., 2019). Low-achieving students tend to have significant gaps in foundational knowledge that prevent them from being successful in their mathematics courses (Siegler & Braithwaite, 2017). At the high school level, existing studies of high impact tutoring programs show smaller effects than those at the elementary school level (Nickow et al., 2020; Hickey et al., 2019) although high school programs that offer more frequent tutoring have exhibited stronger results (Kraft et al., 2024). For example, de Ree et al. (2021) tested 50-minute daily tutoring sessions over a 16-week period and recorded large gains in mathematics test scores (0.28-0.44 SD) in a small sample of 98 high school students in the Netherlands. Students in the tutoring treatment group replaced their regularly scheduled classes (excepting their mathematics courses) with a high impact tutoring class within the school day while control group students took their regular classes. In a significant US-based study, Guryan et al. (2021) performed two large scale randomized controlled trials of high school students in Chicago who received 60-minute tutoring sessions each day of the school week in a 2:1 instructional format for an entire school year. Tutored students in this intensive program made substantial gains in mathematics test scores and course grades compared to control group students though the difference between the treatment and control groups was considerably smaller when the control condition received additional minutes of mathematics instruction as opposed to control samples that included students who took an elective course.

Effective high impact tutoring programs require considerable training, logistical coordination, and financial resources. To understand how to optimize high impact tutoring, there remains a need to investigate different elements of tutoring programs, such as tutor-student group sizes, length of programs, and tutoring models. During high school, few US-based experimental studies explore high impact tutoring models with near-peer tutors of high school students. This gap is significant because existing evidence indicates potential academic benefits from near-peer or peer tutoring for high school students (Colvin, 2007; Duran, 2017; Karsenty, 2010; Roscoe & Chi, 2007). Prior work further suggests that not only is research needed to examine different school contexts and features of high impact tutoring but also study designs are needed to distinguish the effects of high impact tutoring from added instructional time. In most experimental studies of high impact tutoring, control group students participated in a business-as-usual scenario that did not provide instruction in mathematics. As a result, it is difficult to conclude that high impact tutoring as an instructional format is the primary mechanism behind observed academic gains rather than the supplementary instructional time it offers (Kraft, 2020;

Kraft & Novicoff, 2024). If gains are attributable to added instructional time, the use of resources for academic remediation among high school students might be reconsidered.

Remedial Instruction: A Potential Alternative to High Impact Tutoring

There are challenges to bringing in-person high impact tutoring programs to scale. Logistically, recruiting sufficient numbers of high-quality tutors for part-time work and finding tutors to commute long distances to small towns and rural areas can be difficult. Furthermore, virtual models that might extend the reach of high impact tutoring to rural areas have exhibited smaller academic gains compared to in-person models (Kraft et al., 2024). Cost is another barrier. In one influential high-school level analysis, researchers found large achievement effects for a program consisting of one hour of 2:1 tutoring per day for the entire academic year, but the cost of this program was $3,200-4,800 per student each year. This amount is as much as one-third of average annual per-pupil funding ($14,000) in the United States (National Center for Education Statistics, 2022).

Compared to high-impact tutoring, extended instructional time with a classroom teacher is a less resource-intensive approach that may help to accelerate academic growth over the long run. In a well-performed meta-analysis, researchers analyzed 74 causal studies and found that increased instructional time is associated with a range of small to medium effects on student achievement (Kraft & Novicoff, 2024). Other studies have reported that added remedial instructional time in core subjects may be particularly beneficial for low achieving high school students (Angrist et al., 2016; Cattaneo et al., 2017; Cohodes & Parham, 2021; Lester & Naff, 2016; Nomi & Allensworth, 2009). Providing remedial instruction could thus be an approach to raising the mathematics achievement of low-achieving high school students when logistical and cost barriers limit the reach of high impact tutoring. However, remedial instructional courses with a classroom teacher may not be as effective as high impact tutoring. The differential benefits of high impact tutoring and remedial instruction can only be inferred imprecisely by comparing aggregated results of studies (Kraft & Novicoff, 2024). Little work has been done that compares results in the same study samples, which would give more precise estimates of the difference between high impact tutoring and alternative approaches to remediation.

Methods

Design of Treatment and Control Conditions

In this study, we compared the outcomes of ninth-grade students who were randomly assigned to either a remedial mathematics class providing high impact tutoring or to a remedial mathematics class delivered by a classroom teacher only. The primary focus of the high impact tutoring treatment and remedial mathematics control group course was to develop pre-skills for Algebra I (i.e., pre-Algebra). For ninth-grade students, Algebra I is a critical gateway to higher level mathematics classes, and it is required for high school graduation in most states (Chetty et al., 2014; Duncan et al., 2007). Since both groups were enrolled in the same remedial course, treatment and control group students within schools received approximately the same amount of extra instructional time for the academic year with individual class periods typically being 50 to 55 minutes in duration. Remedial class sections were capped at 22 students. In the remedial mathematics control group, the average class size was 19 students over the three-year period of analysis.

The same teacher within schools was usually responsible for leading both treatment and control group sections. Among these teachers, six were novice emergency certified teachers while six other teachers held standard certification in high school mathematics. In the high impact tutoring treatment group, one tutor worked with two/three low-achieving ninth grade students for an entire class period three times a week over the course of the school year. On school days when treatment group students were not tutored, they worked with a classroom teacher. Oklahoma state law requires a certified teacher to oversee all classrooms even if tutors are present. It is important to note that the remedial pre-Algebra classes did not replace students' regular Algebra course that is required for graduation. Students in both the treatment and control groups were double blocked in two mathematics courses (i.e., remedial pre-Algebra and Algebra I) for the school year.

Four faculty and four support staff in the College of Education at the University of Oklahoma created, operated, and evaluated the high impact tutoring program. Three faculty members were mathematics education scholars while program staff comprised undergraduate and graduate research assistants. The program's tutoring director organized tutor recruitment and hiring, tutor training sessions, school site coordination, and data management. Tutor trainers had either high school teaching experience or experience teaching high school mathematics. To support instructional coherence, program faculty created a common pacing guide and instructional materials with the sequence, activities, and curriculum goals for the high impact tutoring treatment group and the remedial mathematics control group. The pacing guide was adapted from Big Ideas – a common core aligned curriculum for middle and high school students (Larson & Boswell, 2019).

Tutors were undergraduate and graduate students from the University of Oklahoma. To be eligible for a position, tutors were required to pass an interview; hold a 3.0 GPA or higher; and have a B grade or higher in a college-level algebra or calculus. As near peers, university students represent a potentially important resource for reaching high school students with tutoring programs as they often have the requisite skills to teach ninth grade mathematics, and many can accept part-time work (Colvin, 2007; Duran, 2017; Kraft et al., 2024; Roscoe & Chi, 2007). The university-led high impact tutoring program offered tutors a compensation package (i.e., $4,700/semester) that was higher than that of regular part-time work that university students tend to do near campus. Program staff worked to monitor tutor performance at sites and respond to performance challenges (e.g., incomplete lesson plans, tardiness, and professional conduct). During the three-year study, three tutors were released mid-semester because of ongoing performance issues while an average of five tutors/semester were not asked to return between semesters because of performance concerns. Graduation, scheduling conflicts, and tutor health issues were the main reasons why most tutors stopped tutoring at the end of a semester or academic year. Table 1 presents the characteristics of tutors. Approximately 87% served at least one academic year (two semesters) while 33% worked as tutors for two academic years (four semesters).

Table 1. Characteristics of tutors

	Mean / Prop. (SD)
Male	0.49 (0.50)
Female	0.51 (0.50)
Graduate	0.07 (0.25)
Undergraduate	0.93 (0.25)
International Student	0.25 (0.43)
Grade Point Average (4.0)	3.7 (0.31)
Semesters tutored	2.92 (1.42)
Tutors Serving Two Semesters	0.87 (0.33)
Tutors Serving Four Semesters	0.33 (0.47)
Major/Minor in Education	0.04 (0.20)

Total Tutors	502

Note. We collected data on the race/ethnicity of tutors in Year 3. Tutors were 41% White, 17% Black, 15% Asian, and 27% other background or multi-racial.

We developed our high impact tutoring program based on the following pillars.

Reflective Instruction and Personalized Learning. The program's instructional strategies were grounded in the principles of mathematical proficiency – conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition – as outlined by the National Research Council and Mathematics Learning Study Committee (2001) and National Council of Teachers of Mathematics (2014). To develop these five strands, tutors used various mathematical communication strategies, such as describing one's thinking and peer-peer teaching. They also utilized models and manipulatives (e.g., fraction strips, base-10 blocks, or algebra tiles) to strengthen students' conceptual knowledge during tutoring sessions. These instructional approaches were part of a common pacing guide and integrated into tutors' lessons to help students connect meaning, understanding, and application when learning (National Council of Teachers of Mathematics, 2014; Rittle-Johnson et al., 2017). To ensure that individual learning needs could be met, lesson plans allocated time in each lesson for tutors to respond to specific needs of their students and build skills that they were ready to learn.

Continuous Training and Monitoring. Prior to the academic year in early August, tutors attended tutoring "boot camp" for four full days. These sessions were followed by weekly 90-minute training and lesson planning sessions every Monday. During weekly training sessions, program staff trained tutors to build (pre)skills needed for grade level mathematics (i.e., Algebra I). Training sessions also emphasized strategies for increasing high school students' self-concept, differentiating instruction, and engaging students through instructional games (e.g., range-game, quick-look, notice and wonder). Drawing from a common pacing guide, each week tutors submitted and received feedback on their lesson plans for the upcoming week. Program staff performed observations and monitoring of tutors at school sites, followed by formative feedback on training days.

Relationship-building and Mentoring. Program staff trained tutors to build relationships with high school students based on trust and psychological safety. Training sessions emphasized creating supportive learning environments where tutored students would feel comfortable asking questions and taking risks. Training sessions contained modules to foster cultural competence, mentoring, and positive tutor-student interactions. These modules aimed to help tutors cultivate asset-based mindsets by having tutors reflect on the backgrounds of tutees at school sites. Because low-achieving ninth-grade students begin high school with varied individual learning needs, each tutor committed to working with the same ninth-grade students for the semester while program staff worked with tutors to tailor their instruction to match students' knowledge and pace of learning. *School Site Support and Coordination*. Tutors provided instruction three times a week while classroom teachers did so on the other two days of the week. To assist with coordinating instruction, program staff did site visits and coordination sessions with teachers.

Sample and Study Setting

Three separate cohorts of ninth-grade students at seven high schools participated in this study from 2021-2024. The research team recruited public and charter high schools that were within a 30 to 45-minute drive from the University of Oklahoma's campus, which is in central Oklahoma and approximately 20 miles from the state's largest metropolitan area – Oklahoma City. This recruitment decision was made to minimize travel costs and give tutors time to return to campus for their college classes. Within the radius of potential school sites, seven of the ten school districts that we contacted agreed to participate in the study. Three school districts declined our invitation to participate in the study because they were unable to agree to a randomized controlled trial design. There were no direct costs for schools to participate in the study, but schools had to allocate classroom space for the program and assign a classroom teacher to both treatment and control group class sections.

Table 2 presents the characteristics of each school site. Two schools are rural and five are midsize or large urban schools. The seven high schools are characterized by a high degree of racial/ethnic diversity. One high school is a charter school while the other high schools are district-run public schools. Nearly every school has a high percentage of students eligible for free-and reduced priced lunch with two schools having more than 90% of their students being eligible for free-and reduced priced lunch. Oklahoma is a low achieving state that ranks among the poorest performing states on the National Assessment for Educational Progress (NAEP, 2024). Still, three of the schools in the study performed below the state average for academic proficiency and all study participants were below the state average at baseline.

Table 2. Study sites – Participating schools

	School 1	School 2	School 3	School 4	School 5	School 6	School 7
School Type	District-run public	District-run public	District-run public	District-run public	District-run public	Public charter	District-run public
Grades	9-12	9-12	9-12	9-12	9-12	9-12	9-12
Academic Proficiency (%)[1]	10-20	50-60	40-50	60-70	70-80	50-60	0-10
Free & Reduced Price Lunch (%)	60-70	60-70	30-40	50-60	40-50	90-100	90-100
Native American (%)	0-10	0-10	0-10	0-10	0-10	0-10	0-10
Latino/a (%)	0-10	10-20	10-20	20-30	10-20	90-100	70-80
White (%)	0-10	60-70	60-70	40-50	50-60	0-10	0-10
Black (%)	80-90	0-10	0-10	0-10	0-10	0-10	1-10
Other (%)	0-10	10-20	10-20	10-20	20-30	0-10	0-10
Locale	Large urban	Rural	Rural	Midsize urban	Midsize urban	Large urban	Large urban

Note. All schools were located in Oklahoma. These descriptive school-level data come from the Oklahoma State Accountability System (Oklahoma Educational Quality and Accountability, 2024). We provide a range for school information to avoid identifying individual school sites. 1. Academic Proficiency is a composite measure of the percentage of students who score basic, proficient, or advanced on state tests ELA, math, and science. The state average on this index is 49%.

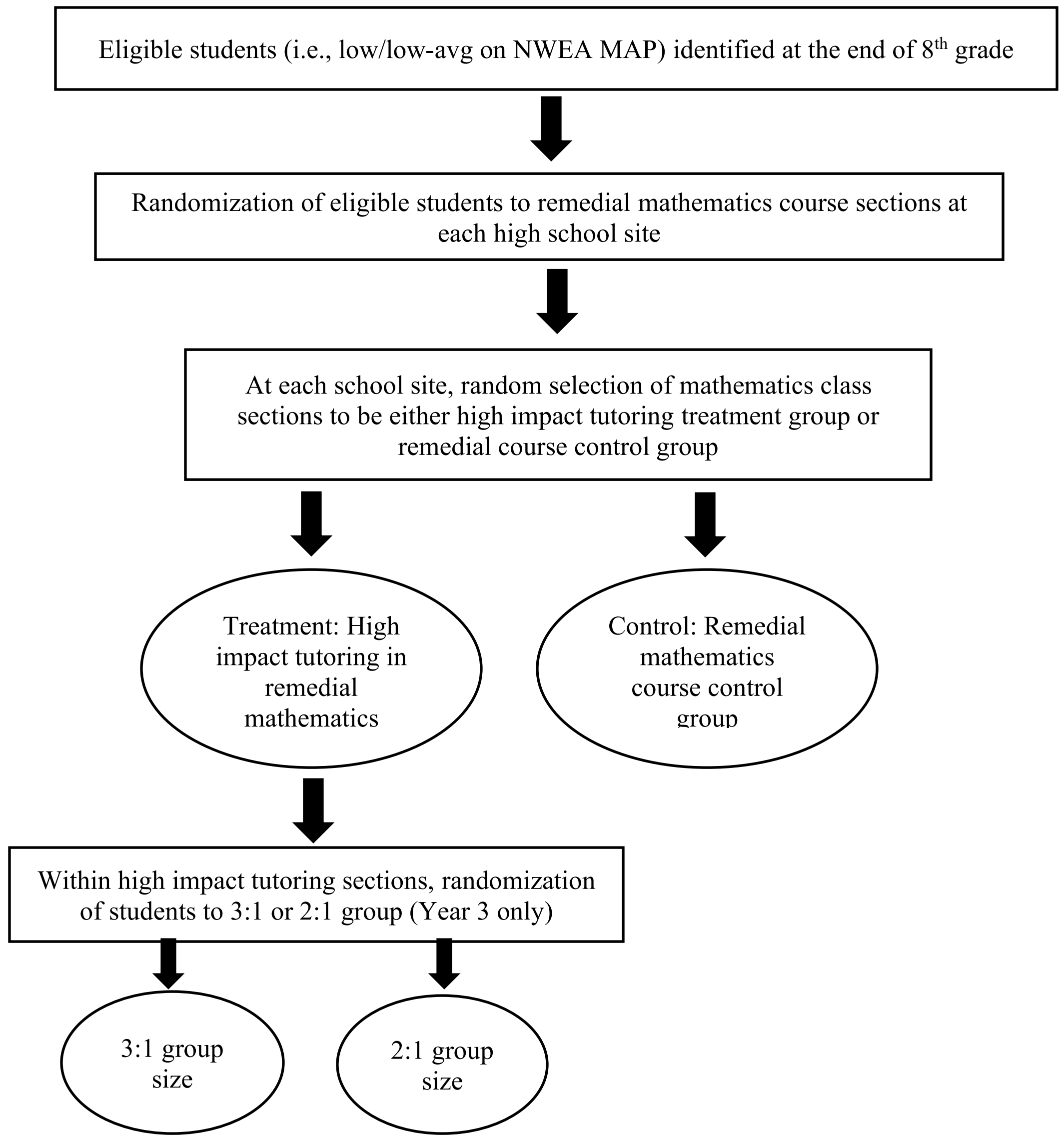
Our initiative was funded to increase tutoring capacity each year. In Year 1, 223 ninth-grade students were randomized to treatment (i.e., 2:1/3:1 high impact tutoring) and control groups (i.e., remedial mathematics course) at one high school where there were three treatment class sections and seven remedial mathematics control group sections. In Year 2, we expanded the initiative to five schools where 370 students were randomized to treatment and control conditions with ten high impact tutoring class sections and nine remedial mathematics class sections. In Year 3, 520 were randomized to treatment and control conditions at seven high schools where there was a total of nineteen high impact tutoring class sections and eleven remedial mathematics class sections. One small school site was unable to establish a control group in Year 3, so robustness checks were performed excluding this school from the main analyses. Results were consistent with those of the main models (see Table 2A).

Attrition analyses showed no statistical differences in attrition between the high impact tutoring treatment and remedial mathematics control group in the pooled sample and within years (see Appendix Table 1A). In the three-year pooled sample, attrition was 7% in the treatment group and 6% in the control group. For each year, attrition was 4% in the treatment group and 6% in the control group in Year 1; 5% in the treatment group and 6% in the control group in Year 2; and 10% in the treatment group and 6% in the control group in Year 3. These attrition rates for treatment and control groups indicate a moderate level of potential bias from differential attrition (Moffatt, 2020). Attrition is related to certain demographic characteristics and not missing at random as students eligible for free and reduced-price lunch were more likely to be lost to attrition than those who were not eligible for free and reduced-price lunch (OR = 2.55, SE = 0.74, $p < 0.01$).

At school sites, we sought to strengthen study participation by holding information nights with families and by communicating routinely with teachers, families, and students about the program. We also use a multi-stage passive consent process that may have supported high study participation rates. Eight students opted out of the study in Year 2 while four students opted out of the analysis in Year 3.

Experimental Design and Randomization Procedure

Figure 1. Randomization Procedures



To estimate causal effects, we used a two-stage clustered randomization procedure with the high impact tutoring treatment being administered at the classroom level within high schools (see Figure 1). To identify potential study participants, we tested students at the end of 8th grade, designating eligible students as "low" or "low-average" in mathematics on the NWEA MAP assessment (Thum & Kuhfeld, 2020). At each school site, cut scores varied to some extent based on school size and tutor availability. Average baseline achievement was at the 25th percentile with school averages ranging between the 17th percentile and 30th percentiles in six of the seven schools. One small school that entered the study in Year 3 started at the 39th percentile, which was a higher baseline average than the other schools in the study sample. With this school excluded, Table 2A in the Appendix shows results that are consistent with the main results.

In late spring, Infinite Campus software was used to randomize eligible students to remedial mathematics class sections at each high school for the coming academic year. These remedial class sections were then randomly selected to be either a high impact tutoring treatment group or to be a remedial mathematics class control group taught solely by a classroom teacher. All study participants were also enrolled in an Algebra I class required for graduation, so study participants took two mathematics courses for the entire school year. The study design is strengthened by having eligible students first randomly assigned to remedial course sections, and subsequently, having course sections randomly assigned to either high impact tutoring or to a remedial mathematics class taught solely by a classroom teacher. In Year 3 of the study, we included an additional treatment condition within the high impact tutoring course sections by randomly assigning students to either 2:1 or 3:1 student-tutor group sizes.

At the start of the academic year, all ninth-grade students took the NWEA MAP assessment, providing us with baseline mathematics achievement data. This assessment was then administered at the end of the academic year to determine the effects of high impact tutoring on students' mathematics test scores. Our program staff trained teachers to administer the NWEA MAP assessment and visited sites on testing days to assist with administering and monitoring the NWEA MAP assessment. Table 3 presents summary statistics for each variable of analysis for the control group, treatment group, and non-study student samples (i.e., students who tested out of the study in eighth grade). Non-study students have much higher baseline mathematics scores, were more likely to be White students, and were less likely to be free/reduced priced lunch status and English language learners.

Table 4 compares the baseline characteristics of the treatment and control groups for each of the three years of analysis and in the pooled sample. In the pooled sample, control group students are more likely to be Latino/a ($p < .001$) and English language learners ($p < .001$). Treatment group students are more likely to be White ($p < .001$). Based on within-year comparisons, baseline differences observed in the pooled sample seem to be driven by differences that occur in Year 3 of the study sample.

Table 3. Summary Statistics

	Control Group Sample				Treatment Group Sample				Non-study 9th Grade Students			
Variable	Mean/ Prop.	SD	Min.	Max.	Mean/ Prop.	SD	Min.	Max.	Mean/ Prop.	SD	Min.	Max
Fall Baseline Math RIT Score	213.31	10.77	173	243	213.12	11.44	167	251	220.79	15.05	150	263
Winter Math RIT Score	215.84	11.49	176	245	216.64	12.03	173	268	221.85	15.18	159	264
EOY Math RIT Score	218.53	12.65	176	251	219.39	12.75	180	282	224.98	15.72	169	290
Fall-Winter Growth	2.53	7.60	-32	33	3.66	8.02	-35	48	1.26	7.53	-24	57
Fall-Spring Growth	5.22	8.89	-27	49	6.27	9.22	-24	85	4.19	8.97	-44	74
EOY GPA	2.52	0.93	0	4.03	2.56	0.94	0	4.08	2.78	1.00	0	4.08
Female	0.51	0.50	0	1	0.52	0.50	0	1	0.49	0.50	0	1
Free/Reduced Price Lunch	0.80	0.40	0	1	0.76	0.43	0	1	0.63	0.48	0	1
Latino/a	0.63	0.48	0	1	0.46	0.50	0	1	0.35	0.48	0	1
Black	0.05	0.21	0	1	0.08	0.27	0	1	0.08	0.28	0	1
White	0.17	0.38	0	1	0.30	0.46	0	1	0.37	0.48	0	1
Native	0.06	0.24	0	1	0.07	0.26	0	1	0.06	0.24	0	1
Other Race	0.09	0.29	0	1	0.09	0.29	0	1	0.13	0.34	0	1
English language learner	0.41	0.49	0	1	0.27	0.45	0	1	0.18	0.38	0	1
Special Education	0.15	0.35	0	1	0.22	0.41	0	1	0.13	0.33	0	1
Students	438				525				1,774			

Note. Non-study students were generally higher performing students who tested out of the study when they took the NWEA MAP assessment at the end of eighth grade.

Table 4. Comparison of the baseline characteristics of the treatment and control groups

	Pooled Control	Treatment	Year 1 Control	Treatment	Year 2 Control	Treatment	Year 3 Control	Treatment
Variable	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)	Mean/Prop. (SE)
Fall Baseline	213.31	213.12	212.75	214.13	214.21	212.14	212.92	213.49
	(0.51)	(0.49)	(0.95)	(1.56)	(0.89)	(0.84)	(0.84)	(0.67)
Female	0.51	0.52	0.47	0.50	0.52	0.55	0.53	0.50
	(0.02)	(0.02)	(0.04)	(0.06)	(0.04	(0.04)	(0.04)	(0.03)
FRL	0.80	0.76	0.95	0.96	0.69	0.68	0.79	0.75
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Latino/a	0.63***	0.46	0.86	0.93	0.50	0.41	0.55***	0.37
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
Black	0.05	0.08*	--	--	0.03	0.07	0.10	0.10
	(0.01)	(0.01)			(0.01)	(0.02)	(0.02)	(0.02)
White	0.17	0.30***	--	--	0.31	0.31	0.18	0.37***
	(0.02)	(0.02)			(0.04)	(0.03)	(0.03)	(0.03)
Native	0.06	0.07	--	--	0.08	0.06	0.10	0.10
	(0.01)	(0.01)			(0.02)	(0.02)	(0.02)	(0.02)
Other Race	0.09	0.09	0.14	0.07	0.07	0.14	0.06	0.07
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
ELL	0.41***	0.27	0.50	0.47	0.35*	0.24	0.38**	0.24
	(0.02)	(0.02)	(0.04)	(0.06)	(0.04)	(0.03)	(0.04)	(0.03)
Special Education	0.15	0.22	0.17	0.18	0.17	0.24	0.11	0.21**
	(0.02)	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)
Students	438	525	133	68	150	177	155	280

Math RIT score and Student Grade Point Average

The Measures of Academic Progress (MAP) assessment was our primary outcome of interest. Developed by the Northwest Evaluation Association (NWEA), the MAP is a computer adaptive test that is used to evaluate mathematics ability and growth throughout the school year. Over 9,700 schools use the MAP assessment in 145 countries and NWEA has a 40-year history of developing such assessments. In this study, the MAP was administered three times each year (i.e., early fall, winter, and late spring). In ninth grade, the MAP is designed to measure students' growth and achievement in the areas of algebraic thinking, numbers and operations, measurement and data, and geometry (NWEA, 2019). MAP results are reported using Rasch Units (RIT) that range from 100 to 350 points. The reliability evidence (test-retest reliability and marginal reliability) for the MAP is strong. For the ninth-grade mathematics assessment, test-retest reliability is greater than 0.9; marginal reliability is 0.96; and (NWEA, 2019). External alignment studies also offer evidence to support the validity of the MAP assessment (Egan et al., 2017). These studies indicate that 97% of MAP items are aligned to Common Core State Standards (ninth grade mathematics: $r = 0.72$). In addition to analyzing the NWEA MAP assessment, we investigated students' grade point average at the end of year (scale 0 to 4.0). We had originally collected data on student tardiness and absences but learned from our program staff that these data were often inconsistently recorded at school sites, so we excluded these indicators.

Administrative Data

Data were collected on the following student background characteristics: gender, free-and-reduced priced lunch status, special education status, English language learner status, and race/ethnic background. Tutors submitted weekly logs that recorded the total amount of time they spent tutoring each high school student during each of the three sessions for the week. Tutored students could receive approximately 4,100 minutes of instruction from a tutor during the academic year. Table 5 presents a breakdown of tutoring minutes for students in the high impact tutoring treatment group. Approximately 50% of students who remained in the study for the ninth-grade school year received 3,000 or more minutes of tutoring. Variation in tutoring minutes was due to individual student absences, suspensions, and school closures (e.g., ice/snow days).

Table 5. Minutes tutored in high impact tutoring treatment group

Range of Minutes Tutored	Percentage of Tutored Students
1-1000	3
1001-2000	7
2000-3000	41
3000-3500	39
3500-4000[a]	11

a. Approximately 0.5% of students received slightly more than 4,000 minutes of tutoring. Minutes tutored calculations are for 508 ninth-grade students who participated in tutoring for the entire school year.

Data Analytic Strategy

To estimate the effects of high impact tutoring, we analyzed data from a pooled sample of three separate annual cohorts of ninth-grade students from 2021-2024. We also estimated results separately for each year of the study. To estimate ITT (Intent-to-Treat) and TOT (Treatment-on-the-Treated) effects, regression models were performed that account for student-level clustering by school site and cohort year (Moffatt, 2020). For TOT effects, compliance with the treatment indicated whether the student was enrolled in high impact tutoring section for the academic year while control group students had to be enrolled in the remedial pre-Algebra class all year and must not have received our program's tutoring within the school day. For the remedial mathematics class control group, class rosters were checked at the start, midpoint, and end of the academic year. Weekly tutor logs allowed us to monitor participation in the treatment group during the academic year. The ITT sample comprised students who were randomly assigned to the high impact tutoring treatment or the remedial mathematics course but left their assigned section, typically for a different elective class within the same school. Students were not permitted to transfer between the treatment and control group sections.

For the main analyses, we estimate the following regression model in Stata, clustering standard errors at the school level:

$$Outcome_{ij}=\beta_0+\beta_1 High\_Impact\_Tutoring_{ij} +\beta 2 Baseline\_MathScore_{ij}+\beta 3 Sex_{ij}+\beta 4 FRL_{ij}+\beta 5 Latino/a_{ij}+\beta 6 Black_{ij}+\beta 7 Native_{ij}+\beta 8 Other\_race_{ij}+\beta 9 ELL_{ij}+\beta 10 IEP_{ij}+\gamma_j + \delta_t+\epsilon_{ij}$$

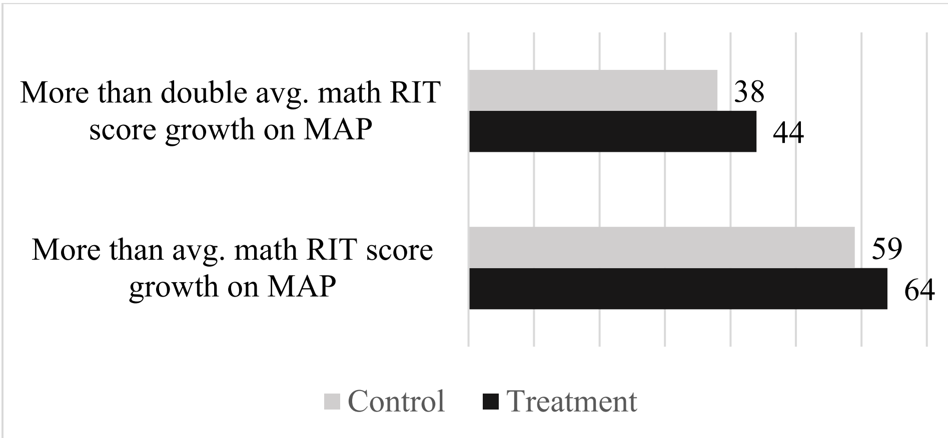
In the model above, $Outcome_{ij}$ is the academic performance (i.e., math RIT score and grade point average) of student *i* in school *j*. $High\_Impact\_Tutoring_{ij}$ is a binary variable indicating whether a student received high-impact tutoring. $Baseline\_MathScore_{ij}$ represents students' baseline math RIT score at the start of ninth grade on the NWEA MAP assessment. FRL_{ij} is a binary variable indicating eligibility for free or reduced-price lunch (FRL). Race/ethnicity indicators are Latino/a, Black, Native American, and Other_race (i.e., bi-racial/multi-racial backgrounds) with White students serving as the reference category. IEP_{ij} denotes whether the student has an Individualized Education Plan, and ELL_{ij} indicates whether the student is an English Language Learner (ELL). The term γj is the school fixed effect, and δt is the cohort year fixed effect accounting for year-specific influences. ϵij is the error term, clustered at the school level to account for the non-independence of students within the same school. The model estimates robust standard errors. We do not account for clustering by class section because the same teacher taught both treatment and control sections at five of the seven school sites. Instead, fixed effects for each school site are included to account for school level differences. We estimate effects separately for each year of the study, removing the fixed effect for cohort year. Statistical models are also used to test for heterogeneous associations based on the number of tutoring hours a student received during the school year. In subsequent models, we performed moderator analyses to examine differential relationships by student and school background characteristics. In Year 3 of the study, students were randomly assigned to 3:1 and 2:1 tutoring group size, enabling us to test for potential differences based on group sizes (2:1 and 3:1).

Results

Descriptive Patterns

In the pooled sample, students in the treatment group grew 6.27 RIT points while students in the control group grew 5.21 RIT points on the NWEA MAP assessment during the academic year. To put this growth into context, average projected growth for ninth-grade students on this assessment is 3.6 RIT points (Thum & Kuhfeld, 2020). However, since these growth rates are high compared to national norms, they are purely descriptive and do not represent causal effects. In Figure 2, students in the high impact tutoring treatment group showed more substantial gains than control group students in the remedial mathematics class. Sixty-four percent of treated students grew more than their expected annual growth in mathematics while 59% of students in the control gained more than expected annual growth. Additionally, 44% of treated students grew more than double expected annual growth, compared to 38% of control group students doing so.

Figure 2. Math RIT score growth on the NWEA MAP Assessment (%)



Note. Average growth in mathematics on the NWEA MAP assessment in ninth grade is 3.6 points.

Treatment Effects: Three-year Pooled Sample

Table 6 presents Intent-to-Treat (ITT) and Treatment Effect on the Treated (TOT) results for students' end-of-year math achievement and grade-point averages. The ITT effect on math achievement is 1.62 RIT points ($p < .05$) while the TOT effect is 1.59 points ($p < .05$) on the NWEA MAP assessment. Since average expected growth in ninth grade is 3.6 points on the NWEA MAP assessment, these differences amount to nearly a half-year of learning in mathematics or a difference of 0.13 SD between the high impact tutoring treatment and remedial mathematics class control group. The similarity between the ITT and TOT effects could be due to the overall high study participation rate and low attrition in the treatment and control groups in this study. In these models, we observe no statistical differences in grade-point averages between the treatment and control groups.

Table 6. Estimated effects on mathematics achievement and grade-point average

	Intent-to-Treat (ITT)		Treatment Effect (TOT)	
Variables	EOY Math RIT Score	EOY GPA	EOY Math RIT Score	EOY GPA
High Impact Tutoring Treatment	1.62*	0.08	1.59*	0.08
	(0.78)	(0.08)	(0.80)	(0.07)
Baseline Math RIT Score	0.78***	0.03***	0.79***	0.03***
	(0.03)	(0.00)	(0.03)	(0.00)
Demographic Factors	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y
Cohort Year Fixed Effects	Y	Y	Y	Y
Control Group Mean	218.11***	2.50***	217.98***	2.49***
	(0.43)	(0.04)	(0.43)	(0.04)
Students	963	963	936	936
Schools	7	7	7	7

Robust Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05
Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status,

and IEP status. School fixed effects are binary indicators for schools in the analysis. Cohort year fixed effects represent binary indicators for the years of analysis.

Table 7 presents TOT results based on the number of hours tutored. In Model 1, there is a marginal positive effect for math RIT score but no difference for grade-point average. In Models 3 and 4, a quadratic term is introduced to test for potential non-linear effects of tutoring time and academic outcomes, but there is no statistical evidence of a curvilinear relationship between tutoring time received and mathematics RIT score while grade-point average. In Appendix Table 3A, we estimate effects by 2,000, 2,600, and 3,000 minutes tutored. Estimates also show little/no difference in effects based on these varying levels of tutoring time.

Table 7. Effect of high impact tutoring on treated students (9th grade) by hours tutored

Variables	EOY Math RIT Score	EOY GPA	EOY Math RIT Score	EOY GPA
High Impact Tutoring Treatment (Hours)	0.03*	0.00	0.08	-0.01*
	(0.02)	(0.00)	(0.07)	(0.01)
High impact treatment^2 (Hours)			-0.00	0.00
			(0.00)	(0.00)
Baseline Math RIT Score	0.79***	0.03***	0.79***	0.03***
	(0.03)	(0.00)	(0.03)	(0.00)
Demographic Factors	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y
Cohort Year Fixed Effects	Y	Y	Y	Y
Control Group Mean	218.34***	2.51***	218.34***	2.51***
	(0.61)	(0.05)	(0.61)	(0.05)
Students	936	936	936	936
Schools	7	7	7	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Note. Models 3 and 4 include a quadratic term for hours tutored. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools. Cohort year fixed effects are binary indicators for the years of analysis.

Effects by Year, Tutor-student Group Size, and Student Subgroup

Table 8 presents ITT results by year. Results demonstrate that test score gains for the high impact tutoring treatment group are concentrated in Year 3. In Year 3, there is a 2.56 RIT score ($p < 0.10$) difference between the treatment and control groups. By contrast, there is a smaller advantage for tutored students in Year 1 and Year 2. Consistent with analyses of the pooled sample, no differences are observed between the treatment and control groups for end-of-year grade point averages.

Table 8. Estimated effects (ITT) on mathematics achievement and grade point average by year

Variables	EOY Math Score (Yr. 1)	EOY Math Score (Yr. 2)	EOY Math Score (Yr. 3)	EOY GPA (Yr. 1)	EOY GPA (Yr. 2)	EOY GPA (Yr. 3)
High Impact Tutoring Treatment	0.71	1.09**	2.56†	0.08	0.11	0.06
	(1.13)	(0.39)	(1.48)	(0.11)	(0.13)	(0.05)
Baseline Math RIT Score	0.93***	0.72***	0.72***	0.04***	0.03***	0.03***
	(0.05)	(0.04)	(0.05)	(0.01)	(0.01)	(0.00)
Demographic Factors	Y	Y	Y	Y	Y	Y
School Fixed Effects	N	Y	Y	N	Y	Y
Control Group Mean	220.52***	218.60***	216.39***	2.71***	2.45***	2.43***
	(0.69)	(0.21)	(0.96)	(0.07)	(0.07)	(0.03)
Students	201	327	435	201	327	435
Schools	1	5	7	1	5	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, † < 0.10
Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools.

Table 9 presents TOT results, which demonstrate similar patterns. In Year 3, there is a difference of 2.53 RIT points ($p < 0.10$) between the treatment and control groups, which amounts to about three-quarters of a year of learning in mathematics for ninth grade students. There are no statistical differences for end-of-year grade point averages.

Table 9. Estimated effects (TOT) on mathematics achievement and grade-point average by year

Variables	EOY Math Score (Yr. 1)	EOY Math Score (Yr. 2)	EOY Math Score (Yr. 3)	EOY GPA (Yr. 1)	EOY GPA (Yr. 2)	EOY GPA (Yr. 3)
High impact tutoring Treatment	0.79	1.00*	2.53†	0.08	0.09	0.07
	(1.13)	(0.44)	(1.49)	(0.11)	(0.13)	(0.05)
Baseline RIT Math Score	0.92***	0.75***	0.72***	0.04***	0.03***	0.03***
	(0.05)	(0.06)	(0.05)	(0.01)	(0.01)	(0.00)
Demographic Factors	Y	Y	Y	Y	Y	Y
School Fixed Effects	N	Y	Y	N	Y	Y
Control Group Mean	220.29***	218.33***	216.41***	2.70***	2.46***	2.43***
	(0.69)	(0.26)	(0.96)	(0.07)	(0.07)	(0.03)
Students	199	303	434	199	303	434
Schools	1	5	7	1	5	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, † < 0.10

Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools in the sample.

In Year 3, students assigned to high impact tutoring were randomly assigned to either a 2:1 or 3:1 student-tutor group within their high impact tutoring class sections. In Table 10, we find no evidence that 2:1 student-tutor groups outperform 3:1 groups. By performing a Wald-test, we find a statistically significant positive effect for 3:1 over 2:1 student-tutor groups ($p <$ 0.05), though the 3:1 sample is relatively small.

Table 10. Effects by 3:1 and 2:1 student-tutor ratios

Variables	EOY Math Score	EOY GPA
High Impact Tutoring Treatment (2:1)	1.10	0.05
	(0.67)	(0.09)
High Impact Tutoring Treatment (3:1)	4.08**	0.23*
	(1.27)	(0.09)
Fall Baseline Math RIT Score	0.79***	0.03***
	(0.03)	(0.00)
Demographic Factors	Y	Y
School Fixed Effects	Y	Y
Cohort Year Fixed Effects	Y	Y
Mean for 3:1 High Impact Tutoring Treatment	222.62***	2.75***
	(1.13)	(0.08)
Mean for 2:1 High Impact Tutoring Treatment	219.63***	2.57***
	(0.38)	(0.05)
Students (3:1)	108	108
Students (2:1)	417	417
Schools	7	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools. Cohort year fixed effects are binary indicators for the years of analysis.

In Appendix Table 4A, analyses of ITT effects across student subgroups are presented for students eligible for free and reduced priced lunch, Latino/a students, Black students, English language learners, students with needs, students achieving at the 15th percentile and below, charter school students, and students in rural schools. The interactions between these different student subgroups and the high impact tutoring treatment are not statistically significant except in the case of students who are eligible for reduced priced lunch who show a statistically significant interaction of 1.85 RIT score points ($p < 0.05$).

Cost Effectiveness

High impact tutoring programs require substantial resources and coordination (Guryan et al., 2021). Cost effectiveness analysis offers a way of estimating how the overall effect (0.13 SD) of this study's university-led high impact tutoring program compares with its costs. In Appendix Table 5A, the per-student cost of high impact tutoring was $5,326 with 84% of total costs arising from tutor compensation. The remedial mathematics course was only $467 per student, so the additional per-student cost of high impact tutoring to achieve a 0.15 SD effect on mathematics RIT scores was $4,859. Based on effect sizes from over 700 randomized controlled trials, this study's high impact tutoring program yields a medium effect size but has a high cost per student (see Kraft, 2020). However, it must be noted that if we had delivered our model using only 3:1 tutor-student group sizes, the additional cost of the program would decline substantially to $3,495 per tutored student.

Discussion

A robust evidence base suggests that high impact tutoring programs are effective (Kraft et al., 2024; Nickow et al., 2020). Yet, few US-based studies have investigated outcomes at the high school level, and even less research has isolated the effects of high impact tutoring from the added instructional time that it provides. Addressing these gaps is significant from a policy standpoint because there can be considerable financial and logistical barriers to bringing high impact tutoring programs to scale (Kraft & Falken, 2021). In this study, we advance the literature by randomly assigning low-achieving ninth-grade students to either a high impact tutoring treatment or to a remedial mathematics course delivered solely by a classroom teacher. Pooled results from three separate ninth-grade cohorts showed that students in the high impact tutoring group gained approximately a half-year of additional learning (0.13 SD) over students in the remedial mathematics control group. Evidence also indicated that students in 2:1 student-tutor groups did not outperform students in 3:1 student-tutor groups, suggesting that 3:1 student-tutor ratios might be used for high impact tutoring during high school with no detrimental effects on academic performance.

This study offers key contributions to the literature (Kraft & Falken, 2021; Robinson et al., 2021). In previous randomized controlled trials, high impact tutoring at the high school level has exhibited larger positive effects than those estimated in this study (de Rhee et al., 2021; Guryan et al., 2021). Although this difference could be driven by varying factors, one plausible contributing factor is that the control group in our study was assigned to receive the same amount of additional instructional time in mathematics as the treatment group, whereas other studies tend to use business-as-usual control conditions (e.g., elective courses) that do not provide added instruction time in the subject area. As an example, de Ree et al. (2021) reported large effects on mathematics achievement (0.44-0.72 SD) for a small sample of 49 high school students participating in high impact tutoring, but the control condition in the study did not receive additional minutes of mathematics instruction. In a larger analysis, Guryan et al. (2021) found large positive effects on mathematics achievement when tutored high school students were compared to control group students enrolled in an elective course, but effects were smaller when tutored students were compared to control group students enrolled in a remedial mathematics course.

This study's results also cohere with findings demonstrating that large high impact tutoring programs tend to generate smaller effects. As a large program, our pooled effect (0.13 SD) is similar in magnitude to effects (0.16 SD) presented in recent meta-analytic work for other large high impact tutoring programs (Kraft et al., 2024). Our study also contributes to existing literature by testing for differences between 2:1 and 3:1 student-tutor group sizes. In prior work, there are only marginal differences in student outcomes among 1:1, 2:1, and 3:1 high impact tutoring models (Kraft et al., 2024), but very little work has examined such differences among high school students who potentially have a greater capacity to work independently within tutor groups than elementary and middle school students do (Nickow et al., 2020). In this study, we found no evidence that raising the student-tutor ratio from 2:1 to 3:1 had detrimental effects on student achievement in our sample of ninth-grade students. Because tutor compensation is the greatest expense for many tutoring programs, increasing student-tutor group sizes by even one student can reduce program costs considerably. Along with these findings, we did not find strong statistical differences across student subgroups in the sample except in the case of students eligible for free and reduced priced lunch who appeared to benefit more from high impact tutoring more than their peers.

In contrast to our findings for test scores, the analyses in this study exhibited virtually no differences in grade point average between treatment and control groups even though other studies have reported positive effects of high impact tutoring on grade-point average at the high school level (Guryan et al., 2021). It is possible that students' overall grade point averages may be less sensitive to a mathematics intervention in a single course. Furthermore, grade point average may be a less objective measure of academic growth, being influenced by variability in grading standards, subjectivity in assessments, school norms, and other non-academic factors (Randall & Engelhard, 2010). Since we observe differences on an independent assessment but not for grade point average might arguably increase confidence in our overall findings.

Nonetheless, there are limitations to this study that must be highlighted. The multi-step randomization procedure that we used strengthens internal validity, but this study may have limited external validity with our findings coming from seven high schools (urban, suburban, and rural) in Oklahoma. Oklahoma ranks in the bottom quintile of states for its academic performance (NAEP, 2024). While the sociodemographic profiles of our schools vary, these schools may be shaped by the context for public education in Oklahoma in ways that make our results less generalizable to high schools in other states. For example, each high school in the study has experienced significant challenges recruiting and retaining teachers. Nearly half of the teachers overseeing the treatment and control group classes in this study were novice emergency certified teachers. Schools in states with stronger teacher labor markets could find that high impact tutoring programs are less effective than remedial mathematics courses with experienced teachers are.

Another important limitation to this study is that our research design does not allow us to make comparisons to students who received no remedial support because we did not have a business-as-usual control group. The unadjusted academic growth for both the high impact tutoring and remedial mathematics group was strong, but our study design only allows us to compare these two groups against each other. When comparing the treatment and control groups, the strongest treatment effects occurred in Year 3. Funding for our model was allocated to expand the reach of tutoring each year, but this expansion reduced the number of control group students relative to those who participated in high impact tutoring at school sites by Year 3. Although the precise mechanisms behind the Year 3 results are unclear, there may have been a stigma associated with participating in the relatively smaller group that could have negatively influenced performance in the control group in that year. Alternatively, Year 3 results could be a result of the program's administrative staff and trainers improving coordination, training lessons, and tutor recruitment strategies over time. If this latter scenario were the case, new high impact tutoring programs may need to launch with an improvement mindset, understanding that there will be a learning curve and ongoing refinements to the program.

Among other study design features, we used varying cut scores for study eligibility at school sites because of capacity and school size differences. While all cut scores required eligible students to be classified as “low” or “low-average” on the MAP assessment, slight differences in cut scores could be a limitation to this study. We also relied on a strong independent measure (i.e., MAP assessment) to evaluate program effects, but we did not employ researcher-designed assessments for comparison as there were concerns about over-testing students.

Spillover effects are also a potential limitation. The classroom teacher for the high impact tutoring treatment group class section was usually the same teacher for the remedial mathematics control group class section. Teachers used the same pacing guide for both treatment and control groups. They also supervised their classrooms on tutoring days, so the pedagogical and instructional strategies used for high impact tutoring sessions may have positively affected how teachers (many of whom were novices) ultimately instructed students in their control group sections. Finally, there was no statistical evidence of differential attrition between treatment and control groups, and study participation rates were high in this study compared to other prominent analyses of high impact tutoring at the high school level (Guryan et al. 2012). Yet, overall attrition after initial assignment ranging from 10% to 16% across the three years of analysis, which could compromise internal validity.

Implications for Policy and Practice

As states and localities have been working to expand access to high impact tutoring programs, this study’s findings have policy and practice implications (Jordan et al., 2022). The results indicate that high impact tutoring can be an effective tool for raising academic growth during high school. Yet complex logistical, financial, and administrative factors must coalesce to make effective high impact tutoring available in schools. In these programs, tutor recruitment, training, and compensation barriers can limit the reach of high quality in-person tutoring within the school day (Robinson et al., 2021). University-led high impact tutoring programs represent a promising way to overcome expansion obstacles of this nature, primarily by recruiting university students to serve as tutors. University students often accept part-time employment, possess foundational high school mathematics skills, and are near peers who can become positive role models for academically struggling high school students (Colvin, 2007); Duran, 2017; Roscoe & Chi, 2007). As a caution, however, university students have less formal training and familiarity with school procedures than paraprofessionals and teachers who commonly serve as tutors for high impact tutoring programs. It is vital that university faculty and staff leading tutoring programs provide ongoing training, supervision, and school site coordination to ensure that university students are effective tutors.

In efforts to expand high impact tutoring, distance is a barrier that university programs may also be capable of overcoming. Recruiting tutors to travel beyond a 30–40-minute radius to rural areas can be difficult for companies providing tutoring services because of time, gas expenses, and other transportation costs. Universities located outside of major urban centers might help to reach schools that otherwise do not have access to in-person high impact tutoring in their areas. State and regional universities might also coordinate efforts and resources to provide tutor training, instructional materials, and communication with schools. Combining expertise and resources among higher education institutions within states could ease capacity challenges faced by small regional universities when launching new high impact tutoring programs. There are also potential benefits to university students themselves who can obtain reasonable part-time compensation and quality work experience when serving as tutors.

Program design decisions can influence the scope of high impact tutoring. Among them, higher student-tutor ratios can expand programs by decreasing costs considerably, but larger student-tutor ratios may not be as effective as smaller ratios are. Optimizing this ratio is critical because tutor compensation is the greatest expense for most programs. Our evidence together with a growing number of studies demonstrates that student-tutor ratios can be expanded to 3:1 student-tutor groups without diminishing effectiveness (Kraft et al., 2024). If 1:1 and 2:1 high impact tutoring programs move to 3:1 student-tutor group sizes, these programs can redirect funds to increase the number of students served. In our study, moving from 2:1 to 3:1 group sizes reduced the total cost per student of the program by 26%. Along with other design considerations, this study's program provided tutoring three days a week, which, for the purposes of tutor recruitment, could be a more scalable model than five-day programs. Finally, effective tutor training and overall program implementation must be paramount for university-led high impact tutoring programs to be successful at scale. Leveraging the expertise of university staff, research assistants, and faculty can help to strengthen program quality while also providing ongoing research on program effectiveness and areas for improvement.

As tutoring programs seek to bring their models to scale, future studies should continue to examine approaches to optimizing high impact tutoring programs. Nevertheless, financial, logistical, and human resource constraints are likely to make it too difficult to deliver these types of programs for many schools. In these cases, other cost-effective alternative approaches are needed (Kraft & Novicoff, 2024). In this study, delivering a remedial mathematics class with a classroom teacher was far less costly and human-resource intensive than high impact tutoring was. Moreover, rigorous studies have shown strong academic effects for supplementary instructional time in core subjects (Angrist et al., 2016; Cattaneo et al., 2017; Cohodes & Parham, 2021; Cortes et al., 2014). Future work is thus needed to test whether remedial high school mathematics courses might be a cost-effective alternative to high impact tutoring programs.

References

Adams, C., & Khojasteh, J. (2018). Igniting students' inner determination: The role of a need-supportive climate. *Journal of Educational Administration*, *56*(4), 382-397. https://doi.org/10.1108/JEA-04-2017-0036

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, *34*(2), 275-318. https://doi.org/10.1086/683665

Aughinbaugh, A. (2012). The effects of high school math curriculum on college attendance: Evidence from the NLSY97. *Economics of Education Review*, *31*(6), 861-870. https://doi.org/10.1016/j.econedurev.2012.06.004

Cattaneo, M. A., Oggenfuss, C., & Wolter, S. C. (2017). The more, the better? The impact of instructional time on student performance. *Education Economics*, *25*(5), 433-445. https://doi.org/10.1080/09645292.2017.1315055

Cerasoli, C. P., & Ford, M. T. (2014). Intrinsic motivation, performance, and the mediating role of mastery goal orientation: A test of self-determination theory. *The Journal of Psychology*, *148*(3), 267-286. https://doi.org/10.1080/00223980.2013.783778

Cha, S. H. (2015). Exploring disparities in taking high level math courses in public high schools. *KEDI Journal of Educational Policy*, *12*(1), 3-17.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review, 104*(9), 2633-2679. 10.1257/aer.104.9.2633

Cohodes, S. R., & Parham, K. S. (2021). *Charter schools' effectiveness, mechanisms, and competitive influence* (No. w28477). National Bureau of Economic Research. https://doi.org/10.3386/w28477

Cohodes, S., Setren, E., & Walters, C. (2019). The long-run impact of childhood access to a high-quality school: Evidence from the opening of Boston's charter schools. *National Bureau of Economic Research.*

Cohodes, S. R., Setren, E. M., & Walters, C. R. (2021). Can successful schools replicate? Scaling up Boston's charter school sector. *American Economic Journal: Economic Policy*, *13*(1), 138-167. https://doi.org/10.1257/pol.20180510

Colvin, J. W. (2007). Peer tutoring and social dynamics in higher education. *Mentoring & tutoring, 15*(2), 165-181. https://doi.org/10.1080/13611260601086345

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *Journal of Human Resources*, *50*(1), 108-158. https://doi.org/10.3368/jhr.50.1.108

Coyle, T. R., Purcell, J. M., Snyder, A. C., & Richmond, M. C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence*, *46*, 18-24. https://doi.org/10.1016/j.intell.2014.04.008

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 416–436). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n21

de Ree, J., Maggioni, M. A., Paulle, B., Rossignoli, D., & Walentek, D. (2021). *High impact tutoring in pre-vocational secondary education: Experimental evidence from Amsterdam*. https://doi.org/10.31235/osf.io/r56um

Dougherty, C., & Fleming, S. (2012). *Getting students on track to college and career readiness: How many catch up from far behind?* ACT Research Report Series.

Duncan, G. J., et al. (2007). School Readiness and Later Achievement. *Developmental Psychology, 43*(6), 1428-1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duran, D. (2017). Learning-by-teaching. Evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International, 54*(5), 476–484. https://doi.org/10.1080/14703297.2016.1156011.

Egan, K. L., & Davidson, A. H. (2017, Nov. 14). Alignment of the NWEA MAP Growth & MAP Growth K–2 to the Common Core State Standards: English language arts and mathematics. EdMetric.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap?. *The High School Journal*, *91*(1), 29-42. https://www.jstor.org/stable/40367921

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125-149. https://doi.org/10.1162/EDFP_a_00180

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., ... & Steinberg, L. (2021). *Not too late: Improving academic outcomes among adolescents* (No. w28531). National Bureau of Economic Research. https://doi.org/10.3386/w28531

Hanushek, E. A. (2023). Generation lost: The pandemic's lifetime tax. Education Next. https://www.educationnext.org/generation-lost-the-pandemics-lifetime-tax/

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902. https://doi.org/10.1126/science.1128898.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics, 24*(3), 411-482. https://doi.org/10.1086/504455

Heinrich, C. J., Darling-Aduana, J., Good, A., & Cheng, H. (2019). A look inside online educational settings in high school: Promise and pitfalls for improving educational opportunities and outcomes. *American Educational Research Journal, 56*(6), 2147-2188. https://doi.org/10.3102/0002831219838776

Hickey, A. J., & Flynn, R. J. (2019). Effects of the TutorBright tutoring programme on the reading and mathematics skills of children in foster care: A randomised controlled trial. *Oxford Review of Education*, *45*(4), 519-537.

Howard, J. L., Bureau, J., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science*, *16*(6), 1300-1323. https://doi.org/10.1177/1745691620966789

Institute of Education Sciences. (2020). What Works Clearinghouse standards handbook (Version 4.0). U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Joensen, J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes?. *Journal of Human Resources*, *44*(1), 171-198. https://doi.org/10.3368/jhr.44.1.171

Jordan, P., DiMarco, B., & Toch, T. (2022, April). *An analysis of local school districts' ambitious post-Covid tutoring plans*. FutureEd. https://www.future-ed.org/an-analysis-of-local-schooldistricts-ambitious-post-covid-tutoring-plans/

Karsenty, R. (2010). Nonprofessional mathematics tutoring for low-achieving students in secondary schools: A case study. *Educational Studies in Mathematics, 74*, 1-21.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253

Kraft, M. A., Schueler, B., & Falken, G. (2024). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability. (EdWorkingPaper: 24-1031). https://doi.org/10.26300/zygj-m525

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

Kraft, M. A., & Novicoff, S. (2024). Time in School: A conceptual framework, synthesis of the causal research, and empirical exploration. *American Educational Research Journal, 61*(4), 724-766.

Kotok, S. (2017). Unfulfilled potential: High-achieving minority students and the high school achievement gap in math. *High School Journal*, *100*(3), 183-202. https://www.jstor.org/stable/90024211

Larson, R., & Boswell, L. (2019). *Big Ideas Math: Algebra 1. A Common Core Curriculum*. Eric, PA: Big Ideas Learning.

Lohmeier, J. H., & Lee, S. W. (2011). A school connectedness scale for use with adolescents. *Educational Research and Evaluation*, *17*(2), 85-95. https://doi.org/10.1080/13803611.2011.597108

Long, M. C., Iatarola, P., & Conger, D. (2009). Explaining gaps in readiness for college-level math: The role of high school courses. *Education Finance and Policy*, *4*(1), 1-33. https://doi.org/10.1162/edfp.2009.4.1.1

Moffatt, P. (2020). *Experimetrics: Econometrics for experimental economics*. Bloomsbury Publishing.

National Center for Education Statistics (NCES). (2022). Revenues and Expenditures for Public Elementary and Secondary Education: FY 20. U.S. Department of Education. https://nces.ed.gov/pubs2022/2022306.pdf

The National Assessment of Education Progress. (2022). *Reading and mathematics scores decline during the pandemic*. https://www.nationsreportcard.gov/highlights/ltt/2022/

The National Assessment of Education Progress. (2024). *NAEP US math score trends*. https://www.nagb.gov/naep/mathematics.html

National Council of Teachers of Mathematics (2014). Principles to Actions: Ensuring success for all. (S. Leinwand et al., Eds.). National Council of Teachers of Mathematics.

National Research Council, & Mathematics Learning Study Committee. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press.

Nickow, A., Oreopoulos, P., & Quan, V. (2020). *The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence* (No. w27476). National Bureau of Economic Research. https://doi.org/10.3386/w27476

Nomi, T., & Allensworth, E. (2009). Double-dose Algebra as an alternative strategy to remediation: Effects on students' outcomes. *Journal of Research on Educational Effectiveness*, *2*(2), 111–148. https://doi.org/10.1080/19345740802676739

Northwest Evaluation Association (NWEA). (2021). MAP growth norms study: Comparing 2020 to 2015. https://teach.mapnwea.org/impl/MAPGrowthNormativeDataOverview.pdf

Northwest Evaluation Association (NWEA). (2019). MAP Growth technical report. Portland, OR: Author.

Office of State and Grantee Relations, OESE. (2024). Elementary and Secondary School Emergency Relief Fund. Formula Grant. Retrieved June 28, 2024, from https://oese.ed.gov/offices/education-stabilization-fund/elementary-secondary-school-emergency-relief-fund/

Oklahoma Educational Quality and Accountability. (2024). Oklahoma School Report Cards. https://schoolreportcards.ok.gov/

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, *26*(7), 1372-1380. https://doi.org/10.1016/j.tate.2010.03.008

Robinson, C. D., Kraft, M. A., Loeb, S., & Schueler, B. E. (2021). *Accelerating student learning with high impact tutoring*. *EdResearch for Recovery Design Principles Series*. EdResearch for Recovery Project. https://files.eric.ed.gov/fulltext/ED613847.pdf

Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534-574. https://doi.org/10.3102/0034654307309920

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, *86*(2), 497-513. https://doi.org/10.1162/003465304323031076

Siegler, R. S., & Braithwaite, D. W. (20r17). Numerical Development. *Annual Review of Psychology*, 68, 187-213. https://doi.org/10.1146/annurev-psych-010416-044101

Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, *39*(4), 342-358. https://doi.org/10.1016/j.cedpsych.2014.08.002

Thum, Y.M. & Kuhfeld, M. (2020). *NWEA 2020 MAP Growth: Achievement status and growth norms tables for students and schools*. NWEA. https://teach.mapnwea.org/impl/MAPGrowthNormativeDataOverview.pdf

Trusty, J., & Niles, S. G. (2003). High-school math courses and completion of the bachelor's degree. *Professional School Counseling*, *7*(2), 99-107. https://www.jstor.org/stable/42732549

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, *12*(3), 243-270. https://doi.org/10.1080/10824660701601266

Appendix

Table 1A. Attrition Analysis

Variables	Pooled Sample	Year 1	Year 2	Year 3
High Impact Tutoring Treatment	0.95	1.30	0.70	1.02
	(0.19)	(0.59)	(0.24)	(0.30)
Female	0.96	2.71	1.10	0.61
	(0.18)	(1.39)	(0.36)	(0.18)
Free and Reduced-Price Lunch	2.55**	--	2.67*	2.33*
	(0.74)		(1.22)	(0.93)
Latino/a	0.92	0.58	1.05	1.09
	(0.32)	(0.36)	(0.71)	(0.64)
Black	0.95	--	2.00	0.71
	(0.47)		(1.77)	(0.51)
White	1.59	--	2.67	1.38
	(0.59)		(1.81)	(0.82)
Native	1.13	--	1.26	1.31
	(0.55)		(1.22)	(0.91)
English language learner	0.72	0.52	1.37	0.55
	(0.18)	(0.25)	(0.65)	(0.21)
Student with Disability	1.00	1.10	0.76	1.26
	(0.25)	(0.75)	(0.34)	(0.44)
Constant	0.07***	0.11**	0.04***	0.09***
	(0.03)	(0.08)	(0.03)	(0.05)
Observations	1,092	223	370	499

Robust standard errors in parentheses. *** $p<0.001$, ** $p<0.01$, * $p<0.05$
Logistic regression predicting study attrition. Coefficients are presented as odds ratios.

Table 2A. Estimated effects on mathematics achievement and grade-point average

	Intent-to-Treat (ITT)		Treatment Effect (TOT)	
Variables	EOY Math Score	EOY GPA	EOY Math Score	EOY GPA
High Impact Tutoring Treatment	1.64*	0.08	1.60*	0.08
	(0.78)	(0.08)	(0.80)	(0.08)
Baseline Math RIT Score	0.78***	0.03***	0.79***	0.03***
	(0.03)	(0.00)	(0.03)	(0.00)
Demographic Factors	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y
Cohort Year Fixed Effects	Y	Y	Y	Y
Control Group Mean	217.94***	2.50***	217.80***	2.49***
	(0.42)	(0.04)	(0.43)	(0.04)
Students	946	946	919	919
Schools	6	6	6	6

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Note. The analyses have a different sample than the main models, excluding one school that did not have a control group in Year 3. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools. Cohort year fixed effects are binary indicators representing the years of analysis.

Table 3A. Effect of high impact tutoring on treated students (9^{th} grade) by minutes tutored

Variables	2,000 Minutes	2,600 Minutes	3,000 Minutes
HIT treatment minutes	1.89*	1.59*	1.67
	(0.80)	(0.73)	(1.04)
Baseline Math Score	0.79***	0.79***	0.78***
	(0.04)	(0.03)	(0.03)
Demographic Factors	Y	Y	Y
School Fixed Effects	Y	Y	Y
Cohort Year Fixed Effects	Y	Y	Y
Control Group Mean	218.16***	218.32***	218.48***
	(0.41)	(0.33)	(0.38)
Students	897	790	691
Schools	7	7	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05
Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools. Cohort year fixed effects are binary indicators for the years of analysis.

Table 4A. Moderator analysis: ITT estimates on mathematics achievement by student subgroup

Variables	FRL Status	Latino/a	Black	ELL Status	Special Ed.	Low-achieving	Charter	Rural
High Impact Tutoring Treatment	0.19	1.67*	1.48*	0.94	1.55	1.35	1.86	2.03*
	(0.90)	(0.84)	(0.74)	(0.85)	(0.89)	(0.70)	(1.62)	(0.93)
Group Main Effect	-2.63***	2.92*	-3.73	-2.13**	-2.22	0.40	-6.84***	-5.32***
	(0.57)	(1.36)	(2.90)	(0.82)	(1.37)	(0.66)	(1.38)	(1.10)
Interaction Effect	1.85*	-0.22	1.37	1.72	-0.00	0.72	-0.64	-2.05
	(0.84)	(0.93)	(3.57)	(1.09)	(2.03)	(0.83)	(1.68)	(1.93)
Baseline Math RIT Score	0.78***	0.78***	0.78***	0.78***	0.78***	0.81***	0.78***	0.78***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Demographic Factors	Y	Y	Y	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Cohort Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Control Group Mean	218.13***	218.14***	218.13***	218.23***	218.15***	218.16***	218.09***	218.17***
	(0.40)	(0.43)	(0.46)	(0.42)	(0.43)	(0.42)	(0.61)	(0.42)
Students	963	963	963	963	963	963	963	963
Schools	7	7	7	7	7	7	7	7

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05. Note. EOY Math Score is the math RIT score on the NWEA MAP assessment. EOY GPA is the end of year grade point average (4.0 scale). Baseline math score is the fall math RIT score on the NWEA MAP assessment. Demographic factors are sex, FRL status, Race/Ethnicity, ELL status, and IEP status. School fixed effects are binary indicators for schools. Cohort year fixed effects are binary indicators for the years of analysis.

Table 5A. Cost effectiveness analysis of university-led high impact tutoring

	Price/Unit ($)	High impact tutoring ($)	Remedial math class ($)
Teacher Compensation	9,300	269,700	204,600
Tutor Stipend & Scholarship (2:1)	9,400	1,959,900	
Tutor Stipend & Scholarship (3:1)	9,400	338,400	
Tutor Management	80,000	80,000	
Tutor Training	40,000	80,000	
Supplies	500	500	
Training Space	5,000	5,000	
Total Cost		2,733,500	204,600
Total Cost / Student		5,207	467
Additional Cost / Student of University- led High Impact Tutoring		4,740	