Impacts of UPchieve On-Demand Tutoring on Students' Math Knowledge and Perceptions

Grady Deacon and Greg Chojnacki

January 2023

Contents

Figures

Table

Middle Years Math Grantee Report Series

This report is one in a series of six reports on math tutoring programs. Over the 2020–2021 and 2021–2022 school years, the Bill & Melinda Gates Foundation invested in rapid-cycle evaluations of a cohort of 10 tutoring providers to learn about their innovative approaches to tutoring as part of its Middle Years Math body of work.[1] The goal of these investments was to understand how different tutoring models might create positive student experiences and lead to improved academic outcomes for students in the foundation's priority communities—those who are Black, Latino, and/or experiencing poverty. These investments were grounded in the substantial body of evidence supporting the effectiveness of tutoring in improving student math knowledge (Nickow et al., 2020).

To build on this existing evidence of effectiveness, the Gates Foundation sought to develop new early evidence about the success of a range of tutoring approaches. Specifically, these investments targeted two key learning priorities. First, the foundation sought to learn how innovative technologies and tutoring program design features might simultaneously improve the quality and lower the cost of tutoring, making high-quality tutoring available to a large number of students in priority communities. The second priority was to learn the extent to which tutoring programs resulted in positive experiences for participating students. To learn about tutoring design features, the foundation invested in tutoring programs with a wide range of approaches, including group and one-on-one tutoring, virtual and in-person models, professional teachers as tutors, or volunteer tutors who shared aspects of identity with tutored students. Tutoring programs also used different approaches to tutoring curriculum and pedagogy. The goal of this report series is to inform the tutoring field more broadly and support the provision of high-quality tutoring to as many students in the priority communities as possible.

To learn rapidly about tutoring providers' innovative approaches, Mathematica worked with each one to identify the most rigorous study design that would be feasible for district partners within a one-to-three-month planning period. Some providers were able to design and implement randomized controlled trials; others used quasi-experimental designs such as matched comparison approaches. One study compared growth in math knowledge among participants to the growth observed in national samples because it was not possible to obtain student-level data for comparison students who did not receive tutoring. These relatively small studies were right-sized to the development stage of the tutoring program and sought to demonstrate early evidence of success before moving on to larger-scale effectiveness studies. To help synthesize findings about student experiences from multiple providers, studies used the same student survey measures of tutor relationship, math confidence, and sense of belonging in tutoring sessions. Most of the studies used standardized math knowledge assessments aligned with Common Core State Standards.

Each study also aimed to inform providers' efforts to refine their programs and support successful implementation. These studies measured the amount of tutoring offered, attendance, and staff impressions about implementation challenges while also gathering qualitative data on students' experiences. Findings from these studies have helped to direct tutoring providers' next steps in refining and scaling their tutoring programs.

---

[1] This publication is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Executive Summary

What is the tutoring program we studied?

UPchieve is an ed-tech nonprofit that provides students in grades 8 to 12 with access to online tutoring, 24 hours a day, seven days a week. To access UPchieve's online tutoring platform, students must attend a Title I school where large shares of students are experiencing poverty or live in a zip code where the median household income is less than $60,000. Using the platform, students request help and get paired with a live, UPchieve-certified volunteer tutor, where students use a chat feature and virtual whiteboard to receive assistance with a topic of their choosing. UPchieve recruits, trains, and certifies unpaid volunteers to become tutors. Tutoring is available across all core subjects (math, English language arts, science, social studies) as well as college counseling and standardized exam preparation. (More details on UPchieve's tutoring are available at upchieve.org and in Appendix A.) This study focused specifically on UPchieve's math tutoring, in classes ranging from pre-algebra to calculus.

What questions does this study answer?

1. To what extent do incentives and encouragement raise students' use of UPchieve tutoring?
2. Does increased use of UPchieve lead to math knowledge gains? If so, how do gains vary across specific student groups, such as students who are Black or Latino, or those who face more barriers to math engagement?
3. Does increased use of UPchieve lead to students reporting higher levels of math confidence and engagement?

How was the study conducted?

Study design. Over 13 weeks from November 2021 to January 2022, UPchieve recruited students in grades 9 and 10 who found UPchieve online, usually by way of an online search engine, and qualified to use the tutoring platform. Each week, a new cohort of students who completed a baseline survey and assessment were randomly assigned to the treatment and control groups. Each cohort comprised 2 to 16 students for a total of 99 eligible students enrolled in the study. The study period lasted 14 weeks for each cohort. The first cohort ended February 7, 2022, and the thirteenth cohort ended May 2, 2022. Students in the treatment group received weekly encouragements and a $5 financial incentive for each study week in which they completed at least one UPchieve math tutoring session. Students in the control group did not receive encouragements or incentives but still had access to UPchieve tutoring. The goal of this study design was to learn about the impact of increased UPchieve use by comparing outcomes from treatment group, who we encouraged to use UPchieve more often, to those of the control group.

Measures and analysis. The study collected several data sources from the treatment and control groups before and after the study period to measure outcomes of the additional tutoring: the Renaissance Star Math assessment of math knowledge, the Expectancy-Value-Cost Scale (EVC) to measure motivation in math, the Patterns of Adaptive Learning Scale (PALS) to measure confidence in math, and the Math Engagement Scales to measure engagement in math. We examined the impact of increased UPchieve use on students' outcomes using a Bayesian linear regression model, which controlled for students' baseline

Star Math scores, EVC scores of math motivation, cohort, and a series of demographic indicators. Eighty-nine students had complete math achievement data, and 91 students had complete survey data measuring math motivation, confidence, and engagement. The study also conducted focus groups with students and teachers to co-interpret preliminary findings and inform which of these findings were most valuable to highlight in the final report.

Limitations. The study's small sample limited the strength of conclusions that can be drawn about whether additional tutoring for students in the treatment group boosted their math knowledge, as described below. Additionally, this study only included students who signed up proactively for UPchieve online and then completed a baseline assessment and survey and were compensated $50. The effect of UPchieve tutoring might differ for students who are less proactive about seeking additional support in math. Last, because the study used encouragements and incentives for treatment group students to measure the effect of increased UPchieve tutoring usage, results do not measure the impact of receiving tutoring related to outcomes for students who had no access to tutoring. Instead, it measures the impact of increased UPchieve tutoring usage relative to lower levels of UPchieve tutoring usage.

What did the study find?

Usage. Students who were encouraged and received incentives to participate in UPchieve tutoring weekly used UPchieve more on average (9 times) than students who were not (3.5 times), leading to 1.6 hours more tutoring. The impact of these encouragements and incentives appeared to be larger among students with lower engagement at baseline.

Math knowledge. Students in the treatment group scored 9 percentile points higher on the Star Math test than control group students, after accounting for baseline differences. The estimated impact of 0.23 standard deviations was large enough that a Bayesian analysis indicated a 91 percent probability that the true impact is positive. This effect size is similar to that of a 2020 review of 24 in-person tutoring interventions led by non-professionals (Nickow et al., 2020)**.**

Student perceptions and beliefs. Students in the treatment group experienced more growth in math motivation, confidence, and engagement. We estimated that they experienced the largest improvements in math motivation as measured by the EVC.

This study presents what we believe to be the first evidence of the impact of on-demand, online tutoring on student math outcomes by measuring differences in test scores resulting from increased tutoring use. UPchieve's impact is outsized compared to other online tutoring models, particularly with regard to the cost and time associated with implementing the program. This is important because rigorous research into virtual tutoring is limited, and the little existing research for online math tutoring that has focused on lower-dosage models has found mixed results (Kraft et al., 2022; Roschelle et al., 2020). These findings underscore the value of a larger randomized controlled trial to validate whether on-demand, online tutoring can produce effects comparable to in-person tutoring interventions. If validated, this could aid efforts to scale tutoring, because recruiting tutors locally is one of the largest barriers to scaling in-person, high-impact tutoring models.

Introduction

UPchieve is a free online and on-demand tutoring platform for students in grades 8 to 12 who attend Title I schools where large shares of students are experiencing poverty. UPchieve matches students with unpaid volunteer tutors one on one and in an online setting. In this setting, students can take advantage of a chat feature and virtual whiteboard to receive assistance with a topic of their choosing, including tutoring in math, science, English, or history; SAT preparation; and college counseling. UPchieve tutors are available 24 hours a day, seven days a week, so that students can receive support at any time during the day or night.

This study aims to provide evidence of the impact of on-demand, online tutoring on student math achievement and perceptions about math using a pilot randomized controlled trial. Specifically, we examined the impact of participating in additional UPchieve math tutoring sessions on 9th- and 10th-grade students' math achievement, motivation, confidence, and engagement. In this evaluation, we sought to answer the following research questions:

1. To what extent do incentives and encouragement raise students' use of UPchieve tutoring?

2. Does increased use of UPchieve lead to math knowledge gains? If so, how do gains vary across specific student groups, such as students who are Black or Latino, or those who face more barriers to math engagement?

3. Does increased use of UPchieve lead to students reporting higher levels of math confidence and engagement?

Key Findings

Students who were encouraged and incentivized to participate in UPchieve tutoring sessions on a weekly basis used UPchieve more than students who were not.

Students in the treatment group received weekly email encouragements to participate in UPchieve tutoring over 14 weeks, as well as $5 for each of those weeks in which they logged at least one tutoring session in math. Control group students did not receive any weekly encouragements or incentives. Students in the treatment group used UPchieve approximately 9.0 times on average over a 14-week period, compared with approximately 3.5 times on average in the control group. Over 14 weeks, nearly two-thirds of the students in the treatment group (62 percent) logged two or more math sessions, whereas less than one-third of the students in the control group (32 percent) logged two or more sessions over the same time period (Figure 1).

In separate focus groups designed to co-interpret analysis results, teachers and students agreed these data demonstrate that reminding and incentivizing students to participate in tutoring outside of school was effective for increasing the amount of tutoring students engage in. In focus groups, multiple students in the treatment group reported that they tended to forget about UPchieve tutoring, but the emails reminded them that they could earn $5 by participating in a math tutoring session that week.

Figure 1. UPchieve math session usage by treatment status
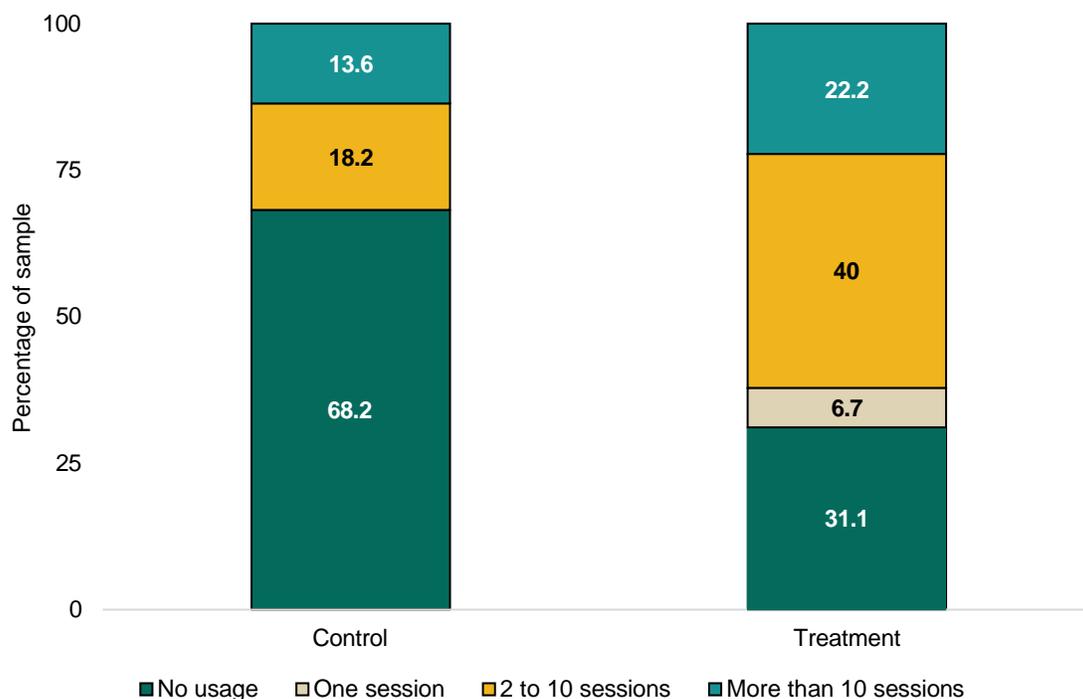
Source: Student data collected from UPchieve randomized controlled trial.

Notes: Sample size was 89 students—44 in the control group and 45 in the treatment group.

Usage patterns suggest that encouragements and incentives were more effective at increasing UPchieve math tutoring usage for students with low reported math engagement at baseline.

Students in the study took a baseline survey measuring math motivation, confidence, and engagement. We examined how students who reported very low and low levels baseline math engagement were influenced by UPchieve's encouragements and incentives to participate in UPchieve tutoring. Math engagement was measured by the Math Engagement Scales (Wang et al., 2016). This survey asked students to rate the extent to which they identify with a series of statements about their behavioral, emotional, cognitive, and social engagement with math in the classroom and at home on a 5-point scale, from "not at all like me" to "very much like me." Students who reported responses of less than 2.75, or more than a quarter point less than "somewhat like me," were defined as having very low engagement in math. Students who reported responses of between 2.75 and less than 3.00, or up to a quarter point less than "somewhat like me" were defined as having low engagement. Each of these categories contained approximately one-fifth of the student sample.

Among students in the control group with very low baseline math engagement scores (a score of less than 2.75 out of 5.00), none completed any UPchieve tutoring sessions in math during the study period. In contrast, 80 percent of students in the treatment group with very low baseline engagement scores completed at least one math tutoring session (Figure 2). Among students in the control group with low baseline math engagement scores (a score greater than or equal to 2.75 and less than

3.00), 14 percent of students completed at least one UPchieve tutoring session in math during the study period, compared with 50 percent of students in the treatment group.

As the baseline engagement level increased across the control group, the proportion of students who logged at least one UPchieve tutoring session also increased. In contrast, among treatment group students, at every level of baseline engagement, at least 50 percent of students completed one or more sessions (Figure 2). Although control group students with higher levels of engagement at baseline tended to use UPchieve more, treatment group students in every engagement category were more likely than their control group counterparts to complete at least one tutoring session. This observation was validated in focus groups designed to co-interpret analysis results. During one of the focus groups, teachers and staff reviewed the data in Figure 2 and commented that encouragements and incentives were particularly effective for students with lower levels of baseline engagement, and that students who already had higher levels of baseline engagement "were going to use it more than others." One of the teachers elaborated on their personal relationship with students with low engagement, saying that encouraging them to engage in the material with incentives serves as an "entry point" that could effectively help these students overcome their barriers.

Together, these results suggest that the email encouragements and the weekly $5 incentive were especially helpful for students who reported low or very low engagement at the beginning of the study. Although these results need to be interpreted with caution—each category of baseline engagement only includes approximately nine students—the overall pattern of usage across categories of baseline engagement provides suggestive evidence that incentives may boost usage, especially among those who report lower math engagement at baseline.

Figure 2. Percentage of students who participated in at least one UPchieve tutoring session, by student baseline engagement and treatment status
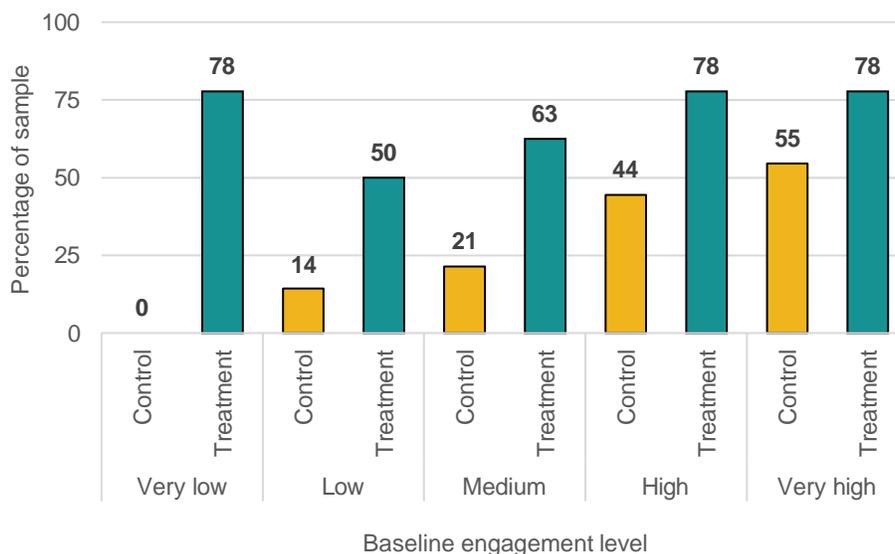
Source: Student data collected from UPchieve randomized controlled trial.

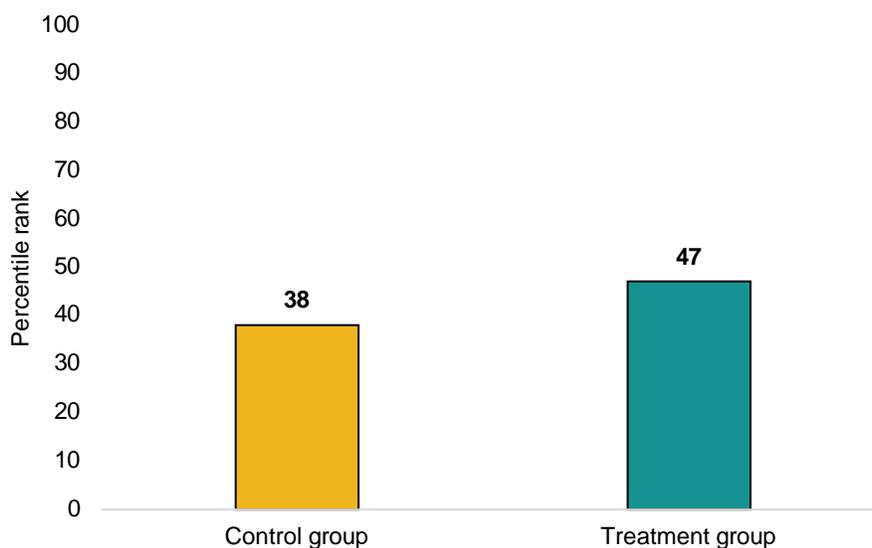Notes: Sample size was 89 students—44 in the control group and 45 in the treatment group. Baseline math engagement level was measured using the global engagement score of the Math Engagement Scales survey module, which included questions such as "I enjoy learning new things about math" and used a scale from 1 (Not at all like me) to 5 (Very much like me). Very low engagement was defined as a score of

less than 2.75. Low engagement was defined as a score greater than or equal to 2.75 and less than 3.00. Medium engagement was defined as a score greater than or equal to 3.00 and less than 3.50. High engagement was defined as a score greater than or equal to 3.50 and less than 4.00. Very high baseline engagement was defined as a score greater than or equal to 4.00.

Students in the treatment group experienced more growth on the Star math assessment.

As discussed above, students who received encouragements and incentives to use UPchieve completed more sessions on average (9 sessions) than students who did not (3.5 sessions). We find this increase in usage of UPchieve math tutoring sessions (about 5.5 additional sessions) resulted in students scoring 9 percentile points higher on the Star Math assessment at follow-up (Figure 3). Although these findings are not statistically significant according to a conventional statistical test (p = 0.12), a Bayesian analysis indicates there is a 91 percent probability that this increase in UPchieve tutoring had a positive impact on students' math knowledge. (See Appendix B for details on analysis methods.) This estimated impact of additional UPchieve usage translates to students scoring 0.23 standard deviations higher on the Star Math assessment at follow-up than their lower-usage counterparts. Given that UPchieve utilizes unpaid volunteer tutors rather than the paid staff that are common in high-dosage tutoring models, this impact on math knowledge underscores the promise of this alternative staffing model for improving student math outcomes among students who proactively sought out math tutoring. It would be useful to understand whether UPchieve can demonstrate a similar impact among a broader set of students, including those who are less inclined to actively seek out math help.

Figure 3. Average percentile rank on the post-study Star math assessment by treatment status
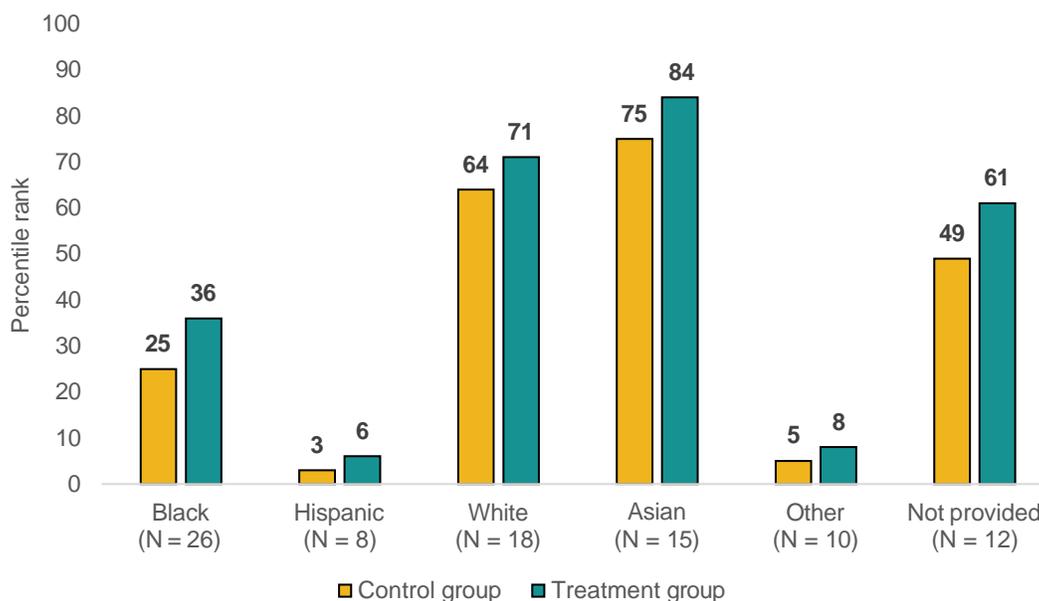

Source: Student data collected from UPchieve RCT.

Notes: Sample size was 89 students—44 in the control group and 45 in the treatment group. Percentile rank indicates how well students score on the Star Math assessment compared to other students in a national sample. These estimates adjust for differences between treatment and control baseline Star Math test scores.

Across race and ethnic groups, students experienced positive effects on their Star Math achievement, on average.

Students of each race and ethnic subgroup experienced higher math achievement scores in the treatment group than in the control group (Figure 4). Students in the treatment group outperformed their control group counterparts by a range of 3 percentile rank points (Hispanic students) to 11 percentile points (Black students). Across all races and ethnicities, the probability that these additional sessions led to a positive impact on students' math knowledge ranged from 85 percent (White students) to 96 percent (Black students). Due to small sample sizes, we caution against comparing the size of these improvements across subgroups.

Figure 4. Average percentile rank on the post-study Star assessment by race and treatment status

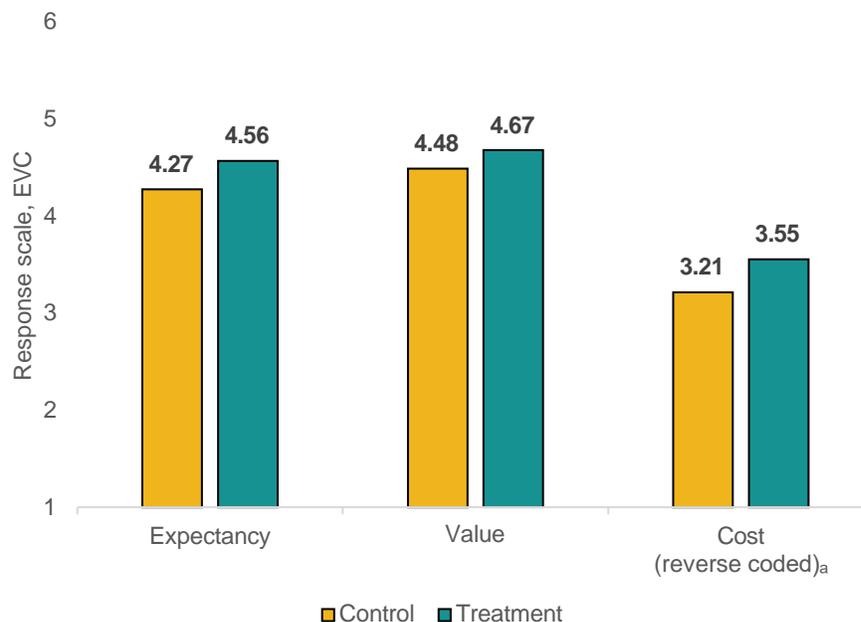Source: Student data collected from UPchieve randomized controlled trial.

Notes: Sample size was 89 students—44 in the control group and 45 in the treatment group. Other students include Middle Eastern, Native American, and Hawaiian students. These estimates adjust for differences between treatment and control baseline Star math test scores.

Students who used UPchieve more experienced more growth in math motivation, confidence, and engagement.

Students in the treatment group experienced large improvements in math motivation as measured by the Expectancy-Value-Cost scale compared to students in the control group (Kosovich et al., 2015). When asked to rate the extent they agree with a series of statements about how successful they could be in math class (known as the "expectancy" of success) on a 6-point scale ranging from "strongly disagree" to "strongly agree," students in the treatment group reported an average response of 4.56, or about midway between "slightly agree" and "agree." This response was 0.29 points higher than for students in the control group, who reported responses of 4.27 on average, or about a quarter of a point above "slightly agree." When responding to questions about how much they value math class, students in the treatment group reported an average response of 4.67. This response was 0.19 points higher than for students in the control group, who reported responses of 4.48 on average. Last, the survey also asked students to respond to questions about the costs of

participating in math class—for example, asking the extent to which they agree that their "math classwork requires too much time." Students in the treatment group reported an average response of 3.55, or 0.34 points more positive than the control group average of 3.21, indicating they perceived a lower cost to participating in math class. Overall, Bayesian analyses indicate the probability that the additional UPchieve tutoring sessions led to a positive impact on students' math motivation ranged from 84 percent for value to 93 percent for cost (conventional statistical tests yielded p-values ranging from 0.26 to 0.05, respectively) (Figure 5).

Figure 5. Average post-study math motivation by treatment status

Source: Student data collected from UPchieve randomized controlled trial.

Notes: Sample size was 91 students—44 in the control group and 47 in the treatment group. Expectancy, value, and cost subconstructs rated on a 6-point scale from 1 (Strongly disagree) to 6 (Strongly agree). These estimates adjust for differences between treatment and control baseline scores on the Expectancy, Value, and Cost Scales survey.

[a] The cost subconstruct has been reverse coded, so that a high value represents a lower perceived cost of participating in math.

Student improvements in math confidence were less pronounced than improvements in math motivation. The Patterns of Adaptive Learning Scales measure asked students how certain or confident they are in their ability to learn and persist in learning math on a 5-point scale, ranging from "not at all" to "very" (Midgley et al., 2000). Students in the control group averaged a score of 3.53, or about a half point above a response of "somewhat." After controlling for baseline differences in student confidence, students in the treatment group reported a score of 3.61 on average, or 0.08 points higher than students in the control group. Bayesian analyses indicate the probability that the additional UPchieve tutoring sessions led to a positive impact on students' math confidence was 70 percent (conventional statistical tests yielded a p-value of 0.45).

Students also experienced benefits from increased UPchieve usage in terms of math engagement. Students were asked to rate the extent to which they identify with a series of statements about their behavioral, emotional, cognitive, and social engagement with math in the classroom and at home on

a 5-point scale, from "not at all like me" to "very much like me" (Wang et al., 2016). Students in the control group reported a score of 3.63 on average, or nearly two-thirds of a point higher than "somewhat like me." After controlling for baseline differences in student engagement, students in the treatment group reported a score of 3.73 on average, 0.10 points higher than the control group and closer to a 4.0, or a response of "mostly like me." Bayesian analyses indicate the probability that the additional UPchieve tutoring sessions led to a positive impact on students' math engagement was 91 percent (conventional statistical tests yielded a p-value of 0.12) (Figure 6).

Figure 6. Average post-study confidence and engagement in math by treatment status
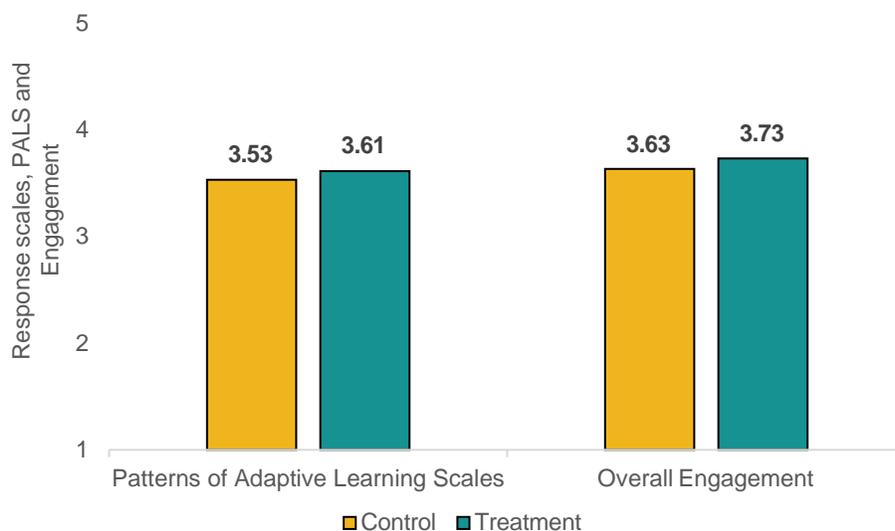
Source: Student data collected from UPchieve randomized controlled trial.

Notes: Sample size was 91 students—44 in the control group and 47 in the treatment group. Patterns of Adaptive Learning Scales (PALS) rated on a 5-point scale from 1 (Not at all) to 5 (Very). Overall engagement rated on a 5-point scale from 1 (Not at all like me) to 5 (Very much like me). These estimates adjust for differences between treatment and control baseline scores on the PALS and Math Engagement Scales surveys.

Overall, these findings indicate that the additional tutoring that treatment group students received led to improvements in their motivation, confidence, and engagement with math. It is possible that in addition to the effect of additional tutoring on these outcomes, the students in the treatment group also benefited from the encouragements they received in weekly emails.

Conclusion and Next Steps

These findings suggest that when students engage in more UPchieve tutoring, it has a meaningful impact on their math knowledge. They also suggest that UPchieve tutoring has large positive impacts on students' math motivation. These findings are useful for the field of math tutoring because they are examples of preliminary evidence that on-demand, online tutoring drawing on unpaid, volunteer tutors improves math achievement and motivation among the students in focus for this study—those who are Black, Latino and/or experiencing poverty. The effect size of the additional UPchieve sessions on students' Star Math assessment scores, 0.23 standard deviations, is similar to that of a 2020 review of 24 in-person tutoring interventions led by non-professionals (Nickow et al., 2020). This is important because rigorous research into virtual tutoring is limited, and the little existing research for online math tutoring that has focused on lower-dosage models has found mixed results (Kraft et al., 2022; Roschelle et al., 2020). During stakeholder focus groups about the study results, students connected their participation in UPchieve tutoring to learning in math class. One student noted that the tutoring they received from UPchieve helped them understand what they did in their math class, consequently improving how they valued math. Additionally, teachers were impressed with the level of improvement that students in the treatment group experienced and actively brainstormed opportunities to implement the tutoring program as a supplement to their teaching.

The findings in this study suggest that UPchieve's impact is outsized for its model, particularly with regard to the time associated with using the program. These findings point to the value of a larger randomized controlled trial to measure more precisely whether on-demand, online tutoring can produce effects comparable to in-person tutoring interventions. If validated, this could have important consequences for how tutoring is scaled, as recruiting tutors locally is one of the largest barriers to scaling in- person, high-impact tutoring models. Ultimately, these findings provide a first glance at the possibility of low-cost tutoring programs as a viable solution to recovering from COVID-19 learning gaps.

To this end, UPchieve has identified the following questions it hopes to answer with future research:

1. What is the impact of UPchieve math tutoring when enrolling exclusively new students who find UPchieve online and a sample size large enough (1,000+ students) to detect small changes in math assessment scores?

2. What is the dosage threshold beyond which additional math achievement and motivational gains start to diminish?

3. Do incentives or encouragements affect usage among students at schools that partner with UPchieve to provide tutoring? How does this compare to the increase in usage that treatment group students—who found UPchieve online—experienced in this study?

References

Bruch, J., Forde, J., Lai, I., & Sotelo Muñoz, V. (2022). *Menu of measures: High-quality measures of Middle Years Math student outcomes.* Mathematica. https://www.mathematica.org/-/media/publications/pdfs/education/2022/menu-of-measures_math_march-2022.pdf

Kosovich, J., Hulleman, C., Kenneth, B., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the Expectance-Value-Cost Scale in middle school. *Journal of Early Adolescence, 35*(5-6), 790–816. https://doi.org/10.1177/0272431614556890

Kraft, M., 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://pubs.aeaweb.org/doi/pdfplus/10.1257/pandp.20221038

Midgley, C., Maehr, M., Hruda, L., Anderman, A., Anderman, L., Freeman, K., Gheen, M., Kaplan, A., Kumar, R., Middleton, M., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales.* The University of Michigan. http://websites.umich.edu/~pals/PALS%202000_V13Word97.pdf

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.* National Bureau of Economic Research. https://www.nber.org/papers/w27476

Renaissance Learning, Inc. (2021). *Star Assessments for Math technical manual.* https://help2.renaissance.com/US/PDF/SM/SMRPTechnicalManual.pdf

Roschelle, J., Cheng, B. H., Hodkowski, N., Neisler, J., & Haldar, L. (2020). *Evaluation of an online tutoring program in elementary mathematics.* Digital Promise. https://files.eric.ed.gov/fulltext/ED604743.pdf

Wang, M., Fredricks, J., Ye, F., Hofkens, T., & Linn, J. (2016). The Math and Science Engagement Scales: Scale development, validation, and psychometric properties. *Learning and Instruction, 43,* 16–26. https://doi.org/10.1016/j.learninstruc.2016.01.008

What Works Clearinghouse. (2022). What Works Clearinghouse standards handbook, version 5.0. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/handbooks

Appendix A. Detailed description of tutoring program

UPchieve uses the combination of on-demand support, around-the-clock availability, a one-on-one student–tutor ratio, and a platform designed for equitable access and collaboration to achieve its outcomes. These components aim to improve student outcomes by centering the needs of students in priority communities and enabling them to access the benefits of one-on-one tutoring more readily. UPchieve does this by meeting students at home or in school, during the day or in the evening, and via chat and virtual whiteboard so that students can take advantage of extra support. Although UPchieve currently offers tutoring in math, English language arts, science, and social studies, this study focused on the impact of UPchieve's math tutoring on students' math outcomes.

UPchieve conducts a variety of activities that contribute to its theory of change, including the following:

- **Maintaining a trained volunteer tutor force**. This includes recruiting and training volunteers and providing quality checks for tutors who are performing poorly. UPchieve provides an introductory course to volunteers that teaches the basics of tutoring on UPchieve, which includes content on diversity, equity, and inclusion; academic integrity; UPchieve safety policies; tutoring best practices; and reporting inappropriate student behavior. UPchieve also provides volunteers with subject-matter materials to review. To be certified as an UPchieve tutor, volunteers must pass an assessment on the content taught in the introductory course and a separate assessment that asks volunteers to demonstrate their knowledge of the subjects they signed up to tutor. Additionally, students provide feedback on their tutors at the end of sessions. If the feedback is poor, an UPchieve community manager is notified and contacts the tutor with suggestions for improvement.

- **Deploying a tutoring platform**. This platform helps match students with available tutors and online tutoring through a collaborative whiteboard or document editor and chat. Students are matched with tutors who are available based on the content and grade level they select.

- **Providing on-demand, one-on-one tutoring on student-requested topics**. Students select the amount of tutoring they receive based on their learning needs and availability.

- **Recruiting students.** UPchieve recruits and retains student participants through online marketing and school partnerships. The majority of UPchieve's students find UPchieve online by searching for phrases like "free online math tutoring."

- **Supporting engagement in their partner schools**. UPchieve supports educators to promote its program with students, offers professional development to educators, and provides outreach and implementation materials to schools.

- **Collecting data to support schools and platform improvement.** UPchieve collects data on students' tutoring topics, the length and frequency of their tutoring sessions, and feedback on tutors to inform platform and tutor development. When students are associated with partner schools, these tutoring usage data are shared with school administrators and teachers to support classroom instruction.

The number of UPchieve users has grown significantly over the past three years. UPchieve provided tutoring to 200 students in fall 2019 and 19,000 students in fall 2022.

Appendix B. Methods

Experimental design. We used a stratified randomized controlled trial to compare the impact of receiving UPchieve tutoring on student math knowledge and math motivation, confidence, and engagement between students in the treatment and control groups. Over 13 weeks from November 2021 to January 2022, we recruited students in grades 9 and 10 who found UPchieve online[2] and qualified to use the tutoring platform (attended a Title I school or live in a zip code with an average household income of less than $60,000). Each week, a new cohort of students who completed a baseline survey and assessment were randomly assigned to the treatment or control group. The study period lasted 14 weeks for each cohort. We stratified our analysis by the cohorts to account for differences in student experiences with UPchieve given the timing of when they joined the study. Students in the treatment group received weekly encouragements and a $5 financial incentive for each study week in which they completed at least one UPchieve math tutoring session. The control group students did not receive encouragements or incentives but still had access to UPchieve tutoring. We found that students in the treatment group used UPchieve math tutoring 9 times on average, while students in the control group used UPchieve math tutoring 3.5 times on average. We then estimated the average impact of the additional 5.5 math tutoring sessions on students' math knowledge and perceptions about math. Treatment group students received weekly encouragements in addition to a $5 incentive for each study week in which they completed at least one UPchieve math tutoring session. Encouragements to treatment group students were sent to students' emails associated with their UPchieve accounts. Below, we provide an example of the encouragement that UPchieve sent treatment group students.

> *Subject: Our Coaches Can Help You with Math (in just 10 minutes!)*
>
> *Hey NAME,*
>
> *UPchieve Academic Coaches are trained, patient, and kind volunteers here to help you when you need it most. You can request help at 6am (before your first math class) or 11:37pm (when you're stuck on that homework problem and your bestie is asleep).*
>
> *Because you're part of our study you can also receive a $5 gift card for receiving 10 minutes of math help on UPchieve this week.*
>
> *Let's work together to level up your math skills and add to your savings.*
>
> *As always, reach out to me if you have any questions.*
>
> *Happy Learning,*

---

[2] Students who found UPchieve online include all students who are not affiliated with UPchieve partner schools, where UPchieve works to implement its tutoring program across students in the same class, grade, or school. UPchieve reports that students who do not attend UPchieve partner schools tend to find the platform through an online search engine.

Measures. We collected several data sources before and after the study period to measure the impact of the additional tutoring: the Renaissance Star Math assessment of math knowledge (Star Math), the Expectancy-Value-Cost Scale (EVC) to measure motivation in math, the Patterns of Adaptive Learning Scale (PALS) to measure confidence in math, and the Math Engagement Scales to measure engagement in math. Details on each of the survey measures are available in Mathematica's menu of high-quality middle years math student outcome measures, which were selected in consultation with external measurement experts (Bruch et al., 2022). Students participating in the study took the Star Math assessment and an online survey measuring motivation, confidence, and engagement in math two weeks before the start of the study and again during the month after the 14-week study period ended. Students in the treatment and control groups continued to have access to UPchieve after the study period, but treatment group students no longer received email encouragements or incentives. The Star Math assessment is a widely used assessment that has a reliability estimate of 0.93 for students in grade 9 and 0.94 for students in grade 10 (Renaissance Learning, Inc., 2021). To assess the reliability of survey measures, we calculated the Cronbach's alpha for each survey measure, and all measures exhibited relatively high values substantially above the U.S. Department of Education's What Works Clearinghouse acceptable threshold of 0.60 (What Works Clearinghouse, 2022). This study's reliability coefficients (alpha) for the EVC before the study were 0.91 for expectancy, 0.87 for value, and 0.80 for cost. After the study, they were 0.94 for expectancy, 0.92 for value, and 0.89 for cost. The reliability coefficient (alpha) for PALS was 0.85 before the study and 0.87 after the study, and the reliability coefficient (alpha) for the Math Engagement Scales both before and after the study was 0.90.

Analysis. To analyze students' Star Math assessment scores and perceptions of math, we used a Bayesian linear regression model facilitated by Mathematica's e2i Coach tool. In this model, we controlled for students' baseline Star Math scores, students' math motivation, their cohort, and a series of demographic and UPchieve activity indicators. This study included 99 students who were new or existing users of UPchieve in 9th or 10th grade. Of the 99 students who entered the study, 89 students completed the post-study Star assessment. This included 45 of the 52 students in the treatment group and 44 of the 47 students in the control group. Ninety-one students had complete survey data measuring math motivation, confidence, and engagement, comprising 47 students in the treatment group and 44 students in the control group. Students' survey responses were invalidated and counted as missing if a student answered the same response option for the entire survey. There were 2 students whose survey responses were invalidated at follow-up.

Baseline equivalence was checked for UPchieve activity that could bias results, including the number of accounts that shared an IP address with another account (a sign of potential fraud or students signing up from the same location); the number of students who took the Star assessment in less than four minutes; and the number of UPchieve tutoring sessions that students logged before the study began. Of 99 student accounts, 28 were identified as sharing an IP address with another account.[3] Most of these students had a matching IP address with one other account. Once this concern was identified, UPchieve began contacting accounts who shared IP addresses with other students in the study to verify their identity before they completed baseline assessments.

There were more students in the treatment group identified as having shared IP addresses (standardized difference of 0.27 standard deviations [SDs]), more students in the treatment group identified for taking the Star Math assessment in less than four minutes (0.31 SDs), and no significant difference between groups in the number of matched UPchieve sessions in math or any

---

[3] Having a shared IP address with another student account is possible, for example, when more than one student signs up for the study at the same school. We decided to keep shared IP address as a covariate in this study's analyses out of an abundance of caution.

other subject before students joined the study (0.09 SDs). Additionally, no differences above 0.25 SDs were detected between groups in baseline Star Math scores (–0.21 SDs) and baseline EVC scores measuring math motivation (–0.04, –0.22, and –0.03 SDs, respectively), and baseline equivalence differences by gender, race, and cohort were reasonable given the small sample size of this study and the number of baseline equivalence measures tested. Our analysis included these variables as controls in the regression model. See Table B.1 for the full results of the baseline equivalence testing. When presenting analysis results in Figures 3 through 6 in the main report, the estimates displayed in each figure adjust for differences between treatment and control baseline values of the outcome presented. For example, the treatment group estimate in Figure 3 is calculated as the control group mean score on the post-tutoring Star Math assessment plus the estimated treatment effect (which controls for treatment-control differences in baseline assessment scores), so that the graphical display aligns with the estimated treatment effect.

Last, we invited teachers, staff, and students who participated in this study to provide their interpretation of the preliminary findings from the study in guided focus groups. Of the five teachers and staff who participated in a focus group, three had previously implemented UPchieve at their schools, and the other two were familiar with UPchieve but had not yet shared it with their students. During the 60-minute focus groups, we presented analysis results on the level of UPchieve tutoring usage, student survey results, and student Star Math scores across treatment and control groups. As we reviewed this information, we asked the participants about the trends they noticed and their suspected causes. The goal of these focus groups was to uncover insights about the results that were unique to these groups of interest and to inform which findings were most valuable to highlight in the final report.

Table B.1. Baseline equivalence of the evaluation sample

Characteristic	Overall	Treatment	Control	Difference	Standardized difference (SDs)
Shared IP address	0.31	0.38	0.25	0.13	0.27
Star test in < 4 min.	0.09	0.13	0.05	0.09	0.31
UPchieve sessions before study	11.16	12.6	9.68	2.92	0.09
Baseline Star Math NCE	45.47	42.74	48.26	-5.51	-0.21
Baseline EVC expectancy	4.14	4.11	4.17	-0.06	-0.04
Baseline EVC value	4.27	4.1	4.44	-0.34	-0.22
Baseline EVC cost	3.51	3.49	3.53	-0.04	-0.03
Current grade	9.46	9.47	9.45	0.01	0.02
Female students	0.71	0.76	0.66	0.1	0.21
Male students	0.18	0.13	0.23	-0.09	-0.24
Trans* students	0.01	0.02	0	0.02	0.21
Nonbinary students	0.08	0.09	0.07	0.02	0.08
Gender not provided	0.03	0.02	0.05	-0.02	-0.13
Black students	0.29	0.29	0.3	-0.01	-0.01
Hispanic students	0.09	0.09	0.09	0	-0.01
White students	0.2	0.16	0.25	-0.09	-0.23

Characteristic	Overall	Treatment	Control	Difference	Standardized difference (SDs)
Asian students	0.17	0.13	0.2	-0.07	-0.19
Race and ethnicity not provided	0.13	0.2	0.07	0.13	0.38
Cohort 1	0.02	0.02	0.02	0	0
Cohort 2	0.04	0.09	0	0.09	0.43
Cohort 3	0.09	0.09	0.09	0	-0.01
Cohort 4	0.16	0.13	0.18	-0.05	-0.13
Cohort 5	0.03	0.04	0.02	0.02	0.12
Cohort 6	0.17	0.22	0.11	0.11	0.29
Cohort 7	0.13	0.13	0.14	0	-0.01
Cohort 8	0.09	0.11	0.07	0.04	0.15
Cohort 9	0.07	0.04	0.09	-0.05	-0.18
Cohort 10	0.12	0.09	0.16	-0.07	-0.21
Cohort 11	0.03	0.02	0.05	-0.02	-0.13
Cohort 12	0.01	0	0.02	-0.02	-0.21
Cohort 13	0.03	0	0.05	-0.05	-0.30

Note: Except as noted, results show the percentages of students with the listed characteristic, the percentage-point differences in the prevalence of the characteristic between the treatment and comparison groups, and standardized difference. Rows corresponding with the EVC survey present the average response on a 6-point scale, and the current grade row reports the mean of students' grade. We considered each characteristic to be balanced between treatment and control groups if the standardized difference was less than 0.25 SDs as stated in the *What Works Clearinghouse Standards Handbook, Version 5.0*. Normal curve equivalent (NCE) scores are Star Math scores that have been scaled to be normally distributed. Trans* is used to indicate gender identities such as transsexual and transgender.