Research in Progress to Better Understand High-Impact Tutoring

NSSA 2023 Conference
High-Impact Tutoring: From Research to Sustainability

May 9, 2023
Stanford University

studentsupportaccelerator.org

PROGRAM EFFECTS

Note: Nickow, Oreopoulos and Quan (2020) provide an excellent meta-analysis of existing published research: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3644077

A Scalable Approach to High-Impact Tutoring for Young Readers: Results of a Randomized Controlled Trial

Kalena Cortes, Texas A&M University; Karen Kortecamp, George Washington University; Susanna Loeb, Stanford University; Carly D. Robinson, Stanford University

Initial analyses for year one of the program done but not finalized. Available draft: https://cdn2.assets-servd.host/ifl-site/production/assets/downloads/Scalable-Approach-to-High-Impact-Tutoring-02_02_23.pdf

Research Question: What is the effect of Chapter One's 1:1 tutoring on early literacy?

Methods: Randomized Controlled Trial

Locations: Broward County Public Schools, Florida

Data: Chapter One assessment and admin data from Broward County Public Schools.

Findings: We find that implementing this program in kindergarten can dramatically improve the reading ability of students at the end of the year. Almost 70% of students who received Chapter One tutoring reached the goal for kindergarten students, Reading Foundation Stage 4, by the end of the year. Comparatively, only 32% of students in the control group reached Stage 4. By reaching Stage 4, where the students can segment and blend CVC words, these students can "hit the ground running" in their reading instruction at the beginning of first grade. Students who enter first grade unable to decode CVC words may be at risk of failing to be fluent readers at the end of first grade.

Closing the income-achievement gap? Experimental evidence from high-dosage tutoring in Dutch primary education

Joppe de Ree, Erasmus University Rotterdam, The Netherlands; Mario A. Maggioni, DISEIS & CSCC Università Cattolica del Sacro Cuore, Milano, Italy; Bowen Paulle, University of Amsterdam, The Netherlands; Domenico Rossignoli, DISEIS & CSCC Università Cattolica del Sacro Cuore, Milano, Italy; Nienke Ruijs, Dutch Inspectorate of Education and VU University Amsterdam, The Netherlands; Dawid Walentek, University of Warsaw, Poland

Complete: https://www.sciencedirect.com/science/article/pii/S0272775723000304

Motivation: We set out to test what we see as the most promising, and potentially scalable approach to reducing income-achievement gaps (in math) in Dutch education.

Research Questions: We wanted to know if a Saga Education-consulted high dosage tutoring program could be successfully built from the ground up and exported to a different national/institutional setting while maintaining substantial effect sizes, and whether existing income-achievement gaps could be substantially reduced by targeting low-income communities with scalable interventions like HDT.

Methods: Randomized controlled trial

Location: A large city in the Netherlands

Data: Most importantly, high stakes standardized achievement tests

Findings: We find treatment effects on math scores of 0.28 national population standard deviations after one school year. These effects can account for 40% of the math achievement gap between low-income and high-income students in the Netherlands.

Implications: As most of the evidence on intensive tutoring programs draws on research from the United States, we conclude that (i.) HDT programs can be successfully built from the ground up and exported to different institutional settings while maintaining substantial effect sizes, and, (ii.) existing income-achievement gaps can be substantially reduced by targeting low-income communities with scalable interventions like HDT.

High dosage tutoring in pre-vocational secondary education: Experimental evidence from Amsterdam

Joppe de Ree, Erasmus University Rotterdam, The Netherlands; Mario A. Maggioni, DISEIS & CSCC Università Cattolica del Sacro Cuore, Milano, Italy; Bowen Paulle, University of Amsterdam, The Netherlands; Domenico Rossignoli, DISEIS & CSCC Università Cattolica del Sacro Cuore, Milano, Italy; Dawid Walentek, University of Warsaw, Poland

Working paper available: https://osf.io/preprints/socarxiv/r56um/

Motivation: This was the first Saga Education-consulted high dosage math tutoring intervention ever carried out, outside the US, at the secondary level. This was also the first effort implemented by The Bridge Learning Interventions, a Dutch non-profit that continues to partner with Saga Education.

Research Questions: We set out to examine whether The Bridge's HDT program could appreciably reduce the gap between, on one hand, what low-achieving students attending stigmatised high-poverty secondary schools need to thrive and, on the other, the educational opportunities that are typically afforded to them.

Methods: Randomized controlled trial

Location: One of Amsterdam's poorest neighborhoods (New West)

Data: Standardized achievement tests

Findings: We document large treatment effects of 0.44 and 0.72 control group standard deviations on a verbal and a nonverbal standardized math test respectively, after 16 weeks of personalized 2-on-1 tutoring..

Implications: Our findings demonstrate considerable unrealised potential among students assigned to the lowest tracks of the Dutch secondary education system. The HDT program can appreciably reduce the gap between what these low-achieving students need to thrive, and the educational opportunities that are typically afforded to them.

Road to Recovery

Dan Goldhaber, AIR; Thomas J. Kane, Harvard University; Andrew McEachin, NWEA; Emily Morton, NWEA; Maria V. Carbonari, Harvard; Miles Davison, NWEA; Michael DeArmond, AIR; Daniel Dewey, Harvard; Elise Dizon-Ross, AIR; Ayesha Hashim, NWEA; Tyler Patterson, Harvard; Douglas O. Staiger, Dartmouth

Working paper available:
https://caldercenter.org/publications/challenges-implementing-academic-covid-recovery-interventions-evidence-road-recovery

Motivation: This study is part of the larger Road to Recovery project, supporting school districts' academic recovery efforts by providing analysis of the participation, dosage, and impacts of their interventions.

Research Questions: We examine participation, dosage, and impacts of academic recovery interventions in 12 mid- to large-sized school districts across 10 states during the 2021–22 school year. We also examine overall school year 2021-22 recovery rates in these districts and discuss implementation challenges that emerged from interviews with district leaders.

Methods: Value-added modeling

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

Data: NWEA MAP Growth assessment data and district-provided student-level demographic data, enrollment data, and intervention eligibility and participation/attendance data.

Findings: Our findings highlight the challenges that recovery efforts faced during the 2021–22 school year. Although, on average, math and reading test score gains during the school year reached the pace of pre-pandemic school years, they were not accelerated beyond that pace. This is not surprising given that we found that districts struggled to implement recovery programs at the scale they had planned. In the districts where we had detailed data on student participation in academic interventions, we found that recovery efforts often fell short of original expectations for program scale, intensity of treatment, and impact. Interviews with a subsample of district leaders revealed several implementation challenges, including difficulty engaging targeted students consistently across schools, issues with staffing and limitations to staff

capacity, challenges with scheduling, and limited engagement of parents as partners in recovery initiatives.

Implications: The implementation challenges district leaders recounted suggest that the simple-sounding logic of academic intervention—identify students in need and provide them extra support—belies a host of complex design and implementation decisions. Under the existing capacity constraints of district staff and the time constraints of typical school days and years, there are no easy solutions to districts' challenges with staffing and scheduling interventions for students. Providing sufficient intervention for all students in need is going to require historic action, beginning with a renewed effort to engage families. States and districts can help by providing transparent and accessible measures of students' academic progress and recovery to families and students. Furthermore, ESSER funds are unlikely to be sufficient for the larger share of districts that spent more time in remote status or serve low percentages of high-poverty students (so received less ESSER dollars). In addition to funding, our findings suggest that districts may need to invest in central office capacity and internal administrative systems (e.g., data systems, hiring procedures) to implement academic recovery interventions at scale. Finally, districts will need help to expand their interventions to be commensurate with their students' losses. In most cases, expansion will mean expanding student participation and dosage in existing programs, as well as layering interventions (e.g., high dosage tutoring and an extended school year) for targeted students.

Breakthrough Collaborative's Tutoring Program: Math Knowledge Gains and Participant Math Perceptions

Breyon Williams, Connor Rooney, and Greg Chojnacki, Mathematica

Complete: https://www.mathematica.org/publications/math-knowledge-gains-and-math-perceptions-among-participants-of-the-breakthrough-collaborative

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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, and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

Research Questions:
1. Among students who participate in the tutoring program, what is the average attendance rate during tutoring sessions?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutor after receiving tutoring?
3. Do students who participate in the tutoring program report high levels of sense of belonging after receiving tutoring?
4. Do students who participate in the tutoring program report higher levels of math confidence after receiving tutoring than before the tutoring began?

5. Do students who participate in the tutoring program demonstrate gains on the district administered i-Ready assessment (Breakthrough Central Texas) or Renaissance Star Math assessment (Breakthrough Greater Boston)?

Methods: This study uses a descriptive approach.

Location: Central Texas and Greater Boston

Data: Data used for this study includes attendance records, data on survey and math knowledge, and qualitative data from focus groups.

Findings: Usage: Breakthrough experienced attendance challenges, although bright spots emerged. Anecdotal evidence from affiliates suggests that the COVID-19 pandemic and the competing after-school commitments of students contributed to attendance challenges. Breakthrough will continue to learn from its attendance bright spots and challenges to inform program design and implementation. Student perceptions and beliefs: Survey responses suggest that students held favorable views about their relationships with their math tutors and about their sense of belonging in tutoring sessions, on average, at the end of programming. Survey responses also provide descriptive evidence that math confidence increased from baseline to the end of programming, although that change was not statistically significant in the small sample of survey respondents. Math knowledge: Participants' math scores grew during the year, on average. This growth was similar to the growth from a national sample during a pre-pandemic school year. Given that the study design did not include a comparison group of similar students who did not participate in the tutoring program, it is not possible to determine conclusively the extent to which Breakthrough's tutoring program contributed to the observed growth.

Implications: Given the prominence of attendance challenges in Breakthrough tutoring this past school year, Breakthrough believes that future research on strategies to improve attendance would benefit the field.

Blueprint Math Fellows Tutoring Program: Math Knowledge Impacts and Participant Math Perceptions

Andrew Gothro and Greg Chojnacki, Mathematica

Complete: https://www.mathematica.org/publications/math-knowledge-impacts-and-math-perceptions-among-participants-in-the-blueprint-math-fellows

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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 (those who are Black, Latino, and/or experiencing poverty), and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

Research Questions:

1. Among students identified to receive the tutoring program, what is the average attendance rate? Does attendance vary by size of the homebase tutoring group?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutors? Does the reported quality of students' relationships vary by size of the homebase tutoring group?
3. Do students who participate in the tutoring program score higher on district assessments than students who do not participate in the tutoring program? Does this difference vary by student characteristics or size of the homebase tutoring group?
4. Do students who participate in the tutoring program report higher levels of math confidence and sense of belonging after receiving tutoring than before? Does this vary by size of the homebase tutoring group?

Methods: The study team used descriptive methods to examine student attendance in tutoring sessions, student–tutor relationships, and student enjoyment of math. We used a Bayesian regression-adjusted comparison group design to compare math performance of students who participated in tutoring with that of students in the same district who did not participate in tutoring (Quasi-Experimental Design). To complement the main analyses, we also used descriptive methods to examine the relationships between size of the homebase tutoring group, student attendance, and student math performance.

Data: Blueprint staff collected student survey data at the beginning (fall) and end (spring) of the school year from students receiving tutoring. Blueprint used the Copilot-Elevate tutor caring subscale (student–tutor relationship construct), Copilot-Elevate classroom belonging subscale (sense of belonging construct), Patterns of Adaptive Learning scale (PALS) (math confidence construct), and Math and Me (math confidence construct) to measure student perceptions about the tutoring and their experience of math. For pre-post analyses of student math confidence and sense of belonging in tutoring, we used data from students who completed both surveys. For post-only descriptive analyses of student–tutor relationships, we used the full set of spring survey data and thus drew on a larger sample than the pre-post analyses. The school district administered the Star Math assessment to Blueprint and comparison group students at the beginning and end of the school year and provided the data to the study team.

Findings: Attendance in tutoring: Student attendance in tutoring sessions averaged 81.5 percent during the school year. Attendance varied only slightly across students in homebase groups of different sizes. Student-tutor relationships: Most students (76.8 percent) reported strong relationships with their tutors. We considered reported relationships to be strong when students' responses averaged 4 or higher on a 5-point scale. A much higher share of students reported strong relationships in Site 2 (92.3 percent) than Site 1 (55.3 percent). A smaller share of students ending the year in large homebase groups (70.3 percent) reported strong relationships with their tutors, compared to students ending in small groups (77.3 percent) or medium-sized groups (83.9 percent). Math knowledge: End-of-year Star Math assessment scores for Blueprint participants were 0.12 standard deviations higher than those of their nonparticipant peers, after controlling for differences in baseline scores and characteristics. Based on this estimate, there is a 99 percent probability that the program improved student scores. Results varied by site, with a larger effect in Site 1 (0.14 standard deviations) than in Site 2 (0.09 standard deviations), and they also varied

by student characteristics. Descriptively, students in large homebase groups showed greater improvement in Star Math scores than those in either medium or small groups. Student math confidence and sense of belonging: Student confidence in math and sense of belonging in tutoring increased modestly during the school year. Responses varied across sites, however, with reported confidence and sense of belonging in tutoring decreasing slightly for students in Site 1 and increasing for students in Site 2. Students ending the year in medium-sized homebase groups reported a meaningful improvement in sense of belonging in tutoring (an increase of 0.8 points on a 5-point scale), compared to no substantial change for students in other group sizes.

Implications: This study provides further evidence of the potential benefits of tutoring on student math knowledge, along with suggestive evidence on how program variations might affect outcomes. In particular, the results suggest that larger tutoring groups—as high as eight or 10 students—might still lead to positive math knowledge gains, at least in an in-school tutoring model implemented by a seasoned tutoring organization. Results suggest smaller groups might be more conducive to forming strong student–tutor relationships, but those results did not correlate with larger improvements in math knowledge scores in the Blueprint tutoring program.

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

Grady Deacon and Greg Chojnacki, Mathematica

Complete: https://www.mathematica.org/publications/impacts-of-upchieve-on-demand-tutoring-on-students-math-knowledge-and-perceptions

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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 (those who are Black, Latino, and/or experiencing poverty), and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

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?

Methods: 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

online2 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.

Data: 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.

Findings: 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.

Implications: 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.

Math Corps' Tutoring Program: Math Knowledge Impacts and Participant Math Perceptions

Silvia Robles, Krista O'Connell, Andrew Gothro, and Kate Place, Mathematica

Complete: https://www.mathematica.org/publications/math-knowledge-impacts-and-math-perceptions-among-participants-in-math-corps-tutoring-program

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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 (those who are Black, Latino, and/or experiencing poverty), and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

Research Questions:
1. Among students identified to participate in Math Corps, what is the average attendance rate? Does attendance vary based on student characteristics?
2. Do students who participate in Math Corps score higher on the Star math knowledge assessment than similar students who do not participate in Math Corps?
3. Do students who participate in Math Corps report having a high-quality relationship with their tutors?
4. Do students who participate in Math Corps report high levels of sense of belonging? Among students participating in Math Corps, are sense of belonging levels correlated with reported quality of their relationships with their tutors?
5. Do students who participate in Math Corps report high levels of math confidence? Among students participating in the program, are math confidence levels correlated with reported quality of their relationships with their tutors?

Methods: The study is a quasi-experimental, mixed methods analysis of the Math Corps program that uses data collected in the 2021–2022 school year at three majority-Black traditional public and charter schools across Georgia and Minnesota. Teachers selected students to participate in Math Corps using students' baseline math assessment scores and other criteria. We estimate the effect of Math Corps on students' math knowledge by comparing students who participated in Math Corps to those with similar baseline assessment scores who did not participate in Math Corps (a regression-adjusted matched comparison design) and report estimates adjusted using Bayesian methods. The report also describes average student attendance and the results of a survey on Math Corps students' perceptions of their relationship with their tutors, sense of belonging in the program, and math confidence. We supplement these with findings from focus group and interviews with participants and tutors.

Location: Three majority-Black traditional public and charter schools across Georgia and Minnesota.

Data: The study collected several data sources before, during, and after the study period to measure outcomes of Math Corps: the Renaissance Star math assessment, a survey of student perceptions and beliefs, student attendance data, and focus groups with tutors and participants. The student survey drew from several existing scales: the relationship scale came from a Search Institute–developed relationships instrument, the math confidence scale was drawn from the

Patterns of Adaptive Learning Scale, and the belonging scale was drawn from the Copilot-Elevate survey. Only students who participated in Math Corps completed a survey. We present end-of-year data for student–tutor relationship quality and sense of belonging in Math Corps and growth from midyear to the end of the year for math confidence.

Findings: Attendance: Average student attendance was 91 percent throughout the year. Attendance varied little by gender. Math knowledge: Students who participated in Math Corps demonstrated moderate growth in math knowledge compared to similar students who did not participate in Math Corps, with a 98 percent probability of a positive effect, based on a Bayesian analysis. The estimated effect was 0.18 standard deviations, the equivalent of a student gaining approximately 7 percentile points on the spring math test. This is consistent with effects observed in prior research on the effect of Math Corps (Parker et al., 2019; Codding et al., 2022). Student perceptions and beliefs: Survey results showed that 94 percent of Math Corps participants had a high sense of belonging in the tutoring program and a positive view of their relationship with their tutor at the end of the year. There was a strong correlation between students' sense of belonging and the reported strength of their relationship with their tutor (r = 0.65). Additionally, students' math confidence grew by 27 percentage points from the middle to the end of the year. Thematic analysis of focus groups and interviews with tutors and students support these findings, as tutors attended to students' emotional needs, were patient and engaging, and had high expectations for students in math and life.

Implications: Unlike previous studies that have focused on majority-White populations, this study focuses on the implementation and success of Math Corps in majority-Black schools in areas experiencing poverty. The findings from this yearlong evaluation contribute to a growing body of research highlighting the positive impact Math Corps has on students' math achievement. Overall, findings from this study will help drive Math Corps' expansion to new states and embed more intentional relationship-building strategies into the program.

Cignition Group Tutoring: Impacts on Students' Math Knowledge and Perceptions

Catherine Pratt, Greg Chojnacki, and Kara Conroy, Mathematica

Complete: https://www.mathematica.org/publications/impacts-of-cignition-group-tutoring-on-students-math-knowledge-and-perceptions

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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 (those who are Black, Latino, and/or experiencing poverty), and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

Research Questions:

1. Do students who participate in group tutoring score higher on a fractions math assessment than students who do not participate in group tutoring?
2. Do students who participate in group tutoring report higher levels of math confidence and enjoyment than students who do not participate in group tutoring?
3. How do students receiving group tutoring rate the quality of the relationship with their tutors? Does the reported quality of students' relationships vary by student characteristics, or by tutor?

Methods: The study used a randomized controlled trial design, comparing the fractions knowledge of students who were randomly selected to participate in Cignition group tutoring to that of similar students who did not participate in tutoring. We analyzed survey data to understand Cignition students' confidence in math and enjoyment of math, relative to similar peers who did not participate in tutoring. We also analyzed survey data from tutoring participants to understand students' relationships with their tutors.

Data: The study collected survey and math assessment data before and after the study period to measure outcomes of the group tutoring: a Cignition fractions assessment to measure math knowledge, the Patterns of Adaptive Learning Scale (PALS) to measure confidence in math, the Math and Me survey to measure enjoyment of math, and the Copilot-Elevate survey to measure the student–tutor relationship. We examined the impact of participation in Cignition's group tutoring using a Bayesian linear regression model, which controlled for students' scores on the baseline fractions assessment, school, and a series of demographic indicators. Students in the treatment group completed a midpoint survey measuring the student–tutor relationship; the survey was based on the Copilot-Elevate teacher caring subscale.

Findings: Math knowledge: Students in the treatment group scored 9 percentage points higher on the Cignition fractions assessment than control group students, after accounting for baseline differences. The 9 percentage point difference between tutoring participants and nonparticipants translates to an effect size of 0.44 standard deviations, substantially larger than the benchmark of 0.2 standard deviations that is often used to identify large effects. A Bayesian analysis indicated with 99 percent confidence that Cignition tutoring had a positive effect on student learning. This effect size is similar to that of a 2020 efficacy study of Cignition's 1:1 tutoring offering (Roschelle et al., 2020). Student perceptions and beliefs: Students in the treatment group experienced higher levels of math confidence and enjoyment than students in the control group. Though not statistically significant, confidence in math among treatment group students exceeded that of students in the control group by 0.05 points on a 5-point scale, and enjoyment of math exceeded that of students in the control group by 0.07 points on a 5-point scale. Students who participated in Cignition group tutoring rated their relationship with their tutor at 4.25 on a 5-point scale, on average, indicating agreement or strong agreement that their relationship with their tutor was positive.

Implications: The findings from this study provide encouraging evidence that the 4:1 math tutoring increases students' math knowledge. Cignition's group tutoring had a large impact on students' performance on a post-test of fractions knowledge, which suggests that 4:1 tutoring is similarly as effective as Cignition's 1:1 offering (Roschelle et al., 2020). The estimated effect on fractions knowledge, 0.44 standard deviations, compares favorably to recent estimates in an

emerging body of evidence documenting the promise of individual and 2:1 online tutoring (see, for example, Carlana and La Ferrara [2021] and Gortazar et al. [2022] for early evidence on individual and 2:1 online tutoring, respectively).

Air Tutors' Online Tutoring: Math Knowledge Impacts and Participant Math Perceptions

Lily Fesler, Anna Gu, and Greg Chojnacki, Mathematica

Complete: https://www.mathematica.org/publications/math-knowledge-impacts-and-math-perceptions-among-participants-in-air-tutors-online-tutoring

Motivation: This study is part of a series of six reports on math tutoring programs funded by the Gates Foundation 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 (those who are Black, Latino, and/or experiencing poverty), and to learn the extent to which tutoring programs resulted in positive experiences for participating students. 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 as possible.

Research Questions:
1. What effects does Air Tutors have on participating students' math knowledge, as measured by performance on standardized math tests?
2. Do Air Tutors students report higher levels of math confidence after receiving tutoring than before?
3. Do Air Tutors students report a high sense of belonging in their tutoring sessions?
4. Do Air Tutors students report positive student–tutor relationships?
5. What is the average attendance rate for participating Air Tutors students

Methods: This study used a randomized controlled trial (RCT) to analyze the impacts of Air Tutors on math knowledge outcomes, pre-post analysis to measure changes in student math confidence, and descriptive analysis (with no comparison group) to measure attendance, student–tutor relationships, and the context in which Air Tutors implemented the solution.

Location: Three schools, including a K-12 online school and an elementary school and middle school in a school district in Texas.

Data: We collected data from several sources before and after the study period to measure the impact of tutoring. From districts, we collected fall and spring scores on the MAP math assessment and student demographics. We collected survey data using the Patterns of Adaptive Learning Scale (PALS) and Math and Me to measure confidence in math and Copilot-Elevate to measure sense of belonging and student–tutor relationships (Adelson & McCoach, 2011; Midgley et al., 2000). Students participating in the study took the fall MAP test prior to the start of the study and the spring MAP test after the end of the tutoring program. MAP scores were standardized using grade-level national means and standard deviations provided by NWEA (2020). We also collected attendance data for tutoring sessions from Air Tutors.

Findings: Math knowledge: MAP scores among students participating in Air Tutors were 0.13 standard deviations higher than those among control group students, after accounting for baseline differences. This is equal to a 4.0 percentile point increase from students who did not participate in Air Tutors. A Bayesian analysis, which provides a probability that the intervention had an effect given the impact estimate and prior evidence on similar interventions, indicated that Air Tutors had a 93 percent chance of boosting math knowledge. Air Tutors increased participating students' MAP scores the most at a program held during the school day at an in-person middle school (0.26 standard deviations). The analysis also showed that Air Tutors had a 66 percent chance of increasing students' MAP scores at an online school where tutoring was held during the school day (with an estimate of 0.10 standard deviations) and a 75 percent chance of increasing students' MAP scores at an elementary school where tutoring was held as an after-school program (with an estimate of 0.08 standard deviations). Student survey: Students participating in Air Tutors reported highly positive relationships with their tutors, with 89 percent of District 1 students and 87 percent of District 2 students rating an average of at least 4 on a 5-point scale at the end of the year, representing agreement or strong agreement with statements about positive relationships. They also reported a high sense of belonging in their tutoring sessions, with 80 percent of students in District 1 and 69 percent of students in District 2 rating an average of at least 4 on a 5-point scale at end-of-year, representing agreement or strong agreement with statements about sense of belonging. Finally, student math confidence increased over the course of the program, with students from both districts reporting an average increase of 0.18 points on a 5-point scale from baseline to the end of the year. Tutoring session attendance: On average, students attended 72 percent of sessions. The attendance rate was 63 percent at the in-person elementary school, where the virtual tutoring was offered after school at the school's campus. The attendance rates were higher at the in-person middle school (78 percent) and the online school (79 percent), both of which offered virtual tutoring during the school day.

When the counterfactual is as important as the factual: Emerging lessons from a virtual tutoring field experiment

Emanuele Bardelli, Santa Rosa School District; Sara White, Carly Robinson, Leiah Groom-Thomas, Susanna Loeb, Stanford University

Initial results presented at AEFP, March 2023; Working paper expected summer 2023

Motivation: Early findings on tutoring implementation in the 2020-2021 school year found tutor recruitment to be a challenge for districts seeking to expand their programs (White et al., 2021). Virtual tutoring where a student meets with a tutor via video conference is one potential solution for expanding the tutor recruitment pool. While a large body of causal research finds in-person tutoring to have consistently large and positive effects on student test scores (Dietrichson et al., 2017; Salvin et al., 2011; Nickow et al., 2020), less is known about the efficacy of virtual programs.

Research Questions: What is the effect of being assigned to receive synchronous, small-group virtual math or reading tutoring on the end-of-year math and reading test scores of students grades 3-8?

Methods: Randomized Controlled Trial - Student-level assignment to treatment within school-grade-subject blocks.

Location: Suburban school district in Texas.

Data: Student-level administrative data, STARR and MAP test scores, tutoring enrollment and attendance records

Findings: Overall, we find that being assigned to receive the virtual tutoring program did not have a statistically significant effect on students' end-of-year (EoY) MAP and STAAR test scores, except for their ELA STAAR scores. In this case, we find that being assigned to virtual tutoring reduced student ELA STAAR scores. Students not assigned to virtual tutoring also received a range of small-group and computer-based interventions. We have suggestive evidence that schools that implemented the virtual tutoring program had more positive student outcomes, possibly because the virtual tutoring allowed for other small group interventions to take place.

Scale-up Experiments

Jonathan M.V. Davis, University of Oregon; Jonathan Guryan, Northwestern University; Kelly Hallberg, University of Chicago; Jens Ludwig, University of Chicago

New working paper coming in summer 2023
https://hechingerreport.org/wp-content/uploads/2023/04/High-Dosage-Tutoring-at-Scale.pdf

Motivation: We would like to know which social programs work at large scale before we scale them. But it would seem the only way to tell whether a program works at scale is to test it at scale. We ' a way out of this Catch-22 for one important scale-up challenge: heterogeneous inputs (like tutors). We show it is possible to operate a program at modest scale but learn about treatment effects at a much larger scale by randomly sampling the inputs the provider would have hired if they operated at the larger scale. We present evidence from a "scale-up experiment" we ran in Chicago where we randomly hired tutors, randomly assigned students to a high-dosage tutoring program or a control group, and randomly matched treatment group students with tutors.

Research Questions: By how much do the impacts of high-dosage tutoring programs change with the scale of the program because of heterogeneous tutor quality?

Methods: We use a randomized controlled trial with three randomizations. First, we randomly assigned students to be in the treatment group which was offered high-dosage tutoring and a control group which was offered status quo services. Second, we asked SAGA tutoring to recruit as though they had to hire three and a half times more tutors than they actually needed and then randomly selected 1 in 3.5 tutors to be given a job offer. Finally, we also randomly assigned pairs of students to tutors.

Location: Chicago and New York City

Data: The study uses administrative data from Chicago Public Schools and New York Public Schools and program data from SAGA tutoring.

Findings: Our main estimate indicates that every 100 rank decline in tutor quality decreases the impact of high-dosage tutoring on math test scores by 0.077 standard deviations (SD). This estimate is relatively imprecise with a 95 percent confidence interval ranging from -0.22 to 0.06. The point estimate is comparable in magnitude to the TOT estimate on math test scores, 0.081, and is about half of the average effect across all years of the study, 0.16 SD, reported in the meta-analysis above. We cannot, however, rule out that tutor quality is uncorrelated with Saga's initial ranking.

Implications: In this paper, we propose a new method - the scale-up experiment - that tries to propose a constructive scientific response to help understand the degree to which rising input prices are a barrier to successful scale-up, and how to avoid that problem. Rising input prices obviously do not account for every failure to scale, but we believe they are an important explanation for many cases since so many social programs hinge on hiring people to work with other people and people vary in their skill. Heterogeneity in worker skill means labor is an inelastically-supplied key input. Scale-up experiments reduce the costs of understanding how worker skill changes in program scale. Even in cases when these sorts of experiments are not possible, analysts could at the very least prioritize collecting program rankings of inputs at the hiring stage and report correlations between those rankings and input "value added" as a standard part of reporting on all social program RCTs. This simple framework helps re-orient our thinking away from the simple yes/no question ``Does this program scale?" to the ultimately much more constructive question ``What cost-per-participant is required to scale this program?"

Tutoring Online Project - TOP

Michela Carlana, Harvard University; Eliana La Ferrara, Harvard University

Working paper available, Updated version - summer 2023
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3777556

Study Type: Estimation of Program Effects

Motivation: The initial motivation was addressing learning losses. Currently, we are expanding tutoring programs to help disadvantaged children and to mitigate inequalities in education.

Research Questions:
- Is tutoring online effective?
- Group vs individual tutoring
- Can tutors help encode positive academic memories to remedial education children?
- How can we improve the effectiveness of the match?

Methods: RCT

Location: Italy and Dominican Republic

Data: Administrative and survey data

Findings: Positive effects on educational achievements of individual tutoring, no effect of group tutoring

Implications: Scaling up individual tutoring

Ravenswood Reads

Rebecca Silverman, Renee Scott, Rebecca Deutscher (Stanford University)

Motivation: Over the past two years, we have been conducting research related to early literacy tutoring in out of school time. In the first year, we studied online tutoring plus text messaging to families (during the pandemic) versus text messaging to families alone. In the second year, we studied in person versus remote tutoring and the role of text messaging to families via a 2X2 factorial design. We are now entering our third year working with the Ravenswood City Schools and the Boys and Girls Club of the Peninsula on a project investigating in person tutoring via high school students in which we will also look at the role of group size.

Methods: Experimental and Mixed Methods

Data: Videotaped observations, pre-test and post-test literacy assessment data, surveys
Findings: In year 1, we did not find pre-test to post-test differences in either condition and there was no effect of condition. In year 2, we did find pre-test to post-test differences in both conditions, but we did not find a differential effect for in person or remote tutoring or texting versus no texting with families.

Implications: Tutoring doesn't always lead to positive effects. The context needs to be considered. In person and remote tutoring may be equally effective. Text messaging in addition to tutoring may not have lead to increased effects.

Effects of Synchronous Virtual Early Literacy Tutoring and the Role of Group Size

Susanna Loeb, Carly Robinson, and Cynthia Pollard, Stanford University

In process - Results expected: August 2023

Motivation: This randomized controlled trial seeks to explore the effects of synchronous, small-group virtual tutoring in Kindergarten through 2nd grade early literacy on student academic performance and school attendance. Students assigned to tutoring will meet with a tutor virtually during the school day as part of a regularly scheduled intervention block. The tutoring provider requested that 1,816 students be identified by participating schools as qualifying for reading tutoring based on their beginning of year DIBELS assessments. The tutoring provider set enrollment quotas for each grade and subject. Stratifying by school and grade, we randomly assigned eligible students to a control group or to receive tutoring. Students assigned to tutoring were then randomly assigned to receive 1:1 or 2:1 tutoring.

Research Questions:
What is the effect of synchronous, virtual early literacy tutoring on end-of-year reading test scores?
What is the effect of 1:1 virtual early literacy tutoring on end-of-year reading test scores?
What is the effect of 2:1 virtual early literacy tutoring on students' end-of-year reading test scores?

Do students assigned to 1:1 early literacy tutoring perform better, the same, or worse compared to students assigned to 1:2 early literacy tutoring?

Methods: This is a randomized controlled trial.

Location: Elementary schools in Texas

Data: Student administrative data; Student beginning-, middle-, and end-of-year reading scores (DIBELS); Tutor demographic data.

Findings: Some preliminary analyses of mid-year DIBELS suggest that: Students assigned to tutoring improved more on their mid year reading scores than students who were assigned to the business as usual condition. The positive effect of tutoring was greater in the earliest grades, with kindergarteners improving more than first or second graders. The effects of tutoring varied across schools.

Implications: Results comparing the effectiveness of 1:1 versus 2:1 tutoring have important implications for the cost effectiveness and scalability of HIT. Additionally, findings could inform the design of effective HIT programs for early literacy and shed light on which students benefit most from synchronous, virtual early literacy HIT.

Khoaching with Khan Academy

Chloe Gibbs, University of Notre Dame; Michael Jensen, University of Notre Dame; Philip Oreopoulos, University of Toronto; Joseph Price, Brigham Young University

Initial analyses done but not finalized - results available in about 3 months.

Motivation: Teach teachers to use Computer Assisted Learning more effectively.

Research Questions: Does teaching teachers to use CAL leads to positive student outcomes?

Methods: RCT

Location: Arlington Independent School District

Data: State test scores

Findings: 0.15-0.25 z-score effects, on average. Higher effects for teachers that implemented with greater fidelity (e.g. at least 40 minutes of average classroom practice a week).

Implications: Computer Assisted Learning, if used effectively to increase personalized practice above 40 minutes a week, can provide baseline additional personalized learning at low cost.

National Student Support Accelerator: Effects of Tutoring Early Literacy Tutoring with School-Based Paraprofessionals (Pilot)

Monica Lee, Susanna Loeb, Carly Robinson, Nancy Waymack, Stanford University

In process - results available Summer/Fall 2023

Motivation: 1) Tutor supply is one of the biggest factors that affect the efficacy of tutoring programs. Hiring and retaining effective tutors who are also well aware of individual and school-specific needs is a key element to ensuring the success of high-impact tutoring. This program uses existing school staff (e.g., paraprofessionals) to serve as early literacy tutors.
2) There is insufficient evidence in the field to understand the cost effectiveness of early literacy interventions and whether different types of products have varying levels of efficacy.

Research Questions:
1. Does a high-impact, early literacy tutoring program implemented by existing school staff result in better student outcomes? Does the use of paraprofessionals in tutoring improve literacy outcomes at early grade levels?
2. Does the effectiveness of the program vary by the underlying student characteristics?

Methods: RCT. Within classrooms, eligible students were randomly assigned to tutoring or control.

Location: District of Columbia Public School (DCPS)

Data: Tutoring provider and DCPS administrative data

Implications: We also hope to strengthen the evidence base for districts as they build a case for continuing and refining student supports after pandemic-related federal funds have been exhausted. Research on the effectiveness of tutoring across contexts and on how to implement tutoring in the most resource-efficient way could inform these important decisions. This tutoring model has the potential to be extremely scalable and cost-effective.

Khan Academy as Tutor (KAT)

Philip Oreopoulos, University of Toronto

Just beginning - Results available in about 2 years

Motivation: Low-cost way of scaling personalized learning, combining computer assisted learning in the classroom with free complementary virtual tutoring at home

Research Questions: What is the effect of the program?

Methods: RCT

Location: Toronto District School Board

Data: Khan Academy

Findings: Pilot results look promising, with pre/post tutoring activity jumping since implementation.

Implications: The program offers a realistic scalable program for any school, math grades 3-8

Improving the Academic Performance of First-Grade Students with Reading and Math Difficulty

Lynn Fuchs, Douglas Fuchs, Marcia Barnes, Sonya Sterba, and Caitlyn Craddock, Vanderbilt University

Status: Ongoing; results expected June 2025

Motivation: This study addresses students with delays in word reading (WR) and mathematics calculations (MC) at start of first grade. Concurrent difficulty across WR and MC occurs frequently. Half of children with low performance in one domain experience low performance in the other, and concurrent WR and MC difficulty creates serious vulnerabilities. Students with concurrent difficulty also experience weaker outcomes in each domain and less adequate response to generally effective intervention compared to peers with difficulty in a single area. This population of learners is also underserved because complex school schedules and the costs of supplemental intervention mitigate against schools providing more than one intervention per student. The purpose of this randomized controlled trial is to test the efficacy and efficiency of Galaxy Star, an innovative tutoring program guided by commonalities between WR and MC development and designed to strengthen WR and MC in coordinated fashion with the same amount of tutoring time as single-focus tutoring.

Research Questions:
Does coordinated tutoring (Galaxy Star) improve learning with efficiency?
1. Does Galaxy Star produce stronger WR & CA outcomes compared to business-as-usual school instruction (including classroom instruction and school-provided supplemental services)?
2. Does Galaxy Star produce equivalent outcomes compared to WR tutoring even though Galaxy Star and WR tutoring allocate the same amount of tutoring time?
3. Does Galaxy Star produce equivalent outcomes compared to MC tutoring even though Galaxy Star and MC tutoring allocate the same amount of tutoring time?
Comparability is formally assessed via statistical equivalency (non-inferiority) testing.

Methods: Randomized controlled trial with four conditions. Students are identified with concurrent difficulty in WR and MC at start of first grade; then pretested; and then randomly assigned to the four conditions: Galaxy Star tutoring, WR tutoring, MC tutoring, and control. In each of the tutoring conditions, tutoring is provided individually for 15 weeks with three 30-minute sessions per week. Children are then posttested.

Location: Nashville TN

Data: Data include performance on cognitive, reading, and mathematics measures and reports of school-provided instruction and intervention.

Findings: Preliminary findings, based on 173 students who have completed the study, are in line with hypothesized effects. On reading and math outcomes, Galaxy Star (coordinated tutoring) is more effective than control. On reading outcomes, Galaxy Star is comparably effective as reading tutoring. On math outcomes, Galaxy Star is comparably effective as math tutoring. Conclusions await completion of the randomized controlled trial.

Implications: If findings hold through completion of this randomized controlled trial, the implications are that the complex needs of this group of learners may be served efficiently via

tutoring that is guided by commonalities between WR and MC development and designed to strengthen WR and MC in coordinated fashion with the same amount of tutoring time as single-focus tutoring. This is an important goal in the context of the scheduling challenges and intervention costs that mitigate against schools providing more than one intervention to the same student, leaving many of these students underserved.

Evaluation of the ASSISTments Tutoring Program

Mingyu Feng, WestEd

Just beginning

Motivation: Through a federal EIR grant, the ASSISTments Foundation is funded to pilot, and then eventually evaluate, a tutoring model that leverages staff within the district, in many cases novice tutors, to deliver high-impact standards-based math tutoring aligned to core instruction. The ASSISTments platform serves as the hub to support the instructional steps of tutoring, including diagnosing students' unfinished learning, engaging in practice, and monitoring student progress.

Research Questions: What's the effect of receiving high dosage, human tutoring via the ASSISTments program on the mathematics learning outcome of 5th and 6th grade students, compared to students in the business-as-usual condition? (note final grade levels are still being determined)

Methods: The study will use a student-level and delayed-treatment randomized controlled trial (RCT) design, to meet What Works Clearinghouse standards without reservations (WWC, 2020). The study will be conducted during one school year.

Location: Still finalizing research partners, as the study does not begin until the 2024-2025 school year.

Data: State test scores

Findings: N/A

Implications: The ASSISTments tutoring program is a pathway for districts to sustainably implement tutoring at scale. For both funding and human capital-related reasons, districts will increasingly need to leverage existing staff to serve as tutors. Often these tutors will be novice, with limited math and/or instructional experience. Districts will have limited capacity to manage these programs. The ASSISTments tutoring program provides a low-cost solution, and will have valuable lessons for other systems interested in taking a similar approach.

EFFECTS OF PROGRAM CHARACTERISTICS

Note: little prior research has tested the effects of program characteristics but Nickow, Oreopoulos and Quan, above, identify common characteristics of high-impact programs as does Robinson and Loeb (2021):
https://studentsupportaccelerator.org/sites/default/files/Accelerator_Research_Agenda.pdf

The inequity of opt-in educational resources and an intervention to increase equitable access

Carly D. Robinson, Stanford University; Biraj Bisht, University of California, Irvine; Susanna Loeb, Stanford University

Working paper available: https://edworkingpapers.org/index.php/ai22-654

Motivation: Billions of dollars are invested in opt-in, educational resources to accelerate students' learning. Although advertised to support struggling, marginalized students, there is no guarantee these students will opt in.

Research Questions:
We addressed two primary research questions:
1) To what extent do students take advantage of a free, on-demand tutoring resource provided by their school?
2) Can we increase take-up of on-demand tutoring with personalized communications to students and/or parents?
We also addressed two exploratory research questions:
3) Does take-up of on-demand tutoring lead to student learning gains?
4) What moderates the effect of personalized communications on take-up of on-demand tutoring and student learning gains?

Methods: Description of take up of an on-demand tutoring program and a randomized controlled trial (N=4,763) testing behaviorally-informed approaches to increase take-up. All middle and high school students enrolled during Fall 2020 were eligible for inclusion in the RCT sample. Students assigned to the control group received no additional communications beyond what their school typically sends. Students assigned to the Student Only or Student+Parent treatment arm received an initial mailer sent to their home address which included a letter that introduced the on-demand tutoring platform and provided information on how to access the platform and tips on how to engage . The letter was accompanied by post-it notes that we branded with the on-demand tutoring platform's logo and "Unlimited 1-on-1 Tutoring." They also received 16 personalized emails to their school-provided email address over the course of the semester. For those students assigned to the Parent Only or Student+Parent treatment
arms, one of their parents received a similar initial mailer and branded post-it notes. Parents, too, received 16 personalized text messages. Each communication was written to align with one of four

types of behavioral strategies: reminders, social norms, accountability, and valuing of the subject, which we randomized.

Location: Charter management organization in California.

Data: Data from the school system on whether students take-up the on-demand tutoring platform, as well as student course performance data.

Findings: The take up was low. At baseline, only 19% of students ever accessed the platform, and struggling students were far less likely to opt in than their more engaged and higher achieving peers. Communications to parents and students together increase the likelihood students access tutoring by 46%, which led to a four-percentage point decrease in course failures. Nonetheless, take-up remained low, showing concerns that opt-in resources can increase—instead of reduce—inequality are valid. We found no effects of the different behaviorally-focused messages.

Implications: Without targeted investments, opt-in educational resources are unlikely to reach many students who could benefit. Directing communications to students and their parents simultaneously had a large impact on whether students engaged with the program.

What motivates college students to serve as tutors? An RCT of a recruitment campaign

Carly D. Robinson, Stanford University; Katharine Meyer, Brookings Institute; Xiaoyang Ye, Amazon; Chasity Bailey-Fakhoury, Grand Valley State University; Susanna Loeb, Stanford University

In progress - working paper expected Summer 2023

Motivation: Recruiting tutors can be challenging. College students are a potentially large source of effective tutors, but we do not know what draws students to tutoring. College students may be driven by the prosocial benefits from working with kids and supporting the local community; by the monetary benefits of earning money; by the career benefits of developing professional skills and building resume; or by the social benefits of meeting other students, joining a community. We randomly vary the messages used to recruit college students to test the importance of these motivations.

Research Question:
1) Are students receiving a recruitment message with a targeted subject line more likely to open the email?
2) Are students receiving a targeted recruitment message more likely to click through to the application?
3) Are students receiving a targeted recruitment message more likely to apply to become a tutor?
4) Are students receiving a targeted recruitment message more likely to be hired as tutors?

Methods: Randomized Controlled Trial. We randomly vary email messages sent to the entire undergraduate student body of a large university.

Locations: Grand Valley State University, Michigan

Data: Email addresses from Grand Valley State University and application and selection data from the K-12 Connect tutoring program

Findings: We found no significant differences across messages for opening of emails but students in the monetary benefits group were approximately 200% more likely to apply and to be selected into the tutoring program. We found no differences between the control and any of the other message types. When students were asked their reasons for apply very few selected monetary benefits, indicating that self-reports are not a good indication of what drives students to apply and become tutors.

Tips4Tutors: A Text Messaging Curriculum For Tutors

Susanna Loeb, Carly Robinson, Kristine Gaffaney, and Cynthia Pollard, Stanford University; Sophie Barnes, Harvard University

In process - Results available August 2023

Motivation: Social-emotional skills are critical for students' quality of life and future success. This project will evaluate a program that provides supports for tutors, building from an understanding of both youth development and adult behavior change. The text-message based tutoring program, Tips4Tutors, seeks to promote effective instruction by sending light-touch, easily operational information over a semester through weekly text messages to tutors. Technology-based interventions provide the opportunity to reduce barriers and effectively deliver evidence-based social-emotional learning content with fidelity (Li et al., 2021).

Research Questions: Our primary research questions are:
(1) How does sending tutors facts and tips about cultivating social-emotional learning skills, differentiated by grade-level, affect tutors' reported behaviors, self-efficacy, and expectations for their students?
(2) How does sending tutors these facts and tips affect student social-emotional skills, academic, and behavioral outcomes?
(3) How do the effects of the program differ across tutors (high schoolers, college students, graduate students) and students (grade level, race/ethnicity, gender, and prior academic behaviors)?

Methods: We are conducting an RCT assessing the effects of Tips4Tutors that involves sending text messages to tutors.

Location: Guilford County, North Carolina, Washington DC

Data: Student administrative data (e.g., demographics, school attendance, tutoring attendance)
Student achievement data
Student perception data
Tutor demographics data
Tutor perception data

Findings: Still in data collection, but anecdotally (based upon some responses to text messages/emails) some tutors have found the tips helpful

Implications: Texting is an easy and inexpensive way to deliver important information to tutors at just the right time, and could increase positive educational behaviors at school and home. Both students and tutors could benefit from the curriculum's suggested skill- building activities, and that these behavioral changes translate into learning gains for children. Moreover, text- messaging programs are scalable, flexible and inexpensive; their content, frequency, and timing can be adapted to meet the needs of heterogeneous student and tutor populations. More broadly, text-messaging programs can be effective in reducing the inequality in educational outcomes, as under-resourced schools can implement text messaging interventions with fidelity.

Helping Tutors Working with High-Need Students: The Impact of Providing Tutors with a Community of Practice (Pilot)

Carly Robinson, Cynthia Pollard, Evan Bennett, and Susanna Loeb, Stanford University

In process - results expected Summer 2023

Motivation: While tutors are, in some ways, like teachers, effectively responding to the needs of a large and varied group of tutors presents at least three challenges that distinguish it from supporting teachers. First, because tutoring is delivered to individual students or small groups of students, schools will need a substantial number of tutors. While teachers share many pre-service requirements such as teacher education and certification, some tutors may be experienced educators while others have very little training or experience. Third, many tutors work part time or virtually, making it difficult for them to cultivate a professional community and learn from their colleagues. Unsurprisingly, the field of education has long recognized the importance of peer collaborations and support for educators' efficacy and growth (see Palincsar, 1999). Participation in high-quality CoPs or teams, for example, have been associated with improvements in instructional practices and student outcomes (Barton & Stepanek, 2012; Papay et al., 2020; Vescio et al., 2008). An online CoP for tutors may help to address tutoring-specific challenges and support tutor improvement. CoPs can provide tutors with an online space to find resources relevant to the district context, while also giving tutors opportunities to reflect on their experiences and crowdsource questions and best practices.

Research Questions:
1. Does having access to a professional community of practice (CoP) increase tutors' sense of support and social belonging?
2. Does having access to a professional CoP increase tutors' self-efficacy?
3. Does having access to a professional CoP increase tutors' interest in the teaching profession?
4. Do students of tutors in a community of practice

Methods: We conduct a randomized control trial (RCT) to test the effect of the digital, social-support intervention for tutors. While we are enthusiastic about the potential of a tutor CoP, we are currently piloting the implementation before launching a full-scale RCT. To do so, we have randomly assigned 135 tutors to one of two conditions, stratifying by the tutors' organization and school placement (1) Control: Tutors are asked to self-reflect about their tutoring experiences but do not have access to the open-forum digital platform and (2) Community of Practice: Tutors are

invited to participate in an open-forum, lightly moderated digital platform where tutors across the district can find curated resources and voluntarily share stories, experiences, or request advice.

Location: This study is being conducted in partnership with Gwinnett County Public Schools (GCPS), a large district in Georgia which plans to provide students with tutoring in 2022-2023.

Data: This study makes use of district data on student and tutor outcomes, as well as data from surveys administered by the district for the tutoring program and administrative purposes. The research team can use this survey data to tutors' sense of belonging and social connectedness, self-efficacy, occupational satisfaction and commitment, and their likelihood of pursuing a career in teaching.

Findings: At this time, we do not have findings to share as we are still in the process of collecting data.

Implications: This study has the potential to provide the field with a greater knowledge and understanding on how to best support tutors, as well as the impact of social support on engagement with a tutoring program, learning outcomes, and engagement with the tutoring program itself.

Effect of Tutor-Student Demographic Matching in High School Math Tutoring

Josh Bleiberg, University of Pittsburgh; Carly Robinson, Evan Bennett, Susanna Loeb; Stanford University

Working paper expected summer 2023

Motivation: Students' performance in Algebra 1 correlates with their future high school graduation rates and other long-term outcomes. We also know that female students are less likely to identify as "math people" or pursue STEM majors or careers in STEM. This randomized controlled trial seeks to explore the gender matching effect in a tutoring program on student beliefs, engagement, and academic outcomes in math in a large school district in New England.

Research Questions:
Do students who have a same-gender math tutor have more positive math-related beliefs, better attendance, and better academic outcomes than those who do not have a same-gender tutor?

Methods: RCT; All Algebra 1 students at 5 sites were eligible to receive tutoring, the majority of whom were 9th graders. Student groups were randomly assigned to tutors, stratified by student- and tutor-gender.

Location: Providence Public School District

Data: Student-level administrative data as well as data provided by the tutoring company

Implications: Findings from this study will provide some additional insight into how tutors can impact math beliefs and outcomes.

Embedding optional resources into the classroom: Testing a teacher-focused intervention to promote student usage of on-demand tutoring (Pilot)

Carly Robinson, Evan Bennett, Susanna Loeb; Stanford University

Pilot in data collection, results expected summer 2023.

Motivation: On-demand tutoring is a potentially beneficial resource for students, but we found most students never access help from a tutor. Even with targeted encouragements to parents and students, less than 30% of students used the resource. Classroom teachers are most

Research Questions:
1) To what extent do students engage with virtual, on-demand, opt-in tutoring services when teachers do not receive direct reminders about the resource?
2) Do students whose teachers have been assigned to receive emails promoting on-demand, virtual tutoring take-up the service more than students assigned to the control condition?

Methods: School-level RCT; 35 schools were randomly assigned to the treatment or control group. The control group continued on with business-as-usual, whereas the teachers in the treatment schools received a series of 4 emails giving information on how to embed on-demand tutoring into their classroom experience and encourage student usage.

Location: Multisite

Data: School-level data provided by the tutoring company

Implications: Many districts are investing in opt-in educational resources, and this will provide preliminary information on whether and how we can leverage teachers to increase student usage.

Testing an intervention to increase student engagement in virtual tutoring sessions

Erin Devers and Chris Devers, JoinIn Ventures ; Carly Robinson, Stanford University

In data collection, working paper expected fall 2023.

Motivations: Past data examining the effectiveness of FEV tutor has found that students who have higher average participation points per tutoring session show more growth on standardized tests (both ELA and math) than students with lower average participation points per session. The consistency of this finding suggests that there may be a causal impact of student participation on the efficacy of tutoring. The following experiment is designed to test the impact of more explicit participation feedback from the tutor on student achievement in math.

Research Questions:
1. Do tutors assigned to the Treatment group award more participation points than tutors assigned to the Control group?
2. Do students working with tutors assigned to the Treatment group score higher on the end-of-year MAP test in the subject in which they are receiving tutoring?

Methods: RCT; students randomly assigned to tutors who received the additional participation point training or not

Location: Large southern school district

Data: Student-level administrative data as well as data provided by the tutoring company

Implications: This study would provide information on how to better engage students in virtual tutoring, and whether increased engagement leads to student learning gains.

Do Student-Tutor Demographic Matches Affect Student Engagement and Learning?

Cynthia Pollard, Carly Robinson, Susanna Loeb, Stanford University

Timeline: In data collection, paper expected fall 2023.

Motivation: A growing collection of studies from educational economists has established that an educator's race plays an important role in Black and Brown students' learning. Leveraging large-scale datasets, these studies have found that Black students succeed more when they are taught by teachers who share their racial identity (Clotfelter et al, 2007; Dee, 2004; Egalite et al., 2015; Gershenson et al., 2018; Goldhaber & Hansen, 2010, Redding, 2019). The evidence of positive effects of demographically similar educators has largely been generated by studies with teachers. However, students are increasingly receiving high-impact tutoring with a consistent tutor, but it's unknown whether and how demographic match between tutors and students affect student learning, motivation, and tutor-student relationships.

Research Questions:
1. Does same-race or same-gender matching of tutors to students improve the effectiveness of high impact tutoring on students' achievement, attendance and views of themselves as learners?
2. Are tutors with different demographic characteristics (gender, race/ethnicity) or different experiences more effective at improving students' achievement, attendance and views of themselves as learners?

Methods: The study leverages the random assignment of tutors to students or groups of students to identify the causal effect of tutors on students. At the beginning of the year, Amplify assigns tutors to a given district/engagement. Within each group tutors were placed in a random order, stratified by race, which directs Amplify's scheduling of tutors to student groups. After receiving a semester of tutoring sessions, students will complete the mCLASS benchmark assessment.

Location: Multiple districts and non-public schools across two US states.

Data: Student-level administrative data as well as data provided by the tutoring company

Implications: Findings from this study will provide some additional insight into the educational– and specifically, tutoring–conditions that are most effective for students of color.

Predicting Maintenance of Tier 2 Reading Intervention Effects

David Klingbeil, University of Wisconsin-Madison; Peter Nelson and David Parker, ServeMinnesota; Ethan Van Norman, Lehigh University

In process - Preliminary results from the first wave of data collection likely available fall 2023.

Motivation: The majority of tutoring related research focuses on whether students respond successfully at a point in time. In addition, the frameworks within which tutoring is often delivered (i.e., MTSS, RTI) assume that students who successfully respond will maintain desired levels of skills over time. Scant research has been conducted on this topic, but the few studies that have been conducted suggest that desired skill levels may not be sustained over time. This federally-funded research project examines factors that promote literacy skill maintenance after tutoring ends.

Research Questions: RQ1: To what extent are students' pre-intervention literacy skills and student engagement during reading instruction associated with growth in reading fluency during tutoring? RQ2: To what extent do students' literacy skills and engagement during Tier 1 instruction predict the maintenance of tutoring effects on reading fluency over time?

Methods: Correlational design using novel data collection

Location: Minnesota

Data: Literacy data (fluency, sight words, phonics, comprehension, linguistic skills)
Observation data (behavioral engagement, classroom literacy environment)
Implementation data (tutoring dosage, intervention type)

Findings: n/a--the study is currently in its second year of a four-year project. The four years consist of two two-year waves of data collection.

Implications: Better understanding of what key literacy skills -- or malleable environmental factors -- help sustain tutoring effects.

Are Certified Teachers Differentially Effective Tutors for Early Literacy?

Kalena Cortes, Texas A&M University; Karen Kortecamp, George Washington University; Susanna Loeb, Stanford University; Carly D. Robinson, Stanford University

In progress - working paper expected winter 2023

Research Question: To what extent does having a certified teacher as a tutor make an effective tutoring program more effective?

Methods: Randomized Controlled Trial. We randomly assigned certified teachers and other tutors to classrooms implementing Chapter One tutoring.

Locations: Broward County Public Schools, Florida

Data: Chapter One assessment and admin data from Broward County Public Schools.

Findings: Not yet available

Randomized controlled trial evaluation of the amount of coaching necessary to ensure high fidelity

Peter Nelson and David Parker, ServeMinnesota, [Might be a university partner, identified in the future]

Just beginning - results expected summer 2024

Motivation: Although coaching is one of the most powerful drivers of accurate implementation, costs are significant for tutoring programs. Practitioners and program providers need better evidence with respect to how much coaching is necessary and for whom it's necessary. This study will contribute to the knowledge base for coaching tutors by randomly assigning tutors with initially high fidelity (i.e., 90+% accuracy on an initial observation) to two conditions (2x monthly vs. biannually) of coaching throughout the year.

Research Questions:
1. What differences (if any) exist in the fidelity scores collected at mid-year and end-of-year as a function of treatment assignment?
2. What differences (if any) exist in student literacy scores collected at mid-year and end-of-year as a function of treatment assignment?

Method: Randomized controlled study, with treatment assignment randomized for all tutors with high initial fidelity on tutoring procedures.

Location: North Dakota

Data: Fidelity data (number and percent of implementation steps accuracy across observations) Student outcome data (performance on grade/age-appropriate literacy measures) Implementation data (for context and potential analytic purposes: dosage, intervention type, etc.)

Implications: Practitioners and program providers will have better information on how much coaching is necessary to ensure accurate program implementation.

Are Humans Necessary? Does pairing a human tutor with an AI tutor increase effectiveness?

Susanna Loeb, Carly Robinson, and Nancy Waymack, Stanford University

Just beginning – piloting in spring 2023, Results available summer, 2024

Motivation: We hypothesize that building a relationship with a caring tutor is a key component of high-impact tutoring. Online, AI-driven tutoring lacks the caring adult but provides personalized instruction that has shown to be effective in a lab setting. Understanding the effect of a human tutor on both academic and social emotional outcomes is important to understand as this technology scales.

Research Questions:

1) What is the effectiveness of AI-driven early literacy tutoring, with and without human tutors, for participating students in grades 1 - 5?
2) What is the impact of acting as a tutor for high school and college tutors?
3) Can novice young adult tutors add value without providing the core reading instruction to students?
4) What is the impact of a human tutor for specific sub-populations (e.g., ELL students, students who are furthest behind)?
5) Does the addition of a human tutor increase student fidelity in using Amira?

Methods: RCT

Location: California (Central Valley)

Data: Student-level administrative data as well as data provided by the tutoring company

Implications: This study will help to isolate the effect of a caring tutor in addition to personalized instruction on students' academic and social outcomes.

Project SPARK

Chris Lemons, Stanford

In process - Available July 2024

Motivation: Students with intellectual and developmental disabilities (IDD) in grades K-8 are in need of intensive instruction in literacy and mathematics. Our aim in this study is to evaluate a training model in which we provide professional development and on-going coaching to paraeducators who then provide one-on-one tutoring to students with IDD.

Research Questions:
1) Are we able to train paraeducators to implement reading/math interventions with fidelity?
2) Do we see increases in paraeducators' knowledge, skill, and fidelity of implementation?
3) Do we see related improvements in students' academic skills?

Methods: Randomized control trial.

Location: Tennessee, Texas, Kentucky, Alamaba

Data: Measures of paraeducator knowledge, skill, and fidelity; Student measures related to literacy and math.

Findings: Positive implementation. Good fidelity. Initial indications of student learning.

Implications: We can train paraeducators to be effective interventionists for this population of learners. Remote methods of training are effective.

Using New Methods to Understand Tutoring Relationships and Student Engagement Over Time

Dora Demsky, Carly Robinson, Susanna Loeb, Stanford University

In process

Motivation: Virtual tutoring programs offer an unprecedented opportunity to collect data on the learning experience. Every second of every session produces data. Researchers can observe educator-student conversations, gestures, facial expressions, and monitor the materials with which students engage. Advances in computational methods make it possible to analyze those data for insights into human interaction and behavior, advancing our understanding of the social experience of teaching and learning. We leverage these data and recent advances in natural language process (NLP) techniques to advance how we measure educator-student interactions and learn how they contribute to student engagement and learning.

Research Questions:
1) What percent of time do tutors vs. students spend talking during sessions?
2) How can we categorize the tone/sentiment of tutor-student interactions?
3) How much time during tutoring sessions are spent on (a) content/direct instruction, (b) content/self-paced work, (c) behavior management, and (d) relationship-building.
4) How can we use these results to develop proximal measures of engagement and relationships in tutor-student interactions? How do these measures correlate with other student outcomes?
5) What are promising points of intervention for encouraging positive educator-student relationships as a means for enhancing student outcomes?

Methods: Descriptive; Natural Language Processing

Location: Multisite

Data: Video and audio recordings; transcripts; school administrative data

Implications: Because we observe real world learning interactions over time, we will be able to capture the changing characteristics of interactions over time (Pianta et al., 2012). The rich outcome data allow us to explore how educator-student interactions predict student engagement and learning. Most educational research studies focus on distal outcomes, such as measuring the impact of classroom educators on students' overall attendance and state assessments. However, we are also able to collect proximal outcomes, utilizing transcript and video data to examine in-the-moment patterns of engagement in virtual settings (e.g., Toti et al., 2001; Slater et al., 2017). By collecting frequent, objective information about educator-student interactions and creating NLP measures, we will be able to identify strengths and targets for continued educator improvement and we will provide a model for future work on a broader range of educator-student interactions, including those that focus on other effective instructional practices (e.g., instructional supports and classroom organization; Pianta, 2017). Our ultimate goal is to design interventions that support educators in interacting with students to accelerate learning and well-being.

IMPLEMENTATION and PROCESS

Note: White, Groom-Thomas and Loeb (2022) are working on a summary of existing published research, a newer version will be available in July 2023: https://www.edworkingpapers.com/ai22-652

AI-Generated Feedback for the Virtual Tutoring Community: Lessons Learned through Design Research

Alyssa Van Camp, TeachFX; Catherine Pratt, Mathematica

Working Paper Available:
https://drive.google.com/file/d/10BraP_avG4SJ0Ln2Ba2CQUhVRlZpQdS7/view

Motivation: In late 2021, TeachFX partnered with Cignition, a math and literacy virtual tutoring provider, to study the usefulness and usability of the existing TeachFX offering for virtual tutors and to gather insight on how to adapt the existing offering to the broader virtual tutoring context, including teachers, tutoring program managers, and parents.

Research Questions: What motivates Cignition tutors and the program management team, and what are their needs? How might TeachFX be useful to them? What information about tutoring sessions is useful for parents and math teachers? How would they like to receive information on tutoring sessions? How useful do Cignition tutors find TeachFX for their tutoring? How can math tutors create an environment where students feel comfortable participating and feel supported and successful in tutoring sessions? How useful do Cignition program managers find TeachFX as an addition to their tutoring program? How easy is it to navigate the TeachFX Administrator Dashboard? How challenging was it to incorporate TeachFX into an existing tutoring system? Are there changes over time in tutor instructional practices and student talk time when Cignition tutors use TeachFX? What is the association between select high-leverage instructional practices and the student–tutor talk ratio?

Methods: TeachFX and Cignition collected two rounds of qualitative data in 2022. They conducted interviews with virtual tutors leading group and 1:1 tutoring sessions, tutoring program managers, students participating in tutoring, and their parents and teachers. Initially, interviews focused on the usability and usefulness of TeachFX's existing offering. The second round of interviews, conducted in fall 2022, focused on the perceived usability and usefulness of mock-ups of new product features and functionalities designed specifically for the virtual tutoring context. Students who participated in Cignition group tutoring completed surveys at the end of each week. which the project team analyzed descriptively. The project team also gathered quantitative data using TeachFX, including data on tutors' use of wait time, open-ended questions, and opportunities to build on students' contributions, to analyze associations between tutor practice and student talk during tutoring sessions. The project team designed this study to explore users' perspectives about applying an existing product in a new educational context. Therefore, all study analyses are exploratory.

Location: TeachFX partnered with Cignition and a suburban/rural school district in the southeastern United States to pilot its tool in a tutoring context. Cignition offered its group tutoring program to students in spring and fall 2022 and offered an additional 1:1 tutoring program during the fall. Students participated in group tutoring during an intervention period during the school day, whereas students who participated in 1:1 tutoring did so after school hours. The fall group tutoring program began in September, and the 1:1 program began in October. Both programs ended in mid-January 2023. The amount of tutoring varied by school site: 3-5 days/week, 35 -55 minute sessions. Among students who participated in tutoring, 51 percent were White, 29 percent were Black, 9 percent were Latino, 10 percent were multiple races, and 1 percent were Asian.

Data: The project team collected qualitative data in two phases during spring and fall 2022 and conducted interviews with Cignition's virtual tutors and program managers, classroom teachers and interventionists from the participating school district, and students who participated in Cignition tutoring and their parents. Initially, interviews focused on the usability and usefulness of TeachFX's existing offering (research questions 1 through 4). The second round of interviews, conducted in fall 2022, focused on the perceived usability and usefulness of mock-ups of new product features and functionalities designed specifically for virtual tutoring context (research questions 2 through 5). All interviews were conducted remotely and recorded using Zoom. The project team analyzed transcripts for common themes across each respondent group and research question. In tandem with the qualitative design research described in this report, TeachFX conducted exploratory analyses of the associations between high-leverage instructional practices and the student–tutor talk ratio measured by TeachFX (research questions 6 and 7). TeachFX used data collected during Cignition's fall 2022 group tutoring program for these analyses.

Findings: After using TeachFX for 14 weeks in fall 2022, Cignition tutors reported TeachFX increased their awareness of their instructional practices and student participation during tutoring sessions. Tutors also reported TeachFX improved the quality of their tutoring. Parents and teachers of participating students responded positively to the Individual Student Dashboard, sharing that it provided clarity about student learning and participation in tutoring sessions. Cignition program managers reported mock-ups of the Administrator Dashboard would improve the quality and effectiveness of feedback they provide to tutors. TeachFX also explored relationships between some of the high-leverage instructional practices its app detects and a key outcome of interest: the student–tutor talk ratio. Quantitative analyses revealed a strong, positive association between the tutors' use of wait time 2 (the amount of time spent waiting after a student speaks) and student talk time (0.60, $p < .001$). This finding suggests providing wait time after a student speaks can promote student participation in tutoring sessions, which is consistent with Rowe's (1986) findings that wait time 2 is particularly powerful for increasing the probability of student elaboration in classroom settings.

Implications: Based on this research, TeachFX sees implications for the broader field of high-dosage tutoring. Feedback from tutors indicates that TeachFX, an automated feedback tool, is a high-value addition to the virtual tutoring context. These findings indicate a need for additional research and evaluation of tools like TeachFX in virtual tutoring contexts to determine whether these tools are a cost-effective way to improve tutoring quality at scale. Conduct research on

TeachFX's impact. This descriptive study provided valuable insight into the user experience; however, TeachFX plans to conduct additional research to validate quantitative findings in larger controlled experiments and to further understand the impact of TeachFX in other tutoring contexts not studied in this research study. Areas of interest for further study include in-person tutoring, 1:1 tutoring, tutors with less professional classroom experience (including near-peer tutoring), tutoring offered after school, and on-demand tutoring. Explore additional use cases for the Individual Student Dashboard. The project team spoke with students to understand the conditions in which they feel comfortable participating and being supported in their tutoring sessions. The Individual Student Dashboard was designed primarily for parents and tutors. However, prior feedback TeachFX received suggests students might enjoy and find motivation in viewing the Individual Student Dashboard. In the second phase of research, the project team shared a mock-up of the Individual Student Dashboard with six students for feedback. One student expressed interest in reviewing their own data in the dashboard format, but the other respondents were indifferent. TeachFX seeks to further understand how providing students with objective data affects their engagement in tutoring sessions. Explore developing automated measures to assess relational dynamics in virtual tutoring. Student respondents shared that tutoring practices that sought to build trusting student–tutor relationships make them feel more comfortable participating in sessions. Routine ice breaker activities and non-academic questions that allow students to share information about themselves with their tutors helped students feel more comfortable participating in tutoring sessions. This finding suggests virtual tutoring providers might consider structures to strengthen student–tutor relationships. Given the importance of relational dynamics in student learning, TeachFX is interested in developing automated measures to assess and track the student–tutor relationship over time.

Scaling Promising Practices: A framework from the expansion of high-impact tutoring

Leah Groom-Thomas, Chung Leung, Susanna Loeb, Cynthia Pollard, Sara White, Stanford University

In process - Paper expected summer 2023

Motivation: Educational initiatives, even the most evidence-based, are renownedly difficult to scale. In this paper, we analyze the use of a particularly well-researched educational approach – high-impact tutoring – as a response to the Covid-19 pandemic and identify both enabling conditions and barriers to effective scaling. Drawing on literature from across the disciplines of public health, communications, and education, we identify core conditions for program scaling within decentralized systems like that of U.S. public education. We propose a framework for identifying facilitators and barriers to scaling related to the intervention design, engagement, implementation support, and the policy landscape. Applying this framework, we analyze efforts to expand and implement high-impact tutoring programs across nine school districts in the United States for which we have collected detailed primary and secondary data.

Research Questions: (1) Program features: What features of tutoring do districts aim for and to what extent do these align with the research base on effective practice and the fit with current

school practices? (2) Engagement with tutoring: How do educators learn about tutoring and whom do they turn to for information on how to effectively scale tutoring? (3) Implementation capacity: What capacities do districts need to begin to implement tutoring and do those capacities come from internal or external sources? And (4) Policy and economic environment: What policies and economic pressures and opportunities influence the choice and initial expansion of high-impact tutoring?

Methods: Interview and document analysis; qualitative.

Location: Nine school districts across the US. Each of the included districts implemented and expanded high-impact tutoring programs during the 2021-2022 school year.

Data: The data come from 112 interviews which took place from January through July of 2022 across 90 key informants including school district administrators, tutoring program administrators, teachers, tutors, and external support organizations. Interviewees provided detailed, in-depth information about the selection, launch, and implementation of district-led tutoring initiatives.

Findings: The findings from this study support both a multidimensional framework for understanding how promising practices can scale in a decentralized education system and provide specific insights into the effective scaling of high-impact tutoring in response to the pandemic. We find that not only are the features of the intervention important in scaling – such as whether it is more effective than current practices and whether it fits easily into organizational practices – but also that the capacity of the institutions to make a change, the ease of access that decision-makers have to useful information from trusted sources about the innovation, and whether policy and economic environments incentivize and facilitate implementation.

Implications: The study provides concrete examples of the ways in which a scaling framework focused on adoption and implementation processes may prove beneficial for our understanding of scaling in education more broadly.

High-Impact Tutoring at Scale: Implementation of District Tutoring Initiatives across the U.S.

Leiah Groom-Thomas, Susanna Loeb, Sara White, Nancy Waymack, Stanford University

Initial analyses done but not finalized - Results expected Summer 2023

Research Questions: The primary research question is: What have been the key facilitators and barriers to successfully implementing tutoring? We also ask: What processes did districts go through to choose tutoring? Who and what have been the key decision makers when selecting tutoring programs and designing roll-out? And what are the characteristics of tutoring programs that districts are providing and how do they align with what research suggests is most effective?

Methods: Mixed methods - Interview and document analysis; descriptive administrative data analysis

Location: Nine school districts across the US. Each of the included districts implemented and expanded high-impact tutoring programs during the 2021-2022 school year.

Data: The data come from 112 interviews which took place from January through July of 2022 across 90 key informants including school district administrators, tutoring program administrators, teachers, tutors, and external support organizations. Interviewees provided detailed, in-depth information about the selection, launch, and implementation of district-led tutoring initiatives.

Findings:
Challenges

- Many districts experience start-up delays when launching tutoring programs due to systems challenges with background checks, procurement processes, lagging technology, and hiring difficulties, such as the insufficient availability of tutors.
- Administrative and interview data provide insights into challenges with student attendance and tutoring dosage, namely lack of school structures for embedding tutoring during the school day, minimal personnel support to aid students logging into virtual sessions, and not enough safeguards in place to ensure students attend tutoring sessions with a consistent tutor.
- Stakeholders experienced challenges achieving and maintaining school-level buy-in and ownership due to inconsistent communication about tutoring initiative and a perceived lack of autonomy for tutoring implementation, compounded by staff exhaustion after a full year of pandemic teaching.

Opportunities

- Districts can increase sustainable pipelines for tutor supply and retention by building mutually-beneficial partnerships with local universities and community organizations for long-term implementation at scale.
- District leaders can promote their commitment to tutoring as a long-term solution for learning acceleration by allocating resources to build capacity and tutoring infrastructure at the school level in year one of implementation, followed by engaging in data-driven continuous improvement cycles for scaling programs in year two.
- Umbrella organizations can serve as liaisons to connect district-level leaders to promote collaboration among stakeholders. Interview findings underscore the importance of shared vision at all levels of implementation, ensuring adequate systems are in place prior to rollout, and/or to first deliver a smaller, manageable tutoring program to students such that districts can troubleshoot challenges prior to scaling.

Creating Possibilities: Online learning through Minecraft

Elizabeth Kozleski, Tamara Handy, and Nick Haber, Stanford University; Kathryn Ringland, UC Santa Cruz

Initial analyses done but not finalized - intend to publish the results in the Summer of 2023.

Motivation: Exploring the world of online gaming and the affordances it offers neurodiverse and neurotypical learners to play and learn together.

Research Questions: To what extent can co-designed MineCraft spaces afford diverse learners an interactive play space in which learners with widely varying approaches to sustained learning can engage in reflexive play? To what extent are these play groups sustained over time? How effective

are peers at designing play spaces that work for all levels of learners? How do these spaces afford reflexive play? What is the variance in reflexive play moves?

Methods: Observation, field notes, design labs, video recording
An iterative design format allowed participants to design, revise, and continue to improve play through design. A single case design across conditions was used to measure change in performance.

Location: Online in a Minecraft space

Data: Narrative, video, and observational data.

Findings: Subtle shifts in play activity by neurodiverse players were perceived and responded to in play with minecraft tools. Players engaged in redesign and discussion of the gaming space. Players wanted to continue the playspace after the end of two 4 week sessions of play. Subtle shifts in movement on the screen were responded to and extended play between the players.

Implications: When children between 8 and 12 co-design MineCraft spaces, they can afford diverse learners an interactive play space in which learners with widely varying approaches to sustained learning will engage and sustain reflexive play. The play groups need support to sustain over time. Peers can be successful and engaged in designing play spaces that work for all levels of learners. Reflexive play can be useful in inclusive learning spaces and benefit all learners.

School-Level Perspectives on Scaling up High Impact Tutoring: Goals, Challenges and Strategic Priorities

Patricia Burch, USC, Professor and Co-Director of USC EdPolicy Hub; Susanna Loeb, Professor, Stanford University; Alvin Makori, Research Assistant

In process - Results available summer 2023

Motivation: To provide district partners with usable data to scale tutoring in ways that is inclusive of school-level perspectives; to understand how challenges to scaling might vary depending on whether "partner" is a CMO or traditional school district.

Research Questions:
1. What are school level perspectives on challenges to scaling up high impact tutoring within the school day? What strategies are being used at school level to address challenges?
2. Do school level perspectives on scaling up look different in context of a CMO (in theory less regulated policy context) as compared to a traditional school district (in theory more regulated policy context)?

Methods: Descriptive survey data from classroom teachers, curriculum coordinators and administrators, interview data

Location: Guilford County and Uplift Dallas

Findings: cannot share findings publicly until we debrief with sites in June 2023

Implementation of High-Impact Tutoring through Grants to Tutoring providers

Susanna Loeb, Cynthia Pollard, and Nancy Waymack, Stanford University

In process: Results available fall 2023

Motivation: A SEA is providing grant funding for tutoring providers to work with low-performing schools. We will study the implementation and provide feedback for improvement.

Research Questions: What are the key features of the tutoring programs and the OSSE HIT initiative? To what extent do the Initiative and the programs reflect research-based features of HIT? What happened during the program? Was the model implemented with fidelity? What are the characteristics of the students who received tutoring? Did they differ from students overall? Did characteristics differ by provider? How much tutoring did students receive? Did dosage differ for different groups of students?
What is the relationship between core tutoring features (attendance, dosage, group size, schedule) and student achievement scores (2023 PARCC scores, and EOY Math and Reading Scores)?

Methods: Mixed-methods, descriptive and correlational

Location: DC

Data: Student-level administrative data and tutoring session data from tutoring providers

Implications: In most cases, schools and / or districts fund high-impact tutoring through grants they receive or other funding. This model flips the script putting the money in the hands of providers and asks them to select schools to be their partners.

From Policy to Implementation: Texas HB4545 as an Example of Education Reform

Nancy Waymack, Stanford University; Sharla Horton-Williams, Ed.D., The Commit Partnership

In process - Results available: Fall 2023

Motivation: Understanding the impact of HB4545 in district implementation of tutoring—benefits, barriers, unintended consequences, recommendations for future approaches to statewide legislation.

Research Questions: Many, including: What was the process for conceptualizing, designing, and authoring the bill? How was high impact tutoring implemented and how did the bill influence this process? What revisions or addenda to the current law would districts or TEA like to see as a result of experiences from current implementation?

Methods: Qualitative; document analysis, interviews.

Location: Virtually and in Dallas

Data: Document analysis, interviews

Findings: TX districts are in need of more secondary math curriculum options. Districts struggled with hiring enough tutors, often used the parent waiver on group sizes to get more students adequate tutoring hours. K-2 literacy results look promising so far, albeit in low sample sizes.

Implications: Could help prepare other states, districts, providers, and umbrella organizations to create effective high-impact tutoring legislation or prepare for implementation challenges.

Wittenberg Community High-Dosage Tutoring Project

Kristin Farley, Wittenberg University; Leiah Groom-Thomas, Stanford University

In process - results expected Summer-Fall 2023

Motivation: Wittenberg University Education Department pre-service teachers provide high-dosage tutoring during in-school intervention periods for students in the Northwestern Local and Springfield City School Districts.

Research Questions: What are tutors' feelings of efficacy and how do these feelings change throughout the year? What are classroom teachers' perceptions of tutoring and do these perceptions change throughout the year? What are students' attitudes towards tutoring and do these attitudes change throughout the year?

Methods: mixed methods; descriptive

Location: In two school districts in Ohio: Northwestern Local and Springfield City School Districts

Data: surveys, pulse checks, and administrative data from the district

Findings: Elementary students who receive tutoring report increasingly positive attitudes toward tutoring and school in general. Tutors report greater confidence in their content knowledge and abilities to respond to individual student's instructional needs. Classroom teachers report improved academic performance based on informal formative assessments.

Implications: Universities can leverage teacher preparation programs to provide high-impact tutoring to students in local school districts.

Literacy Liberators: Educational Impacts and Conditions for Success

Ashley Jochim (Center on Reinventing Public Education, Arizona State University), Travis Pillow (Center on Reinventing Public Education, Arizona State University), Eupha Jeanne Daramola (University of California Santa Barbara), Heather Casimere (Center on Reinventing Public Education, Arizona State University)

In Process, available January 2024

Motivation: With the Literacy Liberator initiative, OUSD, The Oakland REACH and Fluent Seeds have devised a novel approach to meeting an urgent need—mobilizing more adults to tutor student in reading—in support of a broader, more transformational aim: activating families

systemwide to support improved learning conditions for students. Our evaluation of the Literacy Liberator initiative aims to:

1) Assess whether a tutoring program that makes use of community members improves outcomes for students

2) Evaluate how becoming a Literacy Liberator shapes community members' expectations for public education and future career aspirations

3) Identify the conditions for effective implementation, including strategies for collaboration between Liberators and classroom educators as well as among the district and nonprofit partners supporting the Literacy Liberator initiative.

Research Questions:
-Impact on Liberators: How does becoming a Liberator influence participants' beliefs, expectations and sense of efficacy around K-12 education? To what extent does becoming a Liberator activate greater levels of engagement, advocacy or interest in joining the educator workforce?
-Relationship between Liberators and classroom educators: What conditions limit or support productive collaboration between school staff and Liberators? How can partnerships between Liberators and school staff be strengthened to support better learning conditions for students?
-Impacts on students: What early evidence exists of the impact of tutoring by Liberators on student achievement?
-Conditions that enable partnership: What conditions in The Oakland REACH, Fluent Seeds, and OUSD enable and constrain partnership? How do these conditions shape the effectiveness of the program and the impacts on students and Liberators?"

Methods: This is a three-phase, mixed-methods implementation study.
-Phase one: Qualitative data from interviews with Liberators, classroom educators, and principals as well as district and partner staff with responsibilities for managing the initiative.
-Phase two: Survey of Literacy Liberators about recruitment, working conditions, and experience with the tutoring model.
-Phase three: Impact analysis to understand how participating in tutoring influences students' literacy outcomes.

Location: Oakland, California

Data: Data gathered from participants about their experiences as well as data on students' literacy outcomes.

Findings:
#1: Systemic challenges may undermine the success of the Liberator initiative
-"Bandaid over a bullet hole": Uneven implementation of evidence-based Tier 1 instruction may dilute the impact of Liberators on students' literacy outcomes
-Logistical challenges, including covering for staffing gaps and duties outside core responsibilities, may divert focus of literacy coaches and classroom teachers and reduce opportunities to collaborate with Liberators
-Variability in site-level implementation and coaching support coupled with inconsistent evaluation may leave impact to chance

#2: Training provided by The Oakland REACH strengthened pre-existing sense of purpose in Liberators
-Participants in the Literacy Liberator fellowship (a new training program designed by The Oakland REACH for community members they recruited into this role) said the experience activated their sense of agency over improving children's literacy outcomes
#3: Pay and working conditions are a critical barrier to sustainability
-Low pay and inadequate hours, benefits presents recruitment and retention challenges

Implications: First, this study presents a deep dive into the unexpected challenges that come with shifting a district operating system to support improved literacy outcomes via tutoring and other aligned initiatives (e.g., curriculum adoption, coaching). Our findings will have implications for policymakers and advocates concerned with implementation of the "science of reading," and identify the conditions that limit or accelerate the shifts in instruction that are needed to make good on the commitment to improved literacy outcomes. Second, this study will document a novel approach to increasing access to tutoring–by leveraging adults who are connected to the school communities in which they work and empowering them with the skills and mindsets needed to change outcomes for students. The approach in Oakland stands in contrast to other initiatives that rely on college students or volunteers, who are unlikely to have ties to the neighborhoods that most need access to tutors and less likely to share the linguistic and cultural heritage of the students with which they support. Finally, this study will identify the opportunities and obstacles districts confront in their efforts to leverage external partners to support the design and implementation of new initiatives.

Outcomes and Cost Evaluation of Early Elementary School Literacy Tutoring (2022-23)

Norma Ming, SFUSD; Carolina Ramirez, SFUSD

Just beginning, in design phase

Motivation: SFUSD seeks to understand the relative value and impact of two external literacy tutoring services.

Research Questions:
(1) Can we observe differences in outcomes among literacy tutoring participants, when compared to matched non-participants?
(2) Can we observe differences in literacy outcomes participants of one tutoring service, when compared to participants of another service?
(3) What is the cost of each tutoring program?

Methods: Multi-level regression analysis. Economic evaluation

Data: Administrative data, Cost data collected from documentation and cost interviews

Revising Tutor Responses in Math Education: Teacher Insights and NLP Strategies

Authors: Rose E Wang, Dora Demszky, Carly Robinson, Susanna Loeb

Just beginning

Motivations: The [recent national report card in mathematics](#) lays bare the ways the pandemic has disrupted students' math education. It highlights the need for effective interventions to help students catch up. One-on-one tutoring has been shown to be an extremely important approach, especially because it provides personalized support to students. However, tutors often do not receive frequent feedback or aid in teaching students. We are interested in providing support to tutors in the form of improving their responses to student math mistakes to be more useful and caring. We recognize that it is not enough for tutors to simply point out errors—they must also provide meaningful feedback that helps students understand their mistakes and how to avoid them in the future. Moreover, we believe that it is important for tutors to approach their work with a sense of empathy and care, recognizing that students may be struggling and need encouragement and support. Through our research, we hope to provide insights into how tutors can be more effective in their work and how we can better support students in their mathematics education. Ultimately, we believe that by improving the quality of one-on-one tutoring in mathematics, we can help to reverse the trend of declining national test scores and ensure that all students have the opportunity to succeed in this important subject

Research Questions:
1. How do experienced math teachers naturally revise tutor responses?
2. How do they identify student errors? How do their strategies change with respect to the error they've identified?
3. Can NLP tools help provide automated revision feedback to future tutors based on the math teachers' expertise?

Methods: Descriptive; Natural Language Processing

Location: USA

Data: Video and audio recordings; transcripts; school administrative data

Implications:
1. Improved quality of one-on-one tutoring in mathematics: The research can provide insights into how tutors can be more effective in their work, ultimately improving the quality of one-on-one tutoring in mathematics.
2. Enhanced student learning outcomes: By providing tutors with more effective feedback strategies, students may be able to better understand their mistakes and learn how to avoid them in the future, leading to improved learning outcomes.
3. Use of NLP tools in one-on-one tutoring platforms: The research explores the potential for using NLP tools like large language models in education to provide automated feedback in the form of revisions to tutors.