NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Research in Progress to Better Understand High-Impact Tutoring

NSSA 2026 Conference

Human Connections in a Digital Age: The Next Chapter of High-Impact Tutoring

May 6, 2026

Stanford University

IMPLEMENTATION, PROCESS, AND SYNTHESIS

Making Dosage Possible: Administrative Ease and Impediments in the Implementation of Tutoring Programs

Michaela Krug O'Neill, University of Michigan; Patricia Strach, University of Albany SUNY; Lindsey Kaler, Brown University; Susan Moffitt, Brown University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: In Progress

Motivation: In light of calls for increased dosage in the tutoring literature, this qualitative study examines what makes dosage possible. We aim to learn about the burdens and benefits paraeducators, teachers, and administrators perceive in relation to the different tutoring models that we focus on in our study, and what it took for them to make one-to-one and small group supplemental instructional opportunities work in their contexts.

Research Questions: 1) How do frontline providers perceive an intervention's implementation ease and impediments? 2) How do policy and program terrains bear on the implementation of high impact tutoring?

Methods: Qualitative; We interviewed 36 educators (i.e., paraprofessionals, teachers and instructional leaders) across purposively sampled school sites within our study sample, which includes six districts across multiple states.

Location: DC; Florida; Indiana; North Carolina; Tennessee

Data: Qualitative data from semi-structured interviews. We designed and implemented interview protocols for paraeducators, teachers, and instructional leaders aligned with our research questions, which asked these groups about the burdens and benefits of their tutoring models and the roles that paras and other support staff have within schools and communities to support teaching and learning.

Implications/Findings: We find that making dosage possible requires understanding the resources available and the tradeoffs involved. While single factors can lead to the failure of a tutoring program (e.g. poor materials, resistant staff, insufficient funding), programs were more successful when they were aligned with existing resources (e.g. staffing, schedules, curriculum) and commitments at the classroom, school, and district levels, and when they took context and policy history into account. This finding has implications for the implementation of tutoring programs at scale, particularly as policymakers think about which tutoring programs could be most effective in their unique context with the aim of closing longstanding achievement gaps.

Improving AI Tool-based Tutoring Implementation in Classrooms

Briana Jocelyn, Stanford University; Monica Lee, Stanford University; Susanna Loeb, Stanford University

Status: Data Collection

Motivation: This study examines how a tutoring provider using AI-based literacy tools is being implemented in K–2 classrooms in a medium-sized urban school district. Although the tool is broadly available, teacher uptake appears uneven, raising questions about barriers to use, the role of coaching and school structures, how well the screener aligns with other assessments, and whether implementation differs for students with IEPs. The study is motivated by the district's interest in scaling effective, equitable, technology-enabled early literacy instruction.

Research Questions: 1) What barriers and supports shape teachers' use of the AI-based tutoring tool, especially in low-usage settings? 2) What are teachers' attitudes toward the tool, and why do some use or avoid the pre-screener? 3) Which school and classroom structures support strong and sustained implementation? 4) How do coaching models and coach-leader alignment influence outcomes? 5) What routines and practices characterize high-usage classrooms? 6) How does usage and impact differ for students with IEPs compared with general education peers? 7) How accurate is the screener, and how well does it align with other literacy assessments?

Methods: Mixed methods. The study includes a districtwide survey of K–2 teachers using the tool, purposive sampling of higher- and lower-engagement classrooms for interviews and observations, and quantitative analysis of usage and assessment data. Data collection includes a brief teacher survey, semi-structured interviews with teachers, instructional coaches, and school leaders, classroom observations using a shared rubric or observation form, and administrative data analysis. The analytic plan includes descriptive and inferential quantitative analyses plus thematic analysis of interviews and observation notes.

Location: A medium-sized urban school district serving K–2 classrooms.

Data: Data sources include teacher surveys, interviews with teachers, instructional coaches, and school leaders, classroom observations, platform usage data, student-level administrative records, student/classroom/school characteristics, and other literacy assessment results. The study also uses three waves of screener/usage data across fall, winter, and spring to examine change over time.

Implications/Findings: This study is designed to generate implementation guidance rather than report completed findings. It aims to identify the organizational, instructional, and coaching conditions that support effective use of AI-enabled literacy tools; clarify how these tools can be used equitably, including for students with IEPs; and inform district decisions about professional development, coaching, technology adoption, and literacy strategy. The intended implication is to help the district build more sustainable, authentic, and holistic technology-supported literacy instruction.

Navigating the ESSER Cliff: A Mixed Methods Study of Districts and their Strategies to Sustain High Impact Tutoring Programs

Amanda Lu, Georgetown University; Ayesha Hashim, NWEA; Helen Zhou, Stanford University/Harvard University; Ev Fu Gilbert Stanford/Harvard University; Susanna Loeb, Stanford University

Status: Preliminary Findings and Qualitative Data Analysis

Motivation: By concentrating on the relationship between ESSER's availability and HIT offerings across contexts, we will be able to describe the varying trajectories of implementation and identify conditions that

enable take-up, high-quality implementation, and sustainability of effective programs. More broadly, the study will offer the field practical and theoretical knowledge about the social, organizational, and political conditions that allow for policy scaling and sustainable change from resource infusions in school systems. It will provide qualitative accounts with site-level nuance and local context, which will accompany more national-level, summative quantitative research on the impact of ESSER on student achievement. In addition to having causal evidence of ESSER's impact on achievement and descriptive evidence of how ESSER dollars were spent, a comprehensive account of how ESSER dollars contributed to altering pre-existing structures for learning provides us a more expansive understanding of how school systems can productively use federal aid and cash infusions.

Research Questions:

1. How did districts utilize ESSER funding to facilitate the growth of HIT programs? Did these programs successfully target underserved populations?
2. How does a district assess and perceive its HIT programs' value in the presence of the ESSER cliff?
    a. What data do they draw on to make these assessments?
    b. What is the process for making these assessments?
    c. How are these assessments operationalized into a decision-making process about continuing programs?
3. What elements of HIT programs can districts maintain from their tutoring programs post-ESSER?
    a. What funds do they use? What tradeoffs with other programming might be occurring?
    b. What types of programs are more likely to survive?
    c. How do districts change their approaches to providing relationship-based personalized instruction?
4. What enables districts to maintain these approaches?
    a. What is the role of state policy?
    b. What is the role of advocacy?
    c. Who are the key organizational players that were crucial to sustaining the HIT program?

Methods: Survey and Qualitative

Location: National

Data: Surveys and Interviews. Since the project's inception, we have successfully navigated three major research phases. In Phase 1, we completed a comprehensive quantitative analysis using five waves of nationally representative School Pulse Panel (SPP) data (December 2022 through June 2025) alongside our own targeted survey of 35 public school districts administered in Spring 2025. This phase established the longitudinal trajectory of tutoring offerings during the ESSER period. We also concluded Phase 2a (Media Accounts), which involved reviewing public records to identify districts with unique program trajectories, such as those that paused or canceled programs, to ensure our qualitative sample was not limited by survey response bias. Following this, we launched Phase 2b (Qualitative Interviews) in Fall 2025 and reached our target of 15 in-depth interviews with district administrators by January 2026.

Implications/Findings: Our analysis reveals several key themes regarding the post-ESSER tutoring landscape. First, ESSER funding was highly effective in expanding access to High-Impact Tutoring (HIT) for underserved students; offerings grew to 46% of schools by May 2024, with over 92% of those schools using diagnostic data to target participation. However, a "peak" was reached in May 2024, followed by a modest decline in offerings

by June 2025 as relief funds expired. While districts consistently rated HIT as the most effective model, 40% anticipated smaller programs in SY 2025-26, citing funding and capacity constraints as the primary barriers to sustainability.

During the initial expansion phase, ESSER provided a unique policy window that enabled districts to experiment and scale bold, innovative programs with minimal initial budget constraints. As districts transition away from federal relief, long-term program viability is increasingly dependent on shifting toward in-house delivery models—55% of our surveyed districts reported moving services internally to reduce costs. Successful sustenance also seems to require a “district champion,” a role often filled by district administrators who are positioned to advocate for HIT across stakeholder groups, as well as rely on the formalization of tutoring within existing strategic initiatives to secure stable local or state funding.

A significant finding is that scaling back has, for some districts, served as an opportunity for program refinement. While some districts were forced to reduce their program scale, this “right-sizing” often allowed for increased implementation fidelity and better supervision of tutor quality. By reducing tutor pools to “the best of the best,” some districts report higher-quality implementation and increased integration with existing Multi-Tiered Systems of Support (MTSS) and Response to Intervention (RTI) models. We also identified a “catch-22”: districts that struggled to produce robust impact data during the rapid HIT program expansion phase during ESSER now find it difficult to justify local or philanthropic funding to replace ESSER, despite anecdotal evidence of increased program impact over time. To survive, successful districts are “mainstreaming” tutoring—moving away from high-cost external vendors toward in-house pipelines and embedding sessions into the standard school day.

Personalized Learning Initiative - Implementation Study

Barbara Condliffe, MDRC; Monica Bhatt, University of Chicago Education Lab; Shira Mattera, MDRC; Jean Grossman, MDRC; Rebecca Davis, MDRC; PLI Research Team, University of Chicago Education Lab & MDRC

Status: Initial analyses done but not finalized

Motivation: The global pandemic was a once- a-century public health crisis that left us with a once-a-century public education crisis. Despite the $189.5 billion that the federal government sent to schools through the Elementary and Secondary School Emergency Relief (ESSER) Fund, we still see students struggling to catch up. To that end, we launched the Personalized Learning Initiative (PLI) in 2021, and have since partnered with eight education agencies around the country.

The goal of the PLI is to understand whether and how we can scale the benefits of high dosage tutoring such that more students might benefit. PLI includes implementation, impact, and cost studies, as well as an effort to explore the personalized treatment effects (or heterogeneous treatment effects) of these interventions. PLI is led by the University of Chicago Education Lab in collaboration with MDRC, as well as researchers from the University of Toronto and Stanford University.

Research Questions: How can we scale the benefits of tutoring so that more students have access to high quality tutoring options that are most effective for them? How do local educational agencies (e.g., school

districts, states) implement evidence-based tutoring program at scale? What are the characteristics of tutoring models in our study sample? What challenges did local education agencies face in scaling their tutoring models and how did they address them? What factors facilitated or hindered implementation and scaling of tutoring program?

Methods: Implementation, Mixed Methods

The implementation research for the PLI study leverages qualitative and quantitative data. Since 2021, primary data collection activities have included interviews with teachers, tutor coordinators and tutors, observations of tutoring and surveys of tutors and tutor coordinators. The implementation research study also relies on administrative records regarding student attendance in tutoring as well as systematic documentation and observation from the PLI technical assistance teams. Note that not every form of primary data collection was pursued in each year.

Qualitative analysis has involved systematic coding of interview data using both deductive codes derived from our conceptual framework and inductive codes that emerged from the data itself. The coded data was then analyzed thematically to identify recurring patterns and develop key themes by participant type, site and area of implementation.

Descriptive statistics calculated overall, by site, and program type (e.g. virtual vs. in-person programs) have been used to summarize the quantitative data.

Location: Chicago Public Schools (Illinois); Fulton County Schools (Georgia); New Mexico Public Education Department; Miami-Dade County Public Schools (Florida); Greenville Public Schools (South Carolina); Winston-Salem/Forsyth County Schools (North Carolina); Guilford County Schools (North Carolina); Rocketship Public Schools (California)

Data: The implementation research for the PLI study leverages qualitative and quantitative data. Since 2021, primary data collection activities have included interviews with teachers, tutor coordinators and tutors, observations of tutoring and surveys of tutors and tutor coordinators. The implementation research study also relies on administrative records regarding student attendance in tutoring as well as systematic documentation and observation from the PLI technical assistance teams. Note that not every form of primary data collection was pursued in each year.

Implications/Findings: The IR study is embedded in the pre-registered impact study found: https://osf.io/fkjmn/files/osfstorage The MDRC PLI project page includes a number of products based on the implementation research and the insights from the PLI technical assistance teams https://www.mdrc.org/work/projects/personalized-learning-initiative Early implementation research from the PLI study has highlighted the importance of building a strong infrastructure at the district(or state) and school-levels to support the effective launch and management of tutoring programs. Additionally, we've identified promising strategies schools can adopt in ensuring that their students actually attend tutoring sessions scheduled for them. Additional insights from the implementation research will be released in summer 2025 and beyond. https://www.mdrc.org/sites/default/files/PLI_Tutoring_Brief.pdf https://www.mdrc.org/work/publications/support-systems-needed-expand-successful-high-dosage-tutoring-programs

Accelerating Student Success: Statewide Evaluation of High-Dosage Tutoring in Massachusetts

Amanda J. Neitzel, Johns Hopkins University Center for Research and Reform in Education; Joseph Reilly, Johns Hopkins University Center for Research and Reform in Education; Erica Johnson Mulcahy, Johns Hopkins University Center for Research and Reform in Education

Status: In process

Motivation: This study is designed to meet the Commonwealth of Massachusetts' need for rigorous, actionable evidence while providing timely feedback to improve implementation.

Research Questions:

1. What is the effect of participation in the Massachusetts statewide tutoring initiative on student achievement in English language arts, as measured by DESE-selected assessments and other standardized measures, compared to expected growth? To what extent do outcomes vary based on tutoring dosage, format (in-person vs. virtual), timing (school-day or after-school), and vendor?
2. Does participation in the statewide tutoring initiative vary across student subgroups, including grade, gender, race/ethnicity, English learner status, special education status, economic disadvantage, prior achievement, attendance, and geographic region?
3. What barriers, by vendor or program model, affect student participation, and how do these vary across student subgroups?
4. How do tutor characteristics (e.g., education, professional experience, training received) and tutor perceptions of readiness, especially in meeting the needs of multilingual learners, students with disabilities, and alignment to culturally and linguistically sustaining practices commonly used by DESE through their Culturally and Linguistically Sustaining Practices' framework, relate to implementation quality and student outcomes?
5. What are the costs associated with the statewide tutoring initiative, and what is the return on investment in terms of student learning gains?

Methods: Mixed Methods;This mixed methods study relies on descriptive and correlational analyses of student academic growth, benchmarked against expected growth trajectories based on DIBELS. Where local conditions and data availability permit, we will complement these analyses with quasi-experimental designs (QEDs). These may include comparison group designs or regression discontinuity designs, depending on program structures. Implementation quality will be assessed using quantitative fidelity indicators (such as dosage and adherence to vendor models) in combination with surveys of educators, administrators, and tutors. Stakeholder perspectives will be captured through educator, administrator, and tutor surveys. Cost analyses will be conducted using vendorprovided expenditure data and program documentation. Costs will be linked to student outcomes to estimate return on investment.

Location: Massachusetts

Data: Standardized ELA assessments, Assessment and participation data from the data dashboard, Tutor session logs, tutor survey data, Administrator and educator survey data, Cost data from DESE, vendors and schools.

Implications/Findings: The results from this study will inform state support for student learning through tutoring interventions, providing insight into tutoring impact on student academic achievement, the equity of participation and outcomes, tutoring implementation and fidelity, stakeholder perspectives on tutoring, and the cost-effectiveness and scalability of tutoring.

PROGRAM EFFECTS

Assessing the Long-Term Effects of a Scalable Approach to High-Impact Tutoring for Young Readers

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

Status: In progress

Motivation: High-impact tutoring is one of the most effective interventions for students. However, it is unclear whether tutoring consistently benefits students over the long term as initial engagement due to novelty and personalized attention may decline as the experience becomes routine. Fewer studies have examined the effects of multi-year tutoring programs. This study is one of the first randomized controlled trials of a four-year tutoring program for kindergarten to third-grade students.

Research Questions: What is the effect of a short burst tutoring program on students early literacy outcomes and non-academic outcomes, including special education classification and grade retention?

Methods: RCT

Location: Florida

Data: Provider Data: Outcomes -- Reading Foundation Stage and Oral Reading Fluency. District Data: Outcomes -- Star Early Literacy Assessment and Florida State 3rd Grade Reading Assessment, Special Education Specification, Grade Retention. Student Characteristics -- Gender, Race/Ethnicity, English Learner Status, Special Education Status, Kindergarten-entry test score.

Implications/Findings: This is among the first studies to examine the efficacy of a multi-year tutoring program, which will provide valuable insights as many school districts extend their tutoring programs. In addition, the unique aspects of the tutoring program—a frequent, short burst of technology-assisted personalized instruction from consistent tutors embedded in the classroom—provide useful information on which features can be leveraged to implement tutoring programs at scale.

Link to preregistration: https://www.socialscienceregistry.org/trials/10810

Teachers as (virtual) tutors

Susanna Loeb, Stanford University; Carly Robinson, Stanford University; Ana Ribeiro, Stanford University; Melissa Gentry, Texas A&M University

Status: In process

Motivation: Recent research suggests that virtual tutoring can be effective (Carlana & Ferrara, 2021; Gortazar et al., 2023), but these studies are based outside of the US and demonstrate the positive impact of virtual tutoring during remote learning and among parents who agreed to their children receiving after-school tutoring. Opt-in and after-school tutoring initiatives draw in some parents and students but not others, and, in particular, may lose some of the students who could most benefit from tutoring: the least engaged students (Robinson et al., 2022). Evaluating a virtual tutoring program will shed light on the potential for virtual tutoring to meet the great needs of the US student population most affected by the pandemic and reach the students who need the most support.

Research Questions: Does a virtual tutoring program using only teachers as tutors result in improved academic outcomes for students?

Methods: RCT

Location: US

Data: Student-level administrative data from the school district and tutoring provider data

Implications/Findings: This study has provided evidence on the impact certified teachers can have as tutors in virtual settings. We find that the effectiveness of virtual tutoring is not uniform across subjects; students who received tutoring in ELA saw increased reading assessment scores of 0.143 standard deviations (significant at the 5% level), while students assigned to math tutoring did not see any statistically significant improvements in math scores. There was no cross-subject growth - ie students in ELA tutoring did not grow in math, and vice versa. We did observe heterogeneity by student demographics - effects on ELA assessments were strongest among students who received free and reduced price lunch, students who were not enrolled in special education, and students who are English language learners.

Link to preregistration: https://osf.io/z8e74/overview

Evaluating the impact of a community-based math tutoring program for early elementary school.

David Parker, ServeMinnesota; Peter M Nelson, Slate Evaluation; Carly Robinson, Stanford University, Elizabeth Huffaker, Stanford University, Susanna Loeb, Stanford University

Status: In process

Motivation: Early numeracy is important for future math achievement. This is one of the first studies to evaluate the causal effect of an in-school early numeracy tutoring program at scale.

Research Questions: What are the year-end math outcomes for early elementary aged students randomly assigned to a community-based tutoring program compared to those randomly assigned to control conditions?

Methods: RCT

Location: Minnesota, Wisconsin, Mississippi, California

Data: About 1200 students were assigned to the treatment or control condition across 28 schools. We will collect: Fidelity data (number and percent of implementation steps accuracy across observations); Student outcome data (performance on grade/age-appropriate math measures); Implementation data (for context and potential analytic purposes: dosage, intervention type, etc.)

Implications/Findings: There are few effective early elementary tutoring programs for math that can be feasibly (and locally) implemented. Findings will also help drive continuous program improvement.

Link to preregistration: https://osf.io/z6dc2/overview

An evaluation of Cignition's small group, virtual math tutoring in Grades 3-6

Catherine Asher, University of Michigan; Robin Jacob, University of Michigan

Status: Analyses finalized and being written up

Motivation: Virtual tutoring has the potential to benefit students from a broader set of schools, for example those in communities or regions where the supply of in-person tutors may be limited. In this randomized controlled trial we are evaluating the impact of a virtual, small group math tutoring program on student achievement and growth for students who are struggling with math skills. Students in grades 3-6 who were in Tier 2 or 3 in math at the beginning of the school year were randomly assigned to either receive Cignition or to participate in "business as usual" math instruction. Students in the treatment group meet virtually with a Cignition tutor several times per week in a small group.

Research Questions:

1. What is the impact of access to Cignition's virtual small-group math tutoring on math achievement for upper elementary students?
2. What is the impact of receiving a full dose of Cignition's virtual small group math tutoring?
3. How does this impact differ by key student subgroups and programmatic context?

Methods: Quantitative; Experimental — This is a student-level randomized control trial.

Location: Ohio and New Jersey

Data: Implementation records from Cignition's systems; administrative data from the schools (demographics, test scores); survey data from teachers reporting on math instruction at their school

Implications/Findings: Our analysis found that being offered a Cignition tutoring slot results in a small positive, not statistically significant increase in math achievement. This is likely due to low dosage, with only 27% of the treatment group receiving the intended dose of 50+ hours of tutoring. One site (N=31) that had 93% of treatment students reach the 50-hour threshold had a statistically significant positive impact (ES = 0.49 SD), which suggests that Cignition's tutoring model can be effective if implemented as intended. Survey data from teachers indicated that treatment students received ~35 more minutes of total math instruction in a typical

week than control students. However, control students received more time with intervention teachers at their school. This may have also contributed to the null finding if school-based intervention teachers were able to give more tailored instruction than Cignition's tutors. We did not find any statistically significant differences in math achievement between implementation contexts or subgroups, with the exception of a statistically significant decrease among English language learners in the treatment group (N=16). Our findings underscore that implementation and dosage are important for the success of tutoring programs. It also highlights the importance of understanding the control condition when interpreting study results.

Link to Pre-Registration: https://osf.io/sftvb/?view_only=f1f2ef0d1cc54e8cb71b31c4df4047f7

Expanded Evaluation of Reading Partners Tutoring Programs

Robin Jacob, University of Michigan; Catherine Asher, University of Michigan; Carolyn Hill, MDRC

Status: Initial analyses done but not finalized

Motivation: Reading Partners offers one-on-one tutoring to thousands of students in grades K-4 across the country. Reading Partners offers tutoring at school-based reading centers via in-person instruction (RPT) and online through the Reading Partners Connects (RPCx) program. Both in-person and online instruction are delivered by trained volunteer tutors and utilize Reading Partners' evidence-based curriculum. A prior rigorous evaluation has shown that Reading Partners' traditional in-person program is an effective literacy intervention (Jacob et al, 2016; Reichardt et al., 2017). This study will provide evidence of the impacts of the RPCx program, in addition to the original RPT model. The study was designed as an RCT. Because of a variety of recruitment challenges and concerns about lack of power, we revised the design to a 2-year regression discontinuity. The RD allows us to expand our sample to include most of the districts that Reading Partners serves, relying primarily on RP program data rather than direct assessments.

Research Questions:

1. What is the effect of one year of Reading Partners Traditional tutoring (RPT) on student reading performance compared to business-as-usual reading instruction?
2. What is the effect of one year of Reading Partners Connects tutoring (RPCx) on student reading performance compared to business-as-usual reading instruction?
3. We are also conducting an implementation and cost study.

Methods: Quantitative; Quasi-Experimental — Impacts are estimated using a regression-discontinuity, based on eligibility for Reading Partners programming; the implementation study includes interviews and quantitative analysis of existing programmatic data; the cost study will use the ingredients method.

Location: National

Data: Reading Partners' student end of year reading assessments (Star); other Reading Partners' administrative data including program participation data and student demographics; direct assessments (SAT-10, AIMSweb, TOWRE-II) for a subset of students where IRB approval and parental consent were

already obtained under the RCT design; implementation data including tutoring session observations, interviews (including RP staff, tutors, and school personnel); ingredients-based cost data.

Implications/Findings: Preliminary findings from the first year of data collection indicate positive student impacts for both online and in-person one-on-one tutoring to support elementary reading. Online tutoring may be a viable alternative for schools needing to provide support to students in elementary reading.

Link to Pre-Registration: https://osf.io/m42eg/?view_only=c57bc62bbb0d4cdf95c510517b486a2e

An Evaluation of Carnegie Learning's Small Group Virtual Math Tutoring for Algebra I Students

Catherine Asher, University of Michigan; Robin Jacob, University of Michigan

Status: In process

Motivation: The study will evaluate the effectiveness of Carnegie Learning's small-group virtual Algebra I tutoring for middle and high school students.

Research Questions:

1. What is the impact of Carnegie Learning's high-impact math tutoring program on the math achievement for middle and high school students immediately after completing tutoring?
2. What is the impact of receiving an "adequate dosage" of Carnegie Learning's virtual small group math tutoring?
3. How does the impact of Carnegie Learning small group tutoring differ for specific subgroups of students, including gender, students below grade level in math, those with IEPs, English learners, and (if available) economically disadvantaged students?
4. What is the delayed effect of Carnegie Learning math tutoring on 8th grader's end-of-course state assessments?
5. What is the spillover effect of Carnegie Learning on other academic outcomes such as attendance (as available)?

Methods: Quantitative; Experimental — The study will use a student-level randomized control trial, stratified within a designated "What I Need" (WIN) period for the middle school students and by grade level for high school students. Effectiveness will be evaluated using both standardized benchmark assessments and the state's end-of-course Algebra I assessment.

Location: Texas

Data: We will draw on school administrative records and data from Carnegie Learning's proprietary management information system, including a standardized math assessment administered at the end of the tutoring semester, to evaluate the program's effectiveness.

Implications/Findings: None to date

Link to Pre-Registration: https://osf.io/m8xjq/overview?view_only=5f8682bfecc34a2984b9e37be7f6a6ec

Can High-Dosage Literacy Tutoring Put First-Graders on the Road to Success?

Monica P. Bhatt, University of Chicago; Jonathan Guryan, Northwestern University; Salman Khan, University of Chicago; James S. Kim, Harvard University

Status: Analyses finalized and being written up

Motivation: The study aims to evaluate the effectiveness of high-dosage tutoring in improving literacy skills for first-grade students, with a focus on exploring how to optimize and scale this approach. Prior research has shown high-dosage tutoring to be promising for improving student learning, particularly for early elementary literacy skills.

Research Questions: What is the end-of-school-year effect of this high-intensity tutoring program on Northwest Evaluation Association Measures of Academic Progress (NWEA MAP) reading test scores among first grade students? What is the effect of this high-intensity tutoring program on NWEA reading test scores one-year post-intervention and two-year post-intervention among first grade students? What are the end-of-school-year, one-year post-intervention, and two-year post-intervention effects of this high-intensity tutoring program on NWEA reading subtest scores among first grade students? How do outcomes differ among students who receive the program during the first year of program implementation (AY2016-17) versus the second year of program implementation (AY2017-18), if at all? What are the end-of-school-year, one-year post-intervention, and two-year post-intervention effects of this high-intensity tutoring program on NWEA math test scores and attendance rates among first grade students?

Methods: Quantitative, Qualitative, Mixed Methods, Experimental

This study employs a randomized controlled trial design examining the effects of high-dosage literacy tutoring for first-grade students. Students were randomly assigned within classrooms to either receive daily 45-minute tutoring sessions in pairs (treatment) or continue with regular classroom instruction (control). The tutoring intervention used the Leveled Literacy Intervention curriculum developed by Fountas & Pinnell. In addition to quantitative analysis, the study included 167 structured classroom observations of both treatment and control settings to document implementation and identify mechanisms. Primary analyses used OLS regression to estimate both intent-to-treat and treatment-on-the-treated effects, with students tracked for up to two years post-intervention to assess longer-term impacts.

Location: The study took place in Chicago, Illinois, in two charter schools within the LEARN Charter School Network.

Data: The study used:

- Student-level administrative data from Chicago Public Schools
- NWEA MAP assessment data (reading and math scores)
- Classroom observation data (167 observations total)
- Attendance data from Saga Education

Implications/Findings: https://osf.io/by9z5/ The study found:

- Pooled samples showed modest positive effects on reading outcomes (0.13-0.14 SD improvement)
- Cohort 1 (2016-17) showed stronger, statistically significant effects (0.19-0.21 SD)
- Cohort 2 (2017-18) showed smaller, non-significant positive effects (0.08 SD)
- Effects for Cohort 1 persisted and even increased one year post-intervention
- The intervention was particularly effective for foundational reading skills
- No significant effects were found on math outcomes
- Qualitative observations showed higher on-task behavior (93% vs 75%), more connected text reading (97% vs 60%), and more scaffolding (98% vs 51%) in treatment vs control classrooms

The research suggests that high-dosage tutoring can be effective for improving early literacy skills, with effects potentially persisting over time. The difference in outcomes between cohorts highlights the importance of implementation quality, particularly regarding instructional practices like scaffolding. The study contributes to understanding how to optimize tutoring interventions to improve academic outcomes for students in underserved communities, with the ultimate goal of closing educational disparities.

Saga Literacy

Monica Bhatt, University of Chicago Education Lab; Salman Khan, University of Chicago Education Lab; Jon Guryan, Northwestern University, University of Chicago Education Lab

Status: Analyses finalized and being written up

Motivation: Understanding effects of classroom level tutoring for 1st graders using Fountas-Pinnell Curriculum, but also examine mechanisms for why tutoring works through observations.

Research Questions: What is the effect of Saga's 1st grade literacy tutoring on student outcomes? What are the mechanisms that might lead to tutoring effects, based on classroom observations?

Methods: Quantitative, Mixed Methods, Experimental

Location: Chicago, IL

Data: Chicago Public Schools administrative data; classroom observations collected by the research team

Implications/Findings:

Experimental Evidence on the Effects of Academic Remediation on Violence

Monica Bhatt, The University of Chicago; Jens Ludwig, The University of Chicago; Matteo Magnaricotte, The University of Chicago; Fatemeh Momeni, The University of Chicago

Status: In process

Motivation: The pandemic has already caused an increase in gun violence, alongside dramatic learning losses for children of all ages across the country. These losses are concerning for future public health outcomes because of the strong relationship between education and violence (Lochner, 2020). By improving schooling outcomes and increasing attendance, tutoring can 1) increase the opportunity cost of crime through potential labor market returns and 2) reduce the time that students spend in the streets where they are more likely to be involved with violence. Tutoring, which has already proven effective at accelerating learning (Nickow et al., 2024), can then become a policy to reduce violence.

Research Questions: What are the long-term (10-year) impacts of being offered a chance to participate in and participating in a high-dosage tutoring intervention delivered in ideal conditions on violence? What are the short-term (2-year) impacts of being offered a chance to participate in and participating in a high-dosage tutoring intervention delivered at scale on violence?

Methods: Experimental, In process

Through a partnership between the University of Chicago, Chicago Public Schools (CPS), Chicago Police Department (CPD) and Chicago Fire Department (CFD), we aim to use several randomized control trials (RCTs) delivering high-dosage tutoring interventions, two of which were carried out in 2013-15 under ideal field conditions (efficacy RCTs) and an additional larger-scale trial that is currently underway (effectiveness RCTs) to produce short- and longer-term experimental evidence of the violence-reducing effects of tutoring delivered at different scales to answer our research questions.

Location: Illinois

Data: This study uses administrative data from the Chicago Public Schools (CPS) and the Chicago Police Department (CPD).

Implications/Findings: Findings are not available yet. Results showing the effectiveness of tutoring in reducing violence would support the use of this policy to address a major public health crisis and provide experimental evidence of additional benefits of tutoring that haven't been documented so far. At a more basic level, they would also strengthen the body of evidence connecting educational outcomes to violence and crime. We expect complete results to be available in the Fall of 2026.

Personalized Learning Initiative (PLI) – Impact Study

Monica P. Bhatt, University of Chicago Education Lab; Terence Chau, University of Chicago Education Lab; Matteo Magnaricotte, University of Chicago Education Lab; Fatemeh Momeni, University of Chicago Education Lab; PLI Research Team, University of Chicago Education Lab, MDRC, University of Toronto, Stanford University

Status: In process

Motivation: The global pandemic was a once-a-century public health crisis that left us with a once-a-century public education crisis. Despite the $189.5 billion that the federal government sent to schools through the Elementary and Secondary School Emergency Relief (ESSER) Fund, we still see students struggling to catch

up. To that end, we launched the Personalized Learning Initiative (PLI) in 2021, and have since partnered with eight education agencies around the country. The goal of the PLI is to understand whether and how we can scale the benefits of high dosage tutoring (HDT) such that more students might benefit. The eight site partners implemented over 20 different tutoring models during the last four years, reflecting adaptations to local conditions and providing numerous policy insights.

The specific goal of this work is to estimate the impact of HDT models with standard, evidence-based parameters, implemented across all partner sites through a randomized controlled trial. Alongside these traditional HDT models, partner sites implemented more sustainable models that reduced the cost per student by introducing different combinations of educational technology and teaching strategies. These sustainable HDT models (SHDT) will be similarly evaluated.

Research Questions: What are the impacts of being offered a chance to participate in and participating in a high-dosage tutoring intervention or a sustainable high-dosage tutoring intervention on students' tutored subject skills?

Methods: Quantitative, Experimental

We use an RCT to evaluate the impact of two tutoring approaches as described in our pre-analysis plan.

Location: Chicago Public Schools (Illinois); Fulton County Schools (Georgia); New Mexico Public Education Department; Miami-Dade County Public Schools (Florida); Greenville Public Schools (South Carolina); Winston-Salem/Forsyth County Schools (North Carolina); Guilford County Schools (North Carolina); Rocketship Public Schools (California)

Data: The study relies on administrative data collected by partner sites measuring student learning, tutoring attendance, and baseline characteristics of participating students. Learning is measured through a combination of end-of-year short-cycle assessments and summative assessments.

Implications/Findings: https://osf.io/fkjmn/ The study will provide experimental evidence for the effectiveness of high dosage tutoring and alternative, more sustainable models implemented at scale across the nation in a variety of sites with different characteristics. If the sustainable model were to produce a higher benefit-cost ratio, it would provide a solution to bring tutoring to the large number of students who need it. Preliminary and intermediate results are released on our website and disseminated as available. https://educationlab.uchicago.edu/projects/personalized-learning-initiative/

Personalized Learning Initiative (PLI) – Cost Study

Rebecca Davis, MDRC; Jean Grossman, MDRC & Princeton; Monica P. Bhatt, University of Chicago; PLI Research Team, University of Chicago Education Lab and MDRC

Status: Initial analyses done but not finalized

Motivation: The goal of the PLI is to understand whether and how we can scale the benefits of high dosage tutoring such that more students might benefit. PLI includes implementation, impact, and cost studies, as well as an effort to explore the personalized treatment effects (or heterogeneous treatment effects) of these

interventions. PLI is led by the University of Chicago Education Lab in collaboration with MDRC, as well as researchers from the University of Toronto and Stanford University.

The cost study specifically looks at the costs of 25 different formats of tutoring and estimates the full societal costs of those programs.

Research Questions: What are the costs of tutoring programs implemented in the Personalized Learning Initiative?

Methods: Quantitative, Experimental

The cost study follows the Ingredients Method (Levin et al. 2017) to inventory the suite of resources that are used to implement tutoring programs. These resources include resources paid for by school systems as well as those reallocated to the tutoring programs. By using nationally generalizable prices, the resulting cost estimates will be relevant to a variety of stakeholders and can be leveraged to replicate successful programs and to gauge the relative efficiency of different tutoring formats using a cost-effectiveness analysis

Location: Chicago Public Schools (Illinois); Fulton County Schools (Georgia); New Mexico Public Education Department; Miami-Dade County Public Schools (Florida); Greenville Public Schools (South Carolina); Winston-Salem/Forsyth County Schools (North Carolina); Guilford County Schools (North Carolina); Rocketship Public Schools (California)

Data: Survey data, interview data, administrative data, program records, observations, and document analysis

Implications/Findings: The cost study is embedded in the pre-registered impact study found: https://osf.io/fkjmn/files/osfstorage

The MDRC PLI project page includes a number of products based on the implementation research and the insights from the PLI technical assistance teams
https://www.mdrc.org/work/projects/personalized-learning-initiative

This is an unprecedentedly large cost study of tutoring programs and will contribute a detailed understanding of the breadth of tutoring models common in K-12 schools in the US.

Once Efficacy Study Kindergarten, 2024–2025: Comparing the typical intervention toolkit to students who received at least 80 sessions of Once

Juliette Berg, LXD Research, Sofia Jimenez, LXD Research; Rachel Schechter, LXD Research

Status: Analyses finalized and being written up

Motivation: Many students enter Kindergarten below grade level in literacy, and schools need evidence-based interventions to address early reading gaps. This study was motivated by the need to understand whether the Once tutoring program — which provides daily, scripted, one-on-one reading instruction delivered by school support staff — produces measurable literacy gains for the students most at risk, particularly when delivered with sufficient dosage (80+ sessions).

Research Questions:

1. To what extent does Once tutoring impact student outcomes when they receive the program for 80+ sessions?
2. To what extent does Once tutoring improve the outcomes of students who start below 50th percentile when students receive 80+ sessions?
3. What impacts does Once have on different student subgroups?

Methods: Quantitative; Quasi-Experimental — This study uses a quasi-experimental design comparing 114 Kindergarteners who received 80 or more Once tutoring sessions to 454 comparison students. Students were selected for tutoring by their schools' own criteria. A random intercept model was used to estimate outcomes, controlling for fall reading scores, attendance, economic disadvantage, ELL status, SPED status, gender, and race/ethnicity.

Location: The study takes place in an urban school district in the Southeast United States across 11 schools. The specific state is not identified in the study summary.

Data: The study uses FastBridge Reading Composite Scores collected at the beginning and end of the 2024–25 school year. The sample includes 590 Kindergarteners below the 50th national percentile at the start of the year. Demographic data — including race/ethnicity, ELL status, special education status, and economic disadvantage — are also used as statistical controls.

Implications/Findings: Kindergarteners below the 50th percentile who completed 80 or more Once tutoring sessions scored meaningfully higher on the end-of-year FastBridge literacy assessment (mean score of 55) than comparison students (mean score of 51). This difference was statistically significant ($p < .001$) with an effect size of Hedges' $g = .46$, which translates to approximately 2.7 additional months of reading growth beyond typical Kindergarten progress. These findings suggest that high-dosage, one-on-one tutoring delivered by trained school support staff — rather than certified teachers — can produce meaningful early literacy gains for the students most at risk. The results add to a growing evidence base that structured tutoring programs, when implemented with sufficient frequency and fidelity, can meaningfully close early literacy gaps for economically disadvantaged students and English language learners.

Link to Pre-Registration: N/A

Link to Report/Brief: https://canva.link/t1josll18x6lq9m

A Randomized Controlled Trial Study of the Once Early-Reading Program in Multiple School Districts

Amanda J. Neitzel, Johns Hopkins University Center for Research and Reform in Education; Ashley Grant, Johns Hopkins University Center for Research and Reform in Education

Status: In process

Motivation: The proposed study will evaluate the impacts of the Once Early-Reading Program, a tutoring model designed to accelerate early literacy development by leveraging existing school paraprofessionals. Unlike many tutoring programs that rely on external providers or volunteers, Once trains paraprofessionals to deliver

daily 30-minute, one-on-one reading sessions, thereby exceeding high-impact tutoring standards (which typically involve three sessions per week).

Research Questions:

1. What are the demographic and academic characteristics of the kindergarten students participating in the study and the schools where the Once Early-Reading Program is implemented? What are the characteristics of teachers and paraprofessionals delivering instruction in the participating schools?

2. How do tutoring dosage and student engagement vary by student characteristics (e.g., gender, race/ethnicity, English learner status, special education classification, socioeconomic status), or school?

3. Does the Once Early-Reading Program, delivered by trained paraprofessionals, improve early reading outcomes for kindergarten students compared to standard (business-as-usual) instruction? Do the impacts of the Once program on student literacy outcomes differ by student demographic subgroups (e.g., gender, race/ethnicity, socioeconomic status, English language classification, special education classification, free or reduced-price lunch status, prior achievement)?

4. Are student literacy outcomes associated with implementation factors (e.g., tutoring dosage, session attendance, engagement levels)?

5. What is the cost-effectiveness of the program, expressed as the cost per standard deviation of literacy improvement achieved?

Methods: Mixed Methods, Experimental; The study employs a randomized controlled trial (RCT) to estimate the program's impacts on student literacy outcomes. Primary outcomes will include early reading assessments already in use within each district (e.g., i-Ready, Acadience Reading, MAP Growth, Amira Foundational Literacy), measured from the beginning to the end of the school year. In addition to examining confirmatory impacts on student achievement, the study will address exploratory questions about dosage and subgroup effects. A teacher survey will document the nature of literacy instruction in control classrooms, providing evidence on the treatment-control contrast. Finally, a cost analysis will estimate total and per-student program costs, as well as the cost-effectiveness of the model in improving early literacy skills.

Location: Indiana, Missouri, Georgia

Data: Our primary outcome is student achievement on the Texas statewide mathematics assessments administered in spring 2026: (i) STAAR Mathematics (grades 6–8) and (ii) Algebra I End-of-Course (EOC) for students enrolled in Algebra I (typically grade 9). Scores will be taken from official district files. The study will also administer a teacher survey that will document empirically the BAU conditions by collecting the types and amounts of mathematics instruction for students in both the treatment and control conditions.

Implications/Findings: Findings will inform both policymakers and practitioners seeking scalable, sustainable approaches to delivering high-impact tutoring.

Link to Pre-Registration: https://osf.io/9rquz/overview

A Randomized Controlled Trial (RCT) to study Accelerate's Evidence for Impact Grantee Math-A-Matics Tutoring

Amanda J. Neitzel, Johns Hopkins University Center for Research and Reform in Education; Steven Ross, Johns Hopkins University Center for Research and Reform in Education; Ashley Grant, Johns Hopkins University Center for Research and Reform in Education

Status: In process

Motivation: NA

Research Questions:

1. What is the scheduled tutoring dosage (sessions, minutes, lessons) for students assigned to Math-a-Matics Tutoring, and what is the actual dosage achieved? How do tutoring dosage and student engagement vary by student characteristics (e.g., gender, race/ethnicity, English learner status, special education classification, socioeconomic status), or school?
2. Does Math-a-Matics Tutoring improve mathematics outcomes for secondary students in grades 6-9 compared to business-as-usual instruction?
3. Do the impacts of Math-a-Matics Tutoring on student mathematics outcomes differ by student demographic subgroups (e.g., gender, race/ethnicity, socioeconomic status, English language classification, special education classification, free or reduced-price lunch status, prior achievement, grade level)?
4. Are student mathematics outcomes associated with implementation factors (e.g., tutoring dosage, session attendance, engagement levels)?
5. Do the impacts of Math-a-Matics Tutoring differ when delivered virtually versus in-person?

Methods: Experimental, Quantitative; The study is employing a three-arm RCT conducted in two stages, randomizing students to virtual and in-person, with tutors teaching both virtually and in person.

Location: Texas

Data: Our primary outcome is student achievement on the Texas statewide mathematics assessments administered in spring 2026: (i) STAAR Mathematics (grades 6–8) and (ii) Algebra I End-of-Course (EOC) for students enrolled in Algebra I (typically grade 9). Scores will be taken from official district files. The study will also administer a teacher survey that will document empirically the BAU conditions by collecting the types and amounts of mathematics instruction for students in both the treatment and control conditions.

Implications/Findings: Findings will inform both policymakers and practitioners seeking scalable, sustainable approaches to delivering high-impact tutoring.

Link to Pre-Registration: https://osf.io/9kpyg/overview

Heart Math Tutoring's Expanded Model: A Randomized Evaluation of Volunteer Tutoring

Amanda J. Neitzel, Johns Hopkins University Center for Research and Reform in Education; Nathan Storey, Johns Hopkins University Center for Research and Reform in Education

Status: Complete

Motivation: This randomized controlled trial (RCT) evaluates the impact of Heart Math Tutoring (HMT), a school-day intervention providing targeted math support to students in grades 1–5 who are performing below grade level.

Research Questions:

1. What are the one-year effects of HMT's Expanded Model on mathematics achievement for students performing below grade level, compared to students receiving business-as-usual teaching?
2. How do the effects of HMT's Expanded Model differ by gender, ethnicity, English learner status, special education participation, grade level, or program design?
3. To what extent is the fidelity of implementation associated with better student outcomes?

Methods: Quantitative; During the 2024-25 school year, seven elementary schools in a district in the southeastern United States took part in a randomized controlled trial to evaluate the impact of HMT's Expanded Model on mathematics outcomes for first through fifth grade students. At all schools, students were identified as needing additional mathematics intervention. These students were stratified based on school and grade, and then randomly assigned within the strata to participate in HMT's Expanded Model (treatment) or continue with their business-as-usual classroom instruction (control). Following an intent to treat (ITT) approach, students were included in their assigned group, regardless of whether they received tutoring.

Location: North Carolina

Data: Data sources and measures for the current study included student achievement data and Heart Math Tutoring dosage data.

Implications/Findings: On the broad, standardized i-Ready diagnostic, impacts were essentially nil: adjusted endline i-Ready scores for treatment and control students were nearly identical, yielding standardized effect sizes close to zero and not statistically significant. The North Carolina EOG, available only for grades 3–5, produced a small positive effect favoring the treatment group (+0.18 SD). In contrast, HMT produced large, statistically significant impacts on the curriculum-aligned Assessing Math Concepts measures. The results indicate that students who received a high dosage of Heart Math Tutoring demonstrated the strongest performance across all measures. The most striking pattern in these results is the contrast between clear, substantial gains on the program-aligned Assessing Math Concepts (AMC) measure and the near-zero impacts on the broad, adaptive i-Ready diagnostic. Time, dose, and developmental sequencing appear to matter for transfer.

A Mixed-Methods Evaluation of ELP's Learn Strategy with Oakland Unified School District

Amanda J. Neitzel, Johns Hopkins University Center for Research and Reform in Education); Nathan Storey, Johns Hopkins University Center for Research and Reform in Education); Joseph Reilly, Johns Hopkins

University Center for Research and Reform in Education); Xue Wang, Johns Hopkins University Center for Research and Reform in Education; Jennifer Krajewski, Johns Hopkins University Center for Research and Reform in Education

Status: In process

Motivation: In addition to examining confirmatory impacts on student achievement for Tier 3 tutoring, the study will address exploratory questions about dosage and subgroup effects.

Research Questions:

1. How does tutoring operate within existing school structures?
2. What is the impact of Tier 3 tutoring on student outcomes? How do impacts vary across schools, grades, and student subgroups?
3. Are there patterns in implementation (e.g., dosage, scheduling) that correlate with stronger outcomes?
4. What is the impact of Tier 2 supports on student outcomes?
5. Are there patterns in student need, response, or growth that could inform more targeted and efficient use of MTSS resources?

Methods: Mixed Methods, Experimental; During the 2025-2026 school year, schools in OUSD are taking part in a randomized control trial to evaluate the impact of MTSS Tier 3 supports on literacy outcomes for 1,942 students. In participating schools, students were randomly assigned within schools to participate in MTSS Tier 3 support consisting of 1:1 virtual tutoring (treatment) or to continue with their business-as-usual classroom instruction and MTSS Tier 2 literacy supports (control).

Location: California

Data: Reading assessments already in use within each district (i-Ready & DIBELS), student and school demographic characteristics, program dosage data, tutor surveys, teacher and principal surveys, classroom observations and site visits.

Implications/Findings: Mid-year impact results are being analyzed. We are compiling principal, teacher, and teacher coordinator perceptions of program implementation. The findings of this study will inform district-wide implementation and tutoring policy related to Tier 3 and Tier 2 tutoring programs, providing detailed understanding of program implementation, impact, and the intersection between Tier 3 and Tier 2 tutoring, including understanding of how student outcomes vary across models or combinations of support, how schools and staff coordinate tutoring support for individual students, and patterns of student need, response, or growth. The results of this study will inform more targeted and efficient use of MTSS resources.

New Jersey Tutoring Corps Evaluation

Greg Chojnacki, Mathematica; Menbere Shiferaw, Mathematica; Lindsay Fox, Mathematica; Anna Gu, Mathematica

Status: In process

Motivation: This study will contribute to this growing body of research by estimating how a high-dosage tutoring program can improve outcomes for students in New Jersey, as students continue to recover from pandemic learning losses.

Research Questions: What are the impacts of NJTC on students' math achievement? What are the impacts of NJTC on students' literacy achievement? What are the impacts of NJTC on school attendance? Do tutoring impacts vary by students' baseline achievement levels, grade level, or number of tutoring sessions that students attend? What are students' and tutors' perceptions of the tutoring sessions?

Methods: Quantitative,Quasi-Experimental. We will use a difference-in-differences approach to compare NJTC students to students who did not participate in NJTC, before and after the NJTC program began in each school. We will estimate the version of this model developed by Callaway and Sant‚ÄôAnna (2021) that requires less restrictive assumptions to recover the average effect of treatment on the treated students.

Location: New Jersey

Data:

1. New Jersey Student Learning Assessment (NJSLA) in mathematics
2. NJSLA in English Language Arts (ELA)
3. School attendance rate
4. To understand students' and tutors' perceptions of the program, the study team will also construct measures based on survey and attendance data provided by NJTC.

Implications/Findings: NA - currently in data collection

Link to preregistration: NA

Arkansas Statewide Rural Literacy Study: Testing Evidence-Based High-Impact Tutoring at Scale

Susanna Loeb, Stanford University; Nancy Waymack, Stanford University

Status: Beginning Fall 2026

Motivation: This project will rigorously test an evidence-based virtual tutoring model (Air Reading) across 100 diverse Arkansas rural schools through a large-scale RCT.

Research Questions: Do evidence-based programs from other states work in Arkansas' context—with different standards, assessments, and curriculum? How do we effectively scale a proven model across 100 diverse rural schools—and what changes are needed to make the model work reliably? How do we invest early to ensure struggling first-graders reach proficiency by 3rd grade—the critical gateway to all learning? Given severe rural staffing shortages, can college students trained in Science of Reading deliver literacy gains comparable to experienced paraeducators—while exploring whether tutoring experience increases teaching interest?

Methods: Quantitative,Experimental. The evaluation is a multi-site, school-level RCT with intent-to-treat estimation and hierarchical models respecting nesting of students within schools.

Location: Arkansas

Data:

1. State reading assessments
2. School attendance rate
3. Tutoring platform logs, session logs
4. Coach observation rubrics
5. Platform attendance logs

Implications/Findings: NA - currently planning

Link to preregistration: NA

EFFECTS OF PROGRAM CHARACTERISTICS

Early relational investments improve tutoring attendance at scale

Carly Robinson, Stanford University; Lal Chadeesignh, Behavioural Insights Team (UK); Giulia Tagliaferri, Behavioural Insights Team (UK); Cristina Barnard Gonzalez, Stanford University; Todd Rogers, Harvard University; Hunter Gehlbach, Johns Hopkins University

Status: Analyses finalized and being written up

Motivation: Tutoring attendance is a persistent challenge as programs scale. While structural design choices can help, few experimental studies have tested concrete, scalable strategies tutors can use to build relationships with students early in a program. These relationships that may sustain engagement over time.

Research Questions: Does exposing tutor–student dyads to their shared interests and characteristics at the outset of tutoring increase students' session attendance?

Methods: RCT

Location: UK

Data: Administrative attendance records from the UK's National Tutoring Programme (Spring 2021); proportion of scheduled sessions attended for 701 tutors and 8,923 students across 4 tutoring providers

Implications/Findings: Students in the treatment condition attended 4.1 percentage points more of their scheduled sessions than controls. Effects were concentrated in in-person settings (6.1-pp increase) and were small and non-significant for virtual tutoring (0.7-pp). Positive effects extended to economically disadvantaged students.

Early relational investments appear to expand total instructional time — the time spent on connection at the outset is more than recovered through improved attendance. Programs should treat relationship-building

activities as a core component of program design, not an add-on. Virtual programs likely need more deliberate strategies to achieve comparable effects.

Link to preregistration: https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/NTP-RCT1-—-Leveraging-Similarity-to-Improve-Pupil-Attendance-Protocol_SAP_2021-09-06-104614_kdqt.pdf?v=1775154297

Enhancing Student Engagement in Virtual Tutoring Sessions: A Randomized Control Trial of a Tutoring Intervention

Susanna Loeb, Stanford University; Carly Robinson, Stanford University; Erin Devers, Indiana Wesleyan University; Emma Strouse, Stanford University; Ana Ribeiro, Stanford University

Status: Initial analyses done but not finalized

Motivation: Prior correlational analyses by the provider suggest that students who earn higher average tutor-awarded participation points, given to students for demonstrating effort in sessions, experience greater growth on end-of-year (EOY) standardized tests. This experiment explores whether training tutors on how to provide participation-oriented feedback has a causal impact on EOY student outcomes.

Research Questions: Does receiving additional training on how to administer more specific feedback for participation points lead to an increase in student engagement, measured by participation points? Does receiving additional training on how to administer more specific feedback for participation points lead to an increase in student academic achievement, measured by Spring NWEA MAP scores?

Methods: RCT, NLP

Location: Texas

Data: Tutoring session-level data and school administrative data

Implications/Findings: This intervention provides a launchpoint in exploring strategies for increasing student engagement in virtual tutoring programs. Among our initial findings are the following:

At the student level, small but consistently positive effects on participation points suggest it is possible to manipulate student engagement through tutor training intervention; however, we lacked the statistical power to pick up effects on end-of-year math scores. While the estimates are positive and in line with the correlational relationship between participation and learning, they are not statistically different from zero.

At the session level, we found significant increases in student participation within a session and significant positive effects for sessions with treatment students and control tutors suggesting students changed their behavior after exposure to the treatment.

We are in the process of using natural language processing on the tutors session chat to gain better insight into how exposure to the intervention affects teacher feedback, including the language used to indicate when and points are awarded.

Link to preregistration: https://osf.io/7g486/overview

Individualized Tutor Coaching vs. Professional Learning Communities (PLCs) on Tutor and Student Outcomes

Cristina Barnard Gonzalez, Stanford University; Susanna Loeb, Stanford University; Cynthia Pollard, UnboundED; Carly Robinson, Stanford University; Amirpasha Zandieh, University of Colorado, Boulder

Status: Analyses finalized and being written up

Motivation: Tutoring is an effective educational intervention, especially in early literacy, where tutors who are formally trained teachers demonstrate superior performance compared to those without formal training (Nickow et al., 2024). While employing trained teachers as tutors is beneficial, it is also costly and logistically complex. In this study we explore whether one-on-one coaching for tutors can effectively improve tutor quality compared to professional learning communities (PLCs). Previous studies have demonstrated that both coaching and PLCs may enhance educators' instructional skills (Kraft et al., 2018) and improve both program quality and retention rates (Worthy et al., 2003; White et al., 2023). We hypothesize that tutors, like their teacher counterparts, will benefit from one-on-one coaching.

Research Questions: 1. Are tutors assigned to receive personalized 1:1 instructional coaching more likely to persist as a tutor than tutors who worked in PLCs?

2. Do students assigned to receive virtual early literacy tutoring from a tutor who received personalized 1:1 instructional coaching improve more on their end-of-year reading test scores than students assigned to a tutor who worked in PLCs?

Methods: RCT

Location: Texas

Data: Student administrative data, student reading scores from tutoring provider, tutor administrative data from tutoring provider, coaching attendance data, session-level data

Implications/Findings: Our analysis suggests that one-on-one coaching increases tutor retention over the course of the year by 19 percentage points compared to the PLCs. However, we did not find evidence that the one-on-one coaching was more effective as measured by students' end-of-year DIBELS scores, and may have even been less effective than PLCs.

Tips4Tutors: A Text Messaging Curriculum for Tutors

Sophie Barnes, Yale University; Susanna Loeb, Stanford University; Cynthia Pollard, UnboundED; Carly Robinson, Stanford University; Helen Zhou, Stanford University, Jilli Jung, Stanford University

Status: Initial analyses done but not finalized

Motivation: Text messaging and online platforms are low-cost, scalable strategies that can provide tutors with actionable, real-time information for supporting students' social-emotional development. We are evaluating a

curriculum that draws on research and best practices in the field of educational psychology and social emotional learning in the design of the messages. We hypothesize that the texting program and access to an online platform will positively impact tutor and student outcomes, including the tutoring experience, tutor-student relationships, student social-emotional competencies, student course performance and completion, attendance, and other positive school behaviors.

Research Questions: Did tutors and their students in the Tips4Tutors condition report greater self-efficacy and tutor-student relationships than tutors and their students in the control condition? Did students of tutors in the Tips4Tutors experience greater improvement in academic performance and school attendance than students of tutors in the control condition?

Methods: RCT

Location: North Carolina, Other US States

Data: Student administrative data, tutor and student perception survey data, tutor demographic data.

Implications/Findings: We conducted this intervention across several semesters in multiple sites. Initial analyses reveal statistically significant treatment effects on tutors' relational self-efficacy. Overall, we see similar treatment effects across subgroups like tutor type, school grade level, and tutoring by subject. Text-messaging programs like Tips4Tutors have the potential to provide effective support and training to tutors. Forthcoming analyses will provide additional insights into the student experience and potential impacts on student outcomes.

Link to preregistration: https://osf.io/v2my6/registrations

Differences in Tutoring Effectiveness across Tutor Characteristics

Jilli Jung, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: In process

Motivation: While tutor quality and effectiveness are key for a successful tutoring program, research on variation in tutor effectiveness and its relationship with tutor characteristics is scarce. Building on the large body of literature on teacher effectiveness, this study explores differences in tutoring effectiveness across tutor characteristics.

Research Questions: To what extent do tutors vary in their effectiveness in tutoring? What are the characteristics of effective tutors?

Methods: RCT

Location: US

Data: Student administrative data, tutoring attendance data, tutor and student perception survey data, tutor demographic data.

Implications/Findings: Preliminary analysis shows meaningful between-tutor variation in students' reading test scores and tutoring engagement, especially lateness to tutoring sessions. While broad characteristics of tutors do little to predict student outcomes, hiring scores are somewhat more informative, but mainly for outcomes closely related to the dimensions they were designed to assess.

Link to preregistration: https://osf.io/r3yfv/overview

Tutor-to-Teacher Pipeline: Tutoring Experience and the Likelihood of Pursuing a Teaching Profession

Jilli Jung, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: In process

Motivation: Tutoring programs provide their tutors with substantial teaching experience. Research has suggested that gaining experience in teaching and working with students is one of the most effective ways to introduce prospective educators to the classroom, encouraging them to enter the teaching profession. This study examines to what extent the experience of being a tutor is related to the likelihood of pursuing a teaching career.

Research Questions: Whether and how does having tutoring experience influence the likelihood of pursuing a teaching profession?

Methods: Quasi-experiment - Regression Discontinuity

Location: US

Data: Application data on tutor fellowship, tutor demographic data.

Implications/Findings: Preliminary analysis shows that tutoring experience increases the likelihood of applying to the teacher training program by 12 percentage points—nearly tripling the baseline rate. Effects are largest among men and applicants of color, suggesting that tutoring experience is more beneficial to groups who are historically underrepresented in the teacher workforce.

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

Elizabeth Huffaker, University of Florida; Bess Markel, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: Preliminary test-based analyses shared with partner; revised analysis and addition of survey data in progress

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 Question: Does same-race matching of tutors to students - especially for Black tutors and students - improve the effectiveness of high impact tutoring on students' achievement, attendance and views of themselves as learners? 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: RCT

Location: Multiple districts and schools across two US states.

Data: Student-level administrative data, tutor trait survey, achievement data provided by the tutoring company.

Implications/Findings: Findings do not support a clear tutor student race-match effect on standardized test scores. We were unable to explore gender effects, due to the low number of male tutors in the sample. However, additional exploratory analyses of additional tutor characteristics suggest that consistent tutoring with the same provider is associated with improved student reading outcomes, and that tutors currently in college are as effective as those with a BA or higher.

Link to preregistration: https://osf.io/2kfzh/overview

Examining the Effectiveness of Teacher-led Virtual Tutoring on Elementary Literacy Skills

Hsiaolin Hsieh, Stanford University; Monica Lee, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: Completed preliminary analysis

Motivation: Meta-analytic work has shown that trained educators (teachers) tend to produce larger gains than paraprofessionals or volunteers (Nickow et al., 2024). However, implementation of teacher-led tutoring programs traditionally encounters two problems: first, working with a student individually or in a small group prevents a teacher from delivering whole-class instruction simultaneously. Second, if a school wishes to introduce a second certified teacher into the classroom to deliver tutoring, staffing can be challenging due to ongoing teacher shortages in many parts of the country (Edwards et al., 2024). In this study, we partner with a tutoring provider who provides teacher-led tutoring virtually; because virtual tutoring platforms can recruit teachers across the country, this widens the pool of qualified tutors available to students (Kraft & Falken, 2021).

Research Questions: The confirmatory research question is–"Does virtual, one-on-one tutoring delivered by certified teachers affect reading outcomes for students reading below grade level?" Additionally, we will address an exploratory research question regarding heterogeneity–"How does the effectiveness of the program vary by the underlying student characteristics (i.e., student gender, race/ethnicity, grade level, socioeconomic disadvantage, English learner status, receipt of SPED services, prior reading level, student engagement)?"

Methods: RCT

Location: US Mid Western

Data: Student administrative data; Student beginning-, middle-, and end-of-year reading scores (i-Ready Reading Diagnostic and the Missouri Assessment Program); Tutoring session data

Implications/Findings: The study analyzes data from 1,550 students across 14 elementary schools, focusing on grades 1–4. Findings show that the program had a positive and statistically significant effect, particularly for students who scored well below grade-level benchmarks on beginning-of-year assessments (MTSS Tier III students). Tier III students in the treatment group outperformed their peers in the comparison group by 0.08 standard deviations in end-of-year i-Ready Reading Overall Scale Scores. Among Tier III students who were assigned to and actively participated in tutoring, the effect was slightly larger (0.09 SD). Both effects were statistically significant at the $p < .05$ level. Additionally, the tutoring program showed a promising trend in improving academic performance and reducing the need for special education placement for the most struggling students, particularly among upper elementary students. Overall, integrating high-impact tutoring into the MTSS framework significantly enhanced reading outcomes for struggling readers. The findings suggest that schools can reduce disparities in early literacy and improve student learning at scale through structured, technology-supported tutoring without substantial additional cost.

Link to preregistration: https://osf.io/ngucz/overview

The Human Factor: A comparison of AI tutoring with and without a human tutor

Carly D. Robinson, Stanford University; David Gormley, Stanford University; Ana Ribeiro, Stanford University; Susanna Loeb, Stanford University

Status: Analysis complete, manuscript in progress

Motivation: AI tutoring is increasingly prevalent, but AI tutors cannot provide the emotional support, motivation, and accountability that can support learning. Hybrid human-AI tutoring aims to address this gap, but further research is needed to understand the impacts and costs of hybrid human-AI tutoring.

Research questions: Compared to students working independently on an AI reading platform, does being assigned to work with a human tutor increase students' overall use of the platform (in minutes used)? Compared to students working independently on an AI reading platform, does being assigned to work with a human tutor increase students' engagement (i.e., completing more stories) while using the platform? Compared to students working independently on an AI reading platform, does being assigned to work with a human tutor improve literacy and reading achievement?

Methods: RCT

Location: Two large American school districts (midwest and west)

Data: Student usage data (minutes used), student engagement data (stories read), student demographic and achievement data

Implications/Findings: The presence of a human tutor substantially boosted student interaction with the platform: in Study A, tutors increased platform usage by roughly 46 percent and engagement, measured by stories completed, by 72 percent. In Study B, usage increased by 85 percent and engagement by 80 percent. However, both studies had extremely low baseline usage, which means that these large percentage increases translated to less than two hours of additional platform use across the school year.

Link to preregistration: https://osf.io/fe8zg/overview

Powering Personalization: Can an intervention help tutors provide more personalized instruction and relationship building to students?

Carly D. Robinson, Stanford University; David Gormley, Stanford University; Susanna Loeb, Stanford University

Status: Data collection complete, analysis in progress

Motivation: The tutoring research field treats personalized instruction and relationship building as central components of effective tutoring, but there is relatively little research that measures these features or attempts to support tutors in developing these skills. This pilot aims to provide evidence about whether an intervention can support tutors in providing personalized instruction and relationship building.

Research questions: How do measures of personalization and relationship building differ across treatment (tutors who received the intervention) and control groups (tutors who did not receive the intervention)? Does the intervention have an impact on student outcomes?

Methods: RCT

Location: A large east coast school district

Data: Student and tutor audio/video/chat data, student and tutor surveys, student task data, student BOY/EOY achievement data, tutor intervention usage logs

Implications/Findings: TBD

A Tutor Like Me: Examining Tutor-Student Demographic Match in an Early Literacy Tutoring Program

Elizabeth Epstein, The City University of New York Graduate Center; Kathryn Pace Miles, Brooklyn College

Status: Completed

Motivation: This study, the continuation of a project presented last year, examines the role of tutor-student demographic congruence, including linguistic background, in the achievement and relational outcomes of remote and in-person high-impact early reading tutoring through the CUNY Reading Corps (CRC). The CRC, established in the fall of 2020 by Dr. Katharine Pace Miles, is an ongoing initiative that pairs CUNY students with New York City public school students for reading intervention using the research- and evidence-based Reading Go! and Reading Ready programs.

Research Questions: Did students achieve greater reading gains when tutors and students shared demographic characteristics, including a common home language background (1a) and/or common ethnic/racial background (1b)? Were student attendance rates for tutoring sessions higher for demographically congruent pairs? Did tutors in demographically congruent pairs report a stronger bond with their students? Did the impact of demographic congruence vary by modality (remote or in-person delivery)?

Methods: Quantitative

This study is a secondary analysis of a dataset collected by the CRC and Hashim (2024) in the summer of 2023. Student demographic data from the NYCDOE was triangulated with tutor-level demographic data from CUNY Central and tutoring outcome data from the CRC (reading assessment results, attendance, tutor-student bond, and tutor satisfaction). Congruence in each category was treated as a binary categorical variable, with tutoring partnerships coded as congruent (1) for shared home language or incongruent (0) for lack of shared home language; (1) for shared race/ethnicity or (0) for incongruent race/ethnicity. Hierarchical linear modeling (HLM) was used to explore the relationship between the academic and relational outcomes of tutoring (reading achievement, attendance, and bond as measured by closeness/conflict scores) and tutor-student demographic match while accounting for differences between tutors. Tutor GPA was also included in the analysis to account for its potential impact on the outcomes of interest.

Location: New York

Data: This study is a secondary analysis of a dataset collected by Hashim (2024) as part of an examination of the impact of instructional format (remote vs. in person) and tutor characteristics on outcomes of the CUNY Reading Corps (CRC) program in the summer of 2023.

Implications/Findings: Though differences were not statistically significant, on average, students in demographically congruent pairs achieved higher reading growth and attended more tutoring sessions than those in incongruent pairs. Further investigation is needed to determine whether demographic congruence should be considered in program implementation. A data-sharing agreement with New York City Public schools is in the final stages of being approved to receive access to student home language data; this will be used in an upcoming analysis on tutor-tutee language background congruence. Results lay the groundwork for primary data collection during the 2024-25 school year. Understanding best practices for tutor-student matching—and whether the implications of tutor matching practices varies by modality (in person or remote)—will help guide school leaders and policymakers as they implement high-dosage tutoring programs.

Link to study: https://academicworks.cuny.edu/gc_etds/6383/

Tutor Coach: Investigating the use and effectiveness of tutoring practices for students in online math tutoring with human and AI tutors

Julian Bernado, Stanford University; Ana Trindade Ribeiro, Stanford University; Carly Robinson, Stanford University; Xander Beberman, Stanford University; Susanna Loeb, Stanford University; Ryan Knight, Insource Services; Kyle Lo, AI2; Alexis Ross, AI2

Status: Just beginning

Motivation: NLP and ML methods have made it possible to identify and measure an increasing number of pedagogical strategies and markers of student engagement. However, we still know little about how effective these strategies are for student engagement and learning. Using NLP and ML methods, we define a series of measures of tutor moves and student engagement to be used as benchmarks across different tutoring platforms (human and AI) and evaluate their impact on student engagement and learning during and outside of the tutoring session in a series of studies.

Research Questions: What are pedagogical and relationship-building strategies commonly used by human math tutors? Which strategies used by human tutors are replicated by AI tutors in math tutoring? Which strategies are effective to increase student engagement in the tutoring session? Which strategies are effective to improve student learning outcomes?

Methods: Large Language Models, Machine Learning; Descriptive and RCTs studies

Location: National

Data: Tutoring sessions' audio and video recordings; student and tutor characteristics collected by the tutoring provider; internal measures of student learning progress.

Implications/Findings: This project will continue to develop measures of tutoring practices and student engagement and design RCTs to evaluate the effectiveness of these measures.

Evaluating the Effect of Virtual 2:1 Tutoring on Early Literacy Skills

Hsiaolin Hsieh, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: In progress

Motivation: This study builds on our previous study (2022-23SY), in which we found that virtual tutoring can significantly improve young learners' early literacy outcomes, but the 1:1 model was costly and relatively expensive to expand. In the 2025-26 SY, we partner again with the same collaborators to conduct another large-scale randomized controlled trial (RCT) that tests whether a 2:1 virtual model can maintain effectiveness while increasing scalability and reducing per-student cost.

Research Questions: Does a high-impact, virtual early literacy tutoring program delivered in pairs (2:1) lead to improved student outcomes?

Methods: RCT

Location: US South

Data: Student and tutor participation collected by the tutoring provider, student-level demographic and assessment data collected by the district partner.

Implications/Findings: Based on the middle-of-year (MOY) data, students assigned to tutoring outperformed their comparison peers by 0.09 standard deviations (SD) on DIBELS ($p < .01$). These gains are primarily driven by first grade: first-grade students in the treatment group scored 0.13 SD higher than their peers in the comparison group ($p < .01$). Kindergarten and second-grade estimates are positive but do not reach

conventional thresholds for statistical significance. When examining effects by baseline reading level, we find that students who were WBB at BOY scored 0.07 SD higher at MOY than their WBB comparison peers, with effects driven largely by students in higher grades. Estimates for BB students are positive but imprecise, likely reflecting the smaller sample size due to prioritizing WBB students for treatment assignment. These findings suggest the intervention is being implemented successfully and is generating impacts consistent with, or in some cases exceeding, published end-of-year estimates for virtual tutoring in early literacy, including several 1:1 models. If gains persist through the end of year, the results will provide rigorous evidence that a paired (2:1) approach can preserve effectiveness while improving scalability and reducing per-student cost. This evidence would be particularly valuable for districts/CMOs seeking to expand access to high-impact tutoring under realistic budget and staffing constraints. More broadly, the study contributes to the research base on how program design choices, such as tutor-student ratio and virtual delivery, shape the effectiveness of high-impact tutoring for early elementary learners.

Link to preregistration: https://osf.io/uf4vn/overview

Perceptions of AI-assisted tutor feedback: evidence from a professional development program for teachers in Uruguay

María Eugenia Curi, Ceibal; Déborah Zaki, Ceibal; Andreina Colmenares, Ceibal; Vanessa Martinez, Ceibal; Carly Robinson, Stanford University; Susanna Loeb, Stanford University; Cristina Barnard Gonzalez, Stanford University.

Status: Implementation

Motivation: The project aims to analyze whether students' perception of feedback on their assignments is influenced by knowing whether that feedback was generated with AI assistance or exclusively by a human tutor. Specifically, it seeks to explore how transparency about AI use in the grading process affects students' perceived value, trust, and usefulness of the feedback received.

Research Questions: What is the impact of students knowing that their tutor generated feedback with AI assistance on their confidence on the feedback and on perceptions of tutor effort in elaborating the feedback? What is the impact of students being informed that their tutor generated feedback with AI assistance on their academic performance?

Methods: RCT.

Location: Uruguay.

Data: Student survey data.

Evaluating the impact of an additional computational thinking class supported by a teacher-mediated AI chat tutor in primary education: evidence from sixth grade in Uruguay

Camila Porto Dellepiane, Ceibal; Victor Koleszar, Ceibal; Carly Robinson, Stanford University; Susanna Loeb, Stanford University; Cristina Barnard Gonzalez, Stanford University; Lily Fesler, Stanford University.

Status: Implementation

Motivation: We aim to explore the causal effects of a teacher-mediated AI chat tutor introduced as part of an additional computational thinking (CT) class for sixth-grade students in Uruguay. The intervention supplements regular instruction and classroom teachers actively guide students' interactions with the AI tutor during class time. The tutors provide structured prompts, feedback, and practice aligned with the national CT curriculum.

Research Questions: What is the impact of providing additional CT classes supported by a teacher-mediated AI tutor on students' CT learning outcomes and socioemotional outcomes (motivation/interest, self-efficacy, and growth mindset)?

Methods: RCT.

Location: Uruguay.

Data: Student assessments measuring computational thinking skills and student-level data on motivation and engagement.

Accelerating and Sustaining Text Reading Fluency Improvements

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

Status: Analyses finalized and being written up

Motivation: Fade-out is a well-documented phenomenon, but relatively little is known about 'sustaining factors'.

Research Questions: To what extent are students' pre-intervention literacy skills and student engagement during reading instruction associated with growth in TRF during a Tier 2 fluency intervention? To what extent do students' literacy skills and engagement during Tier 1 instruction predict the maintenance of fluency intervention effects over time?

Methods: Longitudinal description of initial and sustained effects. Data were collected from 2nd and 3rd grade students participating in the tutoring program. All students completed the assessment battery at the onset of tutoring; those who exited (due to progress data indicating grade-level skills) completed the assessment battery a second time. Text reading fluency data were used to estimate and analyze growth during and after exiting tutoring, with behavioral engagement data and foundational literacy skills data serving as predictors of growth slopes.

Location: Minnesota: 33 schools in the broader Twin Cities metro area, serving 742 students.

Data: Text reading fluency measures, Systematic Direct Observation of engaged behavior, Classroom literacy environment observations, GRADE-Group Reading Assessment and Diagnostic Evaluation [Measure of comprehension/language skills], TOWRE-Test of Word Reading Efficiency [Measure of word-level reading skills]

Implications/Findings: Engagement (as measured) did not appear to be related to growth (during or after receiving tutoring) in either grade. Word reading skills at onset of tutoring predicted growth rate during tutoring

for 2nd graders. Comprehension skills at onset of tutoring and cessation of tutoring predicted growth rates during and after tutoring; but comprehension skills at onset of tutoring did not predict post-tutoring growth.

Practicing How to Teach With AI: Do simulated tutoring sessions improve tutor instructional skills?

Ana Trindade Ribeiro, Stanford University; Paul Yoo, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: Conducting preliminary analysis

Motivation: Tutoring programs can accelerate student learning (Nickow et al., 2024), but experienced tutors are in short supply amid rising demand (Groom-Thomas et al., 2023). To address this challenge, we study the potential of AI-based tools to strengthen tutor supply and quality at scale. We partnered with a large on-line tutoring platform that connects volunteer tutors to students seeking tutoring support. Our partners developed an AI-powered tutor-training simulator that allows tutors to practice in a virtual classroom, where tutors interact—verbally and via chat—with simulated students (trained on large language models) around a target skill (e.g. eliciting students to talk). Tutors can rehearse instructional practices dynamically and receive immediate feedback from the simulator.

Research Questions: Does requiring a tutor-training simulator to potential volunteer tutors affect their likelihood of following through with volunteering? Does providing access to tutor-training simulators to volunteer tutors increase the use of recommended pedagogical practices during tutoring sessions with real students? Does providing access to tutor-training simulators to volunteer tutors lead to more active engagement from their students?

Methods: Quantitative, Experimental (RCT) – tutor-level randomization.

We randomly assigned tutors to different conditions based on their prior experience (first-time vs. returning). First-time tutors either completed a one-time AI training during onboarding or not, and some tutors (both new and returning) were also encouraged to keep using the AI training during their tutoring period.

Data: Transcript data from the tutoring session; survey data on tutors and students; record data on the session; web-traffic data on using the simulator

Implications/Findings: There are suggestive findings that requiring tutor-training simulators to first-time volunteers increased their likelihood of signing up as a tutor. Access and nudges to tutor-training simulator treatment may have increased the average use of student names in tutoring sessions but also decreased time spent speaking for both tutors and students.

Link to preregistration: https://osf.io/famhw/overview

University Tutoring Collaborative

Matt Kraft, Brown University; Sally Sadoff, UC San Diego; Jeffery Livingston, Bentley University; Katie Pace Miles, Brooklyn College, CUNY

Status: Just beginning

Motivation: We received a grant through the University of Chicago's Gracias Center for Human Development Incubator to support an RCT of The Reading Institute's University Tutoring Collaborative, in which education majors provide structured literacy tutoring using Reading Ready or Reading Go either as part of their coursework or as paid tutors through the university. The study will randomly assign K-3rd grade students to receive no treatment, human tutoring, human tutoring with AI, or AI alone.

Research Questions: Does constructive engagement with AI tools allow for a reduced role of human tutors in ways that maintain effectiveness and enable tutoring to scale to more students?

Methods: Quantitative, Experimental; This will be a randomized controlled trial design. Students' scores on their school administered standardized assessment will be used to randomly assign them to the four conditions. All students will receive an additional standardized literacy assessment pre/post treatment. For students in the treatment conditions, high-impact tutoring in the Reading Ready or Reading Go programs will be delivered by university education majors for a minimum of 20 sessions. Students in the treatment conditions with an AI tool will also use either Amira, Magpie, or ClassDojo to supplement the human tutoring.

Location: We currently have commitments from ten university partners, including CUNY Reading Fellows and cohorts of universities from Minnesota and Iowa. We are still actively recruiting universities from the 34 University Tutoring Collaborative participants.

Data:

- student demographic data
- student reading proficiency benchmark data on school administered assessments
- student AI usage data
- pre/post assessment of intervention
- university tutor demographic data
- university tutor knowledge and linguistic survey data

Implications/Findings: TBD

An AI-driven and Closed-loop Human Tutor Training and Real-life Assessment System

Danielle R. Thomas, Carnegie Mellon University; Marie Cynthia Abijuru Kamikazi, Carnegie Mellon University; Clara Brandt, Carnegie Mellon University; Conrad Borchers, Carnegie Mellon University; Kenneth R. Koedinger, Carnegie Mellon University

Status: Complete

Motivation: There exist numerous tutor training platforms. However, few provide human tutors AI-driven training and evaluation based on real-life performance. We present an AI-driven system that provides tutors feedback during training and evaluates authentic tutoring.

Research Questions:

RQ1 (learning): Do tutors demonstrate significant learning gains on specific pedagogical skills (“tutor moves”) within AI-enhanced scenario-based lessons, as measured by pre-to-post test performance improvements?

RQ2 (application): To what extent does performance in training predict the successful application and quality of those skills in real-life tutoring sessions, and which training formats, open responses versus multiple-choice questions, serve as the strongest predictors of behavioral transfer?

Methods: Quantitative,Quasi-Experimental; Among 86 human tutors instructing students remotely in math, the tutors completed six scenario-based lessons. Aggregate analysis revealed a statistically significant 7.4% learning gain. Using mixed-effects models across 405 session-to-lesson pairs, we found that training performance significantly predicted real-life transcript scores with an effect size of 0.25 SD. Model comparison (AIC/BIC) indicated that holistic measures of training provided the best fit, with open-response performance serving as a stronger predictor of transfer than multiple-choice questions. Exploratory analysis showed that after training, tutors were significantly more likely to encounter pedagogical opportunities (61.1% to 68.9%) and demonstrated higher execution quality within those opportunities (65.5% to 68.1%). Interrupted time series analysis suggested that these tutor improvements were part of a gradual trend over time rather than an immediate intervention effect of training. We illustrate a path for “closing the loop” by moving toward continuous assessment of tutor performance, feedback, and improved student learning. In doing so, we contribute open datasets, AI prompts, and scoring rubrics to support transparency and reproducibility.

Location: Adult tutors come from 3 colleges in PA and provide remote tutoring to students across 5 States.

Data: human tutor online lesson data; human tutor and student transcription data of tutoring sessions

Implications/Findings: Aggregate analysis revealed a statistically significant 7.4% learning gain. Using mixed-effects models across 405 session-to-lesson pairs, we found that training performance significantly predicted real-life transcript scores with an effect size of 0.25 SD. Model comparison (AIC/BIC) indicated that holistic measures of training provided the best fit, with open-response performance serving as a stronger predictor of transfer than multiple-choice questions. Exploratory analysis showed that after training, tutors were significantly more likely to encounter pedagogical opportunities (61.1% to 68.9%) and demonstrated higher execution quality within those opportunities (65.5% to 68.1%). Interrupted time series analysis suggested that these tutor improvements were part of a gradual trend over time rather than an immediate intervention effect of training.

Evaluation of Chapter One Short-Burst Early Literacy Tutoring

Greg Chojnacki, Mathematica

Status: Just beginning

Motivation: This study is designed to yield evidence on whether different dosage models – 2.5 times per week on average versus 5 times per week on average – yield different amounts of additional learning. The study will also provide rigorous evidence on a version of the tutoring model that has explicit cost and staffing requirements, allowing the model to be disseminated more readily if the study finds evidence of positive impacts.

Research Questions:

1. To what extent does being assigned to receive Chapter One tutoring in either dosage model (2.5 or 5 sessions per week) improve students' elementary literacy outcomes, relative to not receiving Chapter One tutoring?
2. To what extent does Chapter One's standard, 2.5 sessions per week dosage model improve students' elementary literacy outcomes, relative to not receiving Chapter One tutoring?
3. What is the difference in impact on student learning outcomes between the 2.5 sessions per week and the 5 sessions per week dosage model?
4. To what extent does each dosage model—2.5 sessions per week and 5 sessions per week—improve elementary literacy outcomes for students who are of different genders; economically disadvantaged; multilingual learners; in different grades; and those who start the school year with different levels of literacy skills?
5. To what extent do students in each intervention group receive more minutes of personalized instruction than the comparison group after accounting for other personalized instruction provided outside of the intervention tutoring?

Methods: Quantitative. The study team will randomly assign students whose parents consented to their participation into each of the three study groups (Intervention 1, or I1; Intervention 2, or I2; and Control group, C) within classrooms. Mathematica will use regression analysis to estimate the impact of each intervention arm, compared to business as usual.

Location: Florida

Data: The study will gather data from three sources: a teacher survey, school administrative data, and Chapter One implementation data. The teacher survey data will provide student-level information on the types of personalized instruction each student received and the estimated amount they received. The administrative data will provide all student demographic, grade, classroom, and achievement data. The Chapter One implementation data will provide student level variables on minutes and sessions of tutoring received and aggregate data on the cost of tutoring per student in each experimental group.

Implications/Findings: NA - random assignment and launch of tutoring have not yet started.

Link to preregistration: NA

EIR Regional Tutoring Initiative

Greg Chojnacki, Mathematica; Lindsay Fox, Mathematica; Phil Gleason, Mathematica

Status: Just beginning

Motivation: The Accelerate Regional Tutoring Initiative was designed to develop experimental evidence on how the effects of tutoring vary with frequency of tutoring and student-tutor ratio, two key drivers of program cost. It also seeks to document factors that support and inhibit successful tutoring implementation at the scale of multiple districts and over 1,000 students per year over four years. By conducting two RCTs, one to test Air Reading virtual tutoring and one to test Enhanced Core Reading Instruction in-person tutoring, the study will

specifically develop new evidence on how tutoring effectiveness varies with the amount of tutoring students receive and the group size in which students receive tutoring.

Research Questions:

1. To what extent do each tutoring program—Air Reading and ECRI—impact student literacy achievement?
2. To what extent does the effectiveness of Air Reading differ when students receive tutoring less frequently (3 vs. 4 tutoring sessions per week)? To what extent does the effectiveness of ECRI differ when tutoring is provided in larger vs. smaller groups of students (2 vs. 4 students)?
3. To what extent do the impacts of the tutoring models vary by student or school characteristics?
4. What are the total and per-student costs of implementing each tutoring model?
5. What is the cost-effectiveness of each tutoring model?

Methods: Mixed Methods. The study will conduct two three-arm RCTs involving over 4,000 students, where students in the four intervention groups are differentiated by either the frequency or group size of the tutoring, as outlined below. Air Reading (virtual tutoring) RCT: Comparison Group 1: Business as usual instruction Intervention Group 1A: Students receive an average of 4 tutoring sessions per week. Intervention Group 1B: Students receive an average of 3 tutoring sessions per week. ECRI (in-person tutoring) RCT: Comparison Group 2: Business as usual instruction Intervention Group 2A: Students receive tutoring in groups of 2. Intervention Group 2B: Students receive tutoring in groups of 4. The study will use linear regression analyses to estimate the impacts of each type of tutoring relative to a comparison group. It will also gather qualitative data on district staff perspectives on supporting strong implementation and cost data to assess the cost and cost effectiveness of each tutoring model.

Location: Oklahoma

Data: The study will use student administrative data to estimate the impact of each tutoring program (and implementation variant) on student learning outcomes. We will gather teacher survey data to understand the instruction and learning supports delivered to comparison group and treatment group students. We will use focus groups to qualitative data on district staff perspectives on supporting strong implementation and cost data to assess the cost and cost effectiveness of each tutoring model.

Implications/Findings: The study is in the planning stage. The project will provide direct evidence on how tutor-to-student ratio and tutoring frequency - which are all significant cost drivers of program implementation - affect the impacts of tutoring programs. It will also develop evidence on the factors that support and inhibit the successful implementation of tutoring at a scale of multiple districts.

Link to preregistration: NA

NOVEL METHODS

Moving past human labels: Best practices for annotating conversational data

Julian Bernado, Stanford University; Ana Ribeiro, Stanford University; Xander Beberman, Stanford University; Susanna Loeb, Stanford University

Status: Presented at AEFP, currently revising and aiming to submit to PNAS

Motivation: Wide-scale discursive utterance annotation was revolutionized by the first round of large neural networks trained on large amounts of language data (e.g. BERT). We are currently in the midst of another tipping point in deciding best practices for utterance annotation on account of the rise of LLMs. This paper aims to provide guidance for researchers studying educational conversational data on which of these two methodologies best suits their task, data, and which models they should choose.

Research Questions: Can prompted LLMs outperform fine-tuned BERT-style models in utterance identification? What task characteristics inform the performance gap between the two classes of methods?

Methods: NLP, LLMs, ML, Annotation

Location: National

Data: Transcriptions or chat logs from 1:1 tutoring sessions

Findings: The best LLMs outperform BERT-style models in all cases and the gap may be informed the number of classes being annotated and distribution of examples within those classes

Understanding Learning Breakdowns and Successes in AI Tutoring

Kirk Vanacore, Cornell University; Ana Ribeiro, Stanford University; Julian Bernado, Stanford University; Joshua Marland; Cornell University; Susanna Loeb, Stanford University

Status: In progress

Motivation: Rapid progress in AI has led many to wonder whether AI-based systems can serve as tutors. However, it's unclear what makes these tutors effective.

Research Questions: How can we identify learning breakdowns at scale? Which tutor moves promote student learning and when?

Methods: NLP, LLMs, Knowledge Tracing

Locations: US/UK

Data: Session-level event logs and transcribed student utterances, all from AI Tutoring

Findings: TBD

Using NLP methods to measure student engagement in online tutoring sessions - An exploratory analysis

Julian Bernado, Stanford University; Ana Trindade Ribeiro, Stanford University; Carly Robinson, Stanford University; Hsiaolin Hsieh, Stanford University; Xander Beberman, Stanford University; Susanna Loeb, Stanford University

Status: Just beginning

Motivation: This study is part of a series of projects using NLP methods to explore patterns of behavior expressed through language during tutoring sessions by tutors and students, such as measures of student engagement and tutor strategies for relationship-building, and how they relate to student outcomes.

Research Questions: What are common behaviors displayed by students through speech during a math tutoring lesson? Which behaviors can be considered signs of (dis)engagement to inform tools that can assist tutors to increase student engagement?

Methods: Large Language Models; Descriptive

Location: National

Data: Transcribed text from tutoring sessions' audio recordings; student and tutor characteristics collected by the tutoring provider; internal measures of student learning progress.

Implications/Findings: This project will continue to develop measures of engagement, relationships, and learning in virtual tutoring sessions.

National Tutoring Observatory

Rene Kizilcec, Cornell University; Rachel Slama, RAND; Ken Koedinger, Carnegie Mellon University; Danielle Thomas, Carnegie Mellon University; Doug Pietrzak, Fresh Cognate

Status: Just beginning

Motivation: The mission of the National Tutoring Observatory is to observe and record great teachers at work in one-on-one and small group interactions with learners. By partnering with a range of tutoring providers, we will create the world's largest repository of video and transcript data about tutoring interactions and the incredibly important work of teachers. Our aspiration is to create a Million Tutor Moves dataset that records at least one million interactions between teachers and students across a range of subjects, grade levels, and educational contexts. These new data will bring new insight into the craft of teaching, advance the science of instruction, and provide important data for technologists developing AI tools

Research Questions: What do tutors say and do when working with learners?

Methods: Quantitative

Our enterprise is more infrastructure than study. Our primary goal is to collect tutoring data at large scales, and make a variety of cleaned, organized data products (tutoring videos, multimedia session transcripts, associated context and outcome data) made available as widely as possible to educational researchers and technologists.

Location: Across the US, UK, and hopefully many more places in the future.

Data: We have seven tutoring providers, and we're working with them to provide access to video recordings and transcripts of their tutoring sessions at scale.

Implications/Findings: We've just begun! Teachers are essential to learning, especially the improvisational interactions between teachers and students. We hope we'll have more things to share within the year.

Identifying Effective Tutoring Strategies Through NLP Analysis of High and Low Value-Added Tutors

Ana Trindade Ribeiro, Stanford University; Xander Beberman, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Status: In progress

Motivation: Tutoring, both in-person and virtual, has been shown to be an effective intervention for improving student learning (Nickow et al., 2024; Gortazar et al., 2024). While changes in program design are known to impact effectiveness, relatively little is known about which specific tutor practices are most effective at improving learning. This is partially due to the historical cost and difficulty associated with classroom observation. However, the recent increase in virtual instruction offers an opportunity to measure strategies much more efficiently. By leveraging modern natural language processing techniques, large datasets of tutoring recordings can be analyzed quickly and reliably, allowing for highly scalable measurement of instructional approaches and educator-student interactions. This study uses a data-driven approach, relying on the methodology for estimating teacher value-added, to rank a group of tutors by their effectiveness and identify differences in pedagogical strategies between the most and least effective tutors.

Research Questions: What are the pedagogical strategies most commonly used by the most effective tutors? What are the pedagogical strategies most commonly used by the least effective tutors? Which tutor pedagogical strategies are predictive of learning growth for students?

Methods: Quantitative

Data: Transcribed text from tutoring sessions' audio recordings; student administrative data including beginning, middle, and end-of-year DIBELS test scores.

Implications/Findings: The study is in progress and will continue to develop insights into the differences in linguistic and pedagogical strategies used by groups of higher and lower value-added tutors.

Sandpiper: Orchestrated AI-Annotation for Educational Discourse at Scale

Daryl Hedley, FreshCognate; Doug Pietrzak, FreshCognate; Jorge Dias, FreshCognate; Ian Burden, FreshCognate; Bakhtawar Ahtisham, Cornell University; Zhuqian Zhou, Cornell University; Kirk Vanacore, Cornell University; Josh Marland, Cornell University; Rachel Slama, Cornell University; René Kizilcec, Cornell University; Justin Reich, MIT; Kenneth Koedinger; CMU

Status: Analyses finalized and being written up

Motivation: Develop a system that could easily process and annotate educational dialogue data and meet the design goals listed below. We will evaluate this system using user testing and quantitative comparisons to other methods.

Research Questions: Guarantee the privacy of educational data through scalable methods; Prevent LLM hallucinations and malformed tags during the coding process; Support iterative refinement and validation of automated annotation

Methods: Descriptive; Digital educational environments are expanding toward complex AI and human discourse, providing researchers with an abundance of data that offers deep insights into learning and instructional processes. However, traditional qualitative analysis remains a labor-intensive bottleneck, severely limiting the scale at which this research can be conducted. We present Sandpiper, a mixed-initiative system designed to serve as a bridge between high-volume conversational data and human qualitative expertise. By tightly coupling interactive researcher dashboards with agentic Large Language Model (LLM) engines, the platform enables scalable analysis without sacrificing methodological rigor. Sandpiper addresses critical barriers to AI adoption in education by implementing context-aware, automated de-identification workflows supported by secure, university-housed infrastructure to ensure data privacy. Furthermore, the system employs schema-constrained orchestration to eliminate LLM hallucinations and enforces strict adherence to qualitative codebooks. An integrated evaluations engine allows for the continuous benchmarking of AI performance against human labels, fostering an iterative approach to model refinement and validation. We propose a user study to evaluate the system's efficacy in improving research efficiency, inter-rater reliability, and researcher trust in AI-assisted qualitative workflows.

Location: NA

Data: NA

Implications/Findings: Thus far we have built the system, but we have yet to fully implement a user study, The system will allow for tutoring providers and researchers to annotate and analyze tutoring data.

Link to Report/Brief: https://arxiv.org/pdf/2603.08406

Multimodal transcription of virtual tutoring sessions

Shivang Gupta, CMU; Ken Koedinger, CMU; Danielle Thomas, CMU; Emma Brunskill, Stanford; Zachary Levonian, Digital Harbor Foundation; Hui Cheng, CMU

Status: In process

Motivation: To create a reproducible process for multimodal transcription of tutor and student interactions in a virtual environment.

Research Questions: Can LLMs generate multimodal transcripts of tutoring sessions effectively? Which LLM models are most suited for this task? What are the optimized parameters for multimodal transcript generation? How do the generated multimodal transcripts differ in usefulness from audio-only transcripts?

Methods: Mixed Methods,Experimental; Experimental design to generate transcripts with different models and parameters, followed by expert comparison and coding to compare with existing transcripts to measure accuracy and usefulness in the context of tutor performance improvement.

Location: Pittsburgh

Data: Tutor-student interaction zoom recordings from PLUS Tutoring sessions

Implications/Findings: Gemini is the only model with a large enough context window to effectively transcribe full sessions directly from video. This technique could be used for automated tutor supervision, compliance monitoring, performance analyses, coaching, and process improvement.

Tutor Move Taxonomy: A Theory-Aligned Framework for Analyzing Instructional Moves in Tutoring

Zhuqian Zhou, Cornell University; Kirk Vanacore, Cornell University; Tamisha Thompson, Cornell University; Jennifer St John, Cornell University; Rene Kizilcec, Cornell University

Status: Analyses finalized and being written up

Motivation: NA

Research Questions: NA

Methods: The study develops a tutor-move taxonomy using a hybrid deductive–inductive design: it first synthesizes prior literature from tutoring, learning sciences, classroom discourse, and intelligent tutoring systems to build a preliminary framework, then refines that framework through iterative coding of authentic tutoring transcripts. Two expert annotators coded sessions from multiple tutoring providers across collaborative and independent rounds, using discussion, discrepancy review, and inter-rater reliability checks to revise definitions and category boundaries until the codebook stabilized. The paper also situates this taxonomy within a broader analytic approach that links coded tutor moves to student outcomes using quasi-experimental machine-learning methods such as leave-one-out potential outcomes and causal forests.

Location: The study uses tutoring transcripts from three different tutoring providers in different contexts: one-on-one chat, chats within an intelligent tutoring platform, and small group tutoring.

Data: The study uses various data sources from tutoring platforms in the US and UK.

Implications/Findings: Through the deductive/inductive coding processes, we have identified 29 tutor moves. Understanding what makes tutoring effective requires methods for systematically analyzing tutors' instructional actions during learning interactions. This paper presents a tutor move taxonomy designed to support large-scale analysis of tutoring dialogue within the National Tutoring Observatory. The taxonomy provides a structured annotation framework for labeling tutors' instructional moves during one-on-one tutoring sessions. We developed the taxonomy through a hybrid deductive-inductive process. First, we synthesized research from cognitive science, the learning sciences, classroom discourse analysis, and intelligent tutoring systems to construct a preliminary framework of tutoring moves. We then refined the taxonomy through iterative coding of authentic tutoring transcripts conducted by expert annotators with extensive instructional and qualitative research experience. The resulting taxonomy organizes tutoring behaviors into four categories: tutoring support, learning support, social-emotional and motivational support, and logistical support. Learning support moves are further organized along a spectrum of student engagement, distinguishing between moves that elicit student reasoning and those that provide direct explanation or answers. By defining tutoring dialogue in terms of discrete instructional actions, the taxonomy enables scalable annotation using AI, computational modeling of tutoring strategies, and empirical analysis of how tutoring behaviors relate to learning outcomes.

Link to Report/Brief: https://arxiv.org/abs/2603.05778

Analyses of multimodal transcriptions of tutoring sessions for tutor coaching.

Ken Koedinger, CMU; Danielle Thomas, CMU; Emma Brunskill, Stanford; Zachary Levonian, Digital Harbor; AJ Scott; Digital Harbor

Status: In process

Motivation: To see if multimodal transcripts can be used to effectively coach tutors and improve their performance in actual sessions.

Research Questions: Does LLM-generated feedback improve tutor performance in virtual tutoring sessions? What is tutors' perception of the utility of LLM-generated feedback? What is tutors' perception of the accuracy of LLM-generated feedback?

Methods: Mixed Methods,Experimental,Quasi-Experimental, Randomized control trial/interrupted time series with pre-post evaluation of tutor behaviors.

Location: PA

Data: Feedback and usage metrics collected from tutors who are give llm-generated feedback.

Implications/Findings: TBD. Will demonstrate the potential of LLMs for automated tutor coaching based on different learning science constructs.

PUBLICATIONS

The following publications have been released since the launch of NSSA in 2021. We included any studies that were submitted for the NSSA Conference In-Progress Research document by the authors and have since been published.

IMPLEMENTATION, PROCESS, AND SYNTHESIS

The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence

Andrew Nickow, Northwestern University; Philip Oreopoulos, University of Toronto; Vincent Quan, JPAL

Abstract: Tutoring ranks among the most versatile and potentially transformative educational tools available. Dozens of randomized experiments have evaluated preK–12 tutoring programs, varying widely in approaches, contexts, and costs. This article presents results from a systematic review and meta-analysis of tutoring field experiments. We develop a framework for understanding variation in tutoring program impact and examine effect sizes (ESs) across a range of characteristics. We find that tutoring programs yield consistently

substantial positive impacts on learning, with an overall pooled ES of 0.288 SD (SE = 0.029, $p < .001$). ESs tend to be largest for programs that use teachers or paraprofessionals as tutors, are held in earlier grades, occur at least 3 days per week, and are held during school.

Outlet: American Educational Research Journal

Link: https://doi.org/10.3102/00028312231208687

High-Impact Tutoring: State of the Research and Priorities for Future Learning

Carly Robinson, Stanford University; Susanna Loeb, Stanford University

Abstract: Research consistently demonstrates that tutoring interventions have substantial positive effects on student learning. As a result, tutoring has emerged as a promising strategy for addressing COVID-related learning loss and affording greater educational opportunities for students living in poverty. The effectiveness of tutoring programs, however, varies greatly, and these variations may drive differential gains in student learning. Therefore, determining the program characteristics that do and do not drive positive student outcomes will be key to providing guidance for policymakers and practitioners who want to implement high-impact tutoring at scale. Our goal is to highlight the programs, characteristics, and conditions that evidence suggests make for effective tutoring and to create an evidence-based framework for delivering and evaluating tutoring interventions. In addition, we identify promising questions for future research.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai21-384

Design Principles for Accelerating Student Learning with High-Impact Tutoring

Carly Robinson, Stanford University; Matthew Kraft, Brown University; Susanna Loeb, Stanford University; Beth Schueler, University of Virginia

Abstract: Research shows that tutoring, especially when delivered intensively and with personalized support, is one of the most effective interventions for improving math and reading outcomes, particularly for students from low-income backgrounds. Successful tutoring programs are characterized by strong tutor-student relationships, frequent sessions, small group sizes (no more than three students per tutor), high-quality materials aligned with the curriculum, and data-informed practices. Programs are more effective when scheduled during the school day and when tutors receive proper training and ongoing support. Finally, the brief provides guidance on implementing and scaling tutoring initiatives to ensure sustainable, impactful outcomes.

Outlet: EdResearch for Action

Link: https://edresearchforaction.org/research-briefs/design-principles-for-accelerating-student-learning-with-high-impact-tutoring/

Generalizing about Effect Sizes when Education Interventions are Taken to Scale: A Meta-Analytic Exploration of the Experimental Evidence on K-12 Tutoring

Matthew Kraft, Brown University; Beth Schueler, University of Virginia; Grace Falken, Brown University

Abstract: U.S. public schools are engaged in an unprecedented effort to expand tutoring in the wake of the pandemic. Broad-based support for scaling tutoring emerged, in part, because of the large effects on student achievement found in prior meta-analyses. We conduct an expanded meta-analysis of 282 randomized control trials and explore how estimates change when we better align our sample with a policy-relevant target of inference: large-scale tutoring programs in the U.S. aiming to improve standardized test performance. Pooled effect sizes from studies with stronger target-equivalence remain meaningful, but are only a third to a half as large as those from our full sample. This result is driven by stark declines in pooled effect sizes as tutoring program scale increases. We explore four hypotheses for this pattern and identify a bundled package of design features that our analyses suggest may help to partially inoculate programs from these attenuated effects at scale.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-1031

Challenges and Strategies to Scale High-Impact Tutoring: Learning from School Districts' Implementation Efforts

Sara White, Vanderbilt University; Leiah Groom-Thomas, Stanford; Amanda Lu, Georgetown University; Susanna Loeb, Stanford

Abstract: High-impact tutoring (HIT) is academic instruction provided in a one-to-one or small group setting by a trained tutor that takes place at least three or more times per week. HIT instruction focuses on cultivating positive student–tutor relationships and supporting students' knowledge and skill development. HIT has consistently shown large, positive effects on student learning in both math and reading. However, HIT is not easy to implement and program design is often sacrificed. This study aims to understand strategies that school districts have used to maintain fidelity to HIT design while expanding the number of students served. Specifically, districts have often faced challenges hiring enough tutors, providing quality tutor training and curricular resources, ensuring student attendance in tutoring sessions, and utilizing data. In response to staffing challenges, districts have developed multiple options for tutor hiring including recruiting community members, university students, existing school staff, and virtual providers. To support both instructional quality and tutor retention, districts have provided tutor training aligned with tutors' existing skills and needs, which differs by tutor type. Districts have also improved student attendance by integrating HIT within existing school schedules and clarifying the relationship between HIT and other student support programs.

Outlet: Education Finance & Policy

Link: https://doi.org/10.1162/EDFP.a.434

Business Not as Usual: Understanding Factors for Organizational Change after a Crisis

Amanda Lu, Stanford University; Nancy Waymack, Stanford University; Susanna Loeb, Stanford University

Abstract: Drawing on the Crisis Management Cycle (CMC) framework, this study examines school systems' organizational adjustments in response to the COVID-19 pandemic, focusing on the implementation of High Impact Tutoring (HIT) to address the pandemic's academic impacts. Analyzing 112 interviews across ten local education agencies, we identify three post-crisis organizational change pathways: stagnation without learning, change through learning, and stagnation after initial learning. Critical to navigating these pathways are stakeholder alignment, external partnerships, access to expertise, effective resource allocation, and organizational readiness for adaptation. Our research highlights how these factors collectively determine an educational institution's resilience and capacity for long-term structural adjustment post-crisis. By elucidating the mechanisms that enable or impede organizational learning and change, this paper contributes insights into overcoming entrenched practices, thereby enhancing schools' preparedness and response capabilities for future crises.

Outlet: National Student Support Accelerator, Forthcoming at Teacher's College Record

Link:
https://nssa.stanford.edu/studies/business-not-usual-understanding-factors-organizational-change-after-crisis

Exploring the Potential of Outcomes-Based Contracting: Findings from Initial Implementations

Amanda Lu, Georgetown University; Leiah Groom-Thomas, Stanford University; Susanna Loeb, Stanford University

Abstract: Outcomes-Based Contracting (OBC) ties vendor payments to performance metrics, aiming to enhance accountability in public education. This study examines its implementation in tutoring services through the Southern Education Foundation pilot program. Interviews with district leaders and vendors reveal that OBC fosters collaboration, improves service alignment with student needs, and enhances data tracking. However, financial risks for vendors and the complexity of implementation pose challenges. While OBC shows promise in strengthening district-vendor relationships, its broader adoption requires capacity building, equitable risk-sharing, and further research on long-term student outcomes.

Outlet: Education Finance & Policy

Link: https://doi.org/10.1162/EDFP.a.28

The Key Resource of Time: Master Schedules and Effective Allocation of Students and Educators

Amanda Lu, Stanford University; Paymon Rouhanifard, Timely Schools; Christopher Cleveland, Brown University; Ev Gilbert, Stanford University; Susanna Loeb, Stanford University

Abstract: A central challenge facing education leaders is allocating limited resources in pursuit of their priorities. Three of their critical resources are time, money, and people. A school's master schedule reflects the allocation of all three of these critical resources and ultimately determines the educational opportunities available to students. A school's schedule dictates who will be teaching them, what they will be learning, where this learning will take place, and how much instruction they will receive. The COVID-19 pandemic reduced learning and led many students to disengage from school, both of which called for greater individualized attention to accelerate student learning and rebuild their overall well-being. Yet, these kinds of supports require complex scheduling. New scheduling tools—such as those utilizing AI optimization technology—allow educators to optimize these complex demands, though even with these tools, educators have to navigate external pressures and maintain focus on their strategic priorities. In this brief, we identify the demands on school leaders when creating a master schedule, note the trade-offs present in trying to meet the needs of diverse student populations, and identify key capacities and supports for schedulers. We conclude by discussing the skills and tools that can help education leaders better navigate demands while allocating resources effectively to meet their strategic priorities.

Outlet: SCALE Initiative

Link: https://scale.stanford.edu/publications/master-scheduling

The Promise of Using HIT to Improve Early Literacy Outcomes

Sarah Novicoff, Stanford University; Susanna Loeb, Stanford University

Abstract: Much tutoring focuses on young students with a particular emphasis on the development of their literacy skills. Learning to read strongly predicts later life outcomes, including high school test scores (Sparks et al., 2014) and high school graduation (Hernandez, 2011), and is often thought of as a key outcome of elementary education. As two-thirds of fourth graders are not meeting proficiency on reading assessments (NAEP Reading, 2022), schools have turned to early literacy tutoring to catch students up. This brief describes early literacy tutoring in the US, illustrating the range of approaches available to meet the opportunities and needs across contexts while providing the individualized support that accelerate young students' literacy skills.

Outlet: Kappan

Link: https://kappanonline.org/lessons-from-the-early-literacy-tutoring-landscape/

Wittenberg Community High-Dosage Tutoring Project

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

Abstract: A total of 129 K-5th grade students received tutoring from 79 tutors in ELA and math across the 2023-24 school year. The number of elementary students receiving math instruction increased from Fall to Spring, as did the number of sessions each student received on average. Elementary students consistently reported positive attitudes towards tutoring experiences across the 2023-24 school year. Tutors (preservice teachers) reported high levels of confidence in their instructional self-efficacy in selecting and implementing ELA, Phonics, and Math strategies during tutoring sessions, though they had mixed perceptions of their

tutoring experience. Wittenberg's tutoring program showcases how institutions of higher education can integrate a tutoring program into reading and math methods coursework for preservice teachers to promote hands-on field experience for future educators. It demonstrates how universities can have local impact by building rapport with surrounding school districts and fostering meaningful relationships with students in need.

Outlet: National Student Support Accelerator

Link: https://nssa.stanford.edu/briefs/2023-2024-wittenberg-university-high-impact-tutoring-program-implementation-report

Scaling High-impact tutoring: School Level Perspectives on Implementation Challenges and Strategies

Patricia Burch, University of California; Susanna Loeb, Stanford University

Abstract: High-impact tutoring has emerged as a primary school district investment for addressing learning loss that occurred during the COVID-19 pandemic. While existing research shows that high-impact tutoring is effective for accelerating student learning, this study examined the school-level facilitators and barriers to scaling high-impact tutoring. Situated in an urban traditional school district and an urban charter management organization, we collected survey and interview data from teachers and administrators to identify scaling challenges. Major barriers to scaling included time and space constraints, tutor supply and quality, updated data systems, and school level costs, while a key facilitator was teacher buy-in. We end the paper with recommendations for how districts can strategically grow their high-impact tutoring efforts.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-923

Implementation of the OSSE High Impact Tutoring Initiative: First Year Report School Year 2022 – 2023

Cynthia Pollard, UnboundED; Amanda Lu, Stanford University; Amirpasha Zandieh, University of Colorado Boulder; Carly D. Robinson, Stanford University; Susanna Loeb, Stanford University; Nancy Waymack Stanford University

Abstract: The District of Columbia (DC) Office of the State Superintendent of Education (OSSE) launched a three-year, $33 million investment in scaling and supporting high-impact tutoring (HIT) across DC, with a particular focus on students classified as "at-risk" or having experienced disrupted instruction during the COVID-19 pandemic. This investment is a core part of the city's strategy to address interrupted schooling, given the strong body of evidence demonstrating the effectiveness of HIT as an intervention that can produce dramatic gains in student learning. Through this investment, OSSE strives to accelerate learning for students who were impacted by disrupted learning as well as to address persistent achievement gaps present before the pandemic. This report focuses on implementation and effects of the grant program from July 2022 – June 2023. Using available data collected by tutoring providers and administrative data shared by OSSE, we describe program reach, impact on student outcomes including attendance and achievement indicators, and facilitators and barriers to implementation.

Outlet: National Student Support Accelerator

Link: https://nssa.stanford.edu/briefs/implementation-osse-high-impact-tutoring-initiative

Implementation of the OSSE High Impact Tutoring Initiative - School Year 2023 – 2024 Second Year Report

Amanda Lu, Stanford University; Nancy Waymack, Stanford University; Demetra Kalogrides, Stanford University; Carly D. Robinson, Stanford University; Monica G. Lee, Stanford University; Susanna Loeb, Stanford University

Abstract: The second full school year (2023-24) of the OSSE High Impact Tutoring Initiative expanded the reach of an already ambitious program. The Initiative served 7,274 students, approximately 8% of students in DC schools and 12% of students classified at-risk. The Initiative was able to increase participation by 2,000 students from its first year of implementation while also increasing the successful targeting of at-risk students who stand to benefit most from the program. The Initiative also increased the average dosage level to 33.86 sessions. Collectively, this is a significant improvement in program scale and program delivery, ensuring that increases in tutoring continue to serve students who are most in need of potential benefits.

We continue to find evidence of a positive impact of OSSE-funded HIT on school attendance. This is more evidence that, while alone high-impact tutoring is not a silver bullet for attendance issues, it can be part of the solution. The Initiative has continued to be successful in serving students with lower academic performance and students from historically marginalized groups, who represent the majority of OSSE-funded HIT students. Students with lower state standardized test scores (as measured by Partnership for Assessment of Readiness for College and Careers or PARCC) in the 2022-23 school year, as well as Black or African American students and Hispanic or Latino students were more likely to receive tutoring in the Initiative's second year.

Outlet: National Student Support Accelerator

Link: https://nssa.stanford.edu/briefs/implementation-osse-high-impact-tutoring-initiative-year-2023-2024

Road to Recovery: Impacts of Academic Recovery Interventions on Student Achievement in 2022-23.

Maria V. Carbonari, Michael DeArmond, Daniel Dewey, Elise Dizon-Ross, Dan Goldhaber, Thomas J. Kane, Anna McDonald, Andrew McEachin, Emily Morton, Atsuko Muroga, Alejandra Salazar, Douglas O. Staiger

Abstract: The COVID-19 pandemic devastated student achievement, with declines rivaling those after Hurricane Katrina. These losses widened achievement gaps between historically marginalized students and their peers. Three years later, achievement remains behind pre-pandemic levels for many students. This paper examines 2022-23 academic recovery efforts across eight districts, including tutoring, small group instruction, after-school, extended year, double-dose, digital learning, and expert teacher interventions. Across 22 math and reading interventions, most were delivered to fewer students and for less time than planned. We find positive effects for one tutoring program on math scores and two tutoring programs on reading scores, ranging from 0.22 to 0.33 SD. Each of these programs served a very small share of the district's students and was unlikely to play a major role in district-wide academic recovery. Finally, we find that having an "expert" teacher with high evaluation scores as opposed to a non-expert teacher significantly improves student achievement by https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/road_to_recovery_report_june_2024_v6_01.pdf0 .06 SD in math and 0.11 SD in reading. While highlighting the promise of intensive academic interventions, our

findings underscore the challenges districts face in scaling such interventions to match their recovery needs. The field needs better evidence regarding successful implementation of large-scale interventions.

Outlet: CEPR

Link: https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/road_to_recovery_report_june_2024_v6_01.pdf

Embedded Tutoring in California Community Colleges: Perspectives from the Field on a Promising Practice

Mark Duffy, Research for Action; Kri Burkander, Research for Action

Abstract: Drawing on qualitative data collected in a sample of colleges as part of a larger study on the implementation and impact of Assembly Bill 705 in California, this paper explores the rollout of corequisite reforms, focusing on the use of embedded tutors in introductory math and English courses as a strategy to meet to the needs of students. This paper highlights promising practices identified through extant research and fieldwork at study institutions, provides additional evidence on the value of the reform, discusses challenges, and makes recommendations for the field.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-984

Interpreting Null Effects in Education: Insights from a Tutoring Experiment

Elizabeth Huffaker, University of Florida; Carly D. Robinson, Stanford University; Emanuele Bardelli, Santa Rosa School District; Sara White, Vanderbilt University; Susanna Loeb, Stanford University

Abstract: As school districts focus on improving learning, they can learn not only from when and where interventions work—but also from why they sometimes do not. Policymakers widely embraced high-impact tutoring as an evidence-supported strategy to address learning delays from the COVID-19 pandemic. However, scaling these promising practices can be difficult, and not all implementations will be effective. Many districts have turned to third-party virtual tutoring providers to deliver student supports during the school day. Using random assignment, we evaluate the impacts of one such program for 3rd through 8th grade students in a suburban Texas school district. Compared with students assigned to the comparison interventions, we find no effect of assignment to virtual tutoring on math achievement, and, for reading, we find a moderate negative effect on the state end-of-year assessment (i.e., -0.09 SD) and no effect on a low-stakes exam. Drawing from frameworks for interpreting null or unexpected results in education experiments, we find further evidence of subject-specific heterogeneity in the implementation and efficacy and identify coverage of standards-aligned material as a moderator of estimated effectiveness relative to “business-as-usual” interventions. This paper offers strategies to identify factors contributing to null or unexpected results and highlights implications for designing policy-relevant studies to assess educational interventions.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1295

Virtual 1:1 Literacy Tutoring in Oakland Unified School District: Implementation and Effectiveness of a Pilot at Scale

Mary Laski, Center on Reinventing Public Education; Calen Clifton, Center on Reinventing Public Education

Abstract: In 2024–25, Oakland Unified School District (OUSD) launched a districtwide pilot of virtual, high-dosage, 1:1 literacy tutoring in partnership with three providers: Hoot, Ignite Reading, and OpenLiteracy. This initiative aimed to address early reading gaps, particularly in phonics, for K–2 students who were below grade level. Key findings: High implementation fidelity: Over 80% of tutored students received at least 20 sessions, and nearly 60% received 900+ minutes of tutoring. Targeted and equitable reach: Schools followed district guidelines in selecting students most in need of support. Meaningful academic impact: On the i-Ready Phonics assessment, tutored students gained the equivalent of 1.3 additional months of learning, and those who received 15+ hours saw gains of 1.6 months. On the DIBELS Oral Reading Fluency-Accuracy assessment, students gained up to 9 months of additional learning, though this result is based on a smaller sample and should be interpreted with caution. Mixed results elsewhere: Tutored students showed similar or slightly lower gains in overall literacy and fluency on other assessments. Provider variation: Students with Ignite Reading outperformed others on the i-Ready Overall assessment, with gains equivalent to 1.5 extra months of learning.Implications: Oakland's pilot stands out as a rare, large-scale, real-world example of effective virtual tutoring. Even at modest doses, students showed measurable gains in foundational skills—suggesting that virtual 1:1 tutoring, when implemented strategically and equitably, can be a powerful tool for early literacy recovery.

Outlet: Center on Reinventing Public Education

Link: https://crpe.org/wp-content/uploads/Literacy-Tutoring-in-Oakland-Unified-School-District_V4.pdf

Insights into the Tutor Workforce

Barbara Condliffe, Rebecca Davis, Jean Grossman, and Shira Mattera; MDRC

Abstract: What is the profile of K-12 tutors? How do tutors feel about their tutoring experiences? What are tutors' future career goals? How might schools and districts facilitate effective tutor-teacher pipelines? This brief addresses these questions using survey data about the experiences of over 400 tutors who worked in 81 schools across seven states that were participating in the Personalized Learning Initiative (PLI) during the 2023-2024 school year. Additionally, the brief reflects on information collected from a subsample of schools about successes and challenges regarding tutoring recruitment and retention, which could promote or undercut tutoring's potential as a source of teachers. It also provides an overview of different strategies policymakers and educational leaders are using to promote tutoring as a pipeline into teaching.

Outlet: MDRC

Link: https://www.mdrc.org/sites/default/files/Insights_Into_Tutor_Workforce.pdf

Tutoring the tutors: Piloting online modules for tutor training

Andrew Kwok, Texas A&M University; Brendan Bartanen, University of Virginia; Michelle Kwok, Texas A&M University; Kathy Ogden Macfarlane, Texas A&M University; Tracey Weinstein, Deans for Impact

Abstract: Background: Despite intense national focus on the positive impact of tutoring, there is little empirical evidence on how best to train tutors. This is particularly pertinent given that tutors could serve as a potential pipeline into teaching. Objective: This mixed-methods study explores the implementation of modules about developing knowledge for high-quality instructional materials for tutor professional development. Research Design: The modules were distributed and completed within four tutoring-based sites in the United States that are part of a national network for aspiring teachers. Quantitatively, we examined pre-/post-assessment data to probe changes in tutor knowledge and career plans. Qualitatively, we interviewed tutors and their program trainers about their module experience. Conclusions: Results indicate growth in tutor knowledge as well as participant recommendations for specific improvements to technology and module engagement to enhance curricular experience. Findings suggest the value and limitations of tutor training within this model of training intervention.

Outlet: Teachers College Record

Link: http://journals.sagepub.com/doi/full/10.1177/01614681251389032

Accelerate's 2024-25 Call to Effective Action: A Synthesis of Lessons Learned

Sophie Bright, Gregory Chojnacki, Kamillah Smith (Mathematica), Jason Godfrey, Accelerate — The National Collaborative for Accelerated Learning

Abstract: This synthesis elevates findings from eight focal grantees that address critical evidence gaps (for example, curriculum alignment and understudied learner profiles) as well as emerging approaches in personalized learning, including artificial intelligence (AI)-enabled tools. Five of the focal grantees with rigorous comparison-group research designs reported positive, statistically significant impacts of the tutoring and personalized learning models on student achievement. These effects were substantial, ranging from about 1.5 to 15 months of additional student learning. A sixth grantee reported marginally statistically significant impacts equivalent to about 1.3 months of additional learning. For focal grantees with statistically significant impacts, tutoring efficiency (or the hours of tutoring associated with one additional month of student learning) ranged from 0.7 to 14 hours. Across the focal grantees, several implementation and measurement lessons emerged, including (1) the value of aligning tutoring content to core instructional materials and pacing, (2) key levers to achieve high dosage such as setting clear expectations and monitoring dosage regularly, (3) the value of drawing on data from tutoring sessions to inform timely instructional adjustments and maintain school buy-in, (4) lessons for scaling at while maintaining tutoring quality, and (5) the value of using standardized measures of the counterfactual condition services, program cost, and dosage.

Outlet: Accelerate: The National Collaborative for Accelerated Learning

Link:
https://accelerate.us/wp-content/uploads/2026/02/Accelerate-CEA_Synthesis-Report_24-25-FINAL-020826-1.pdf

PROGRAM EFFECTS

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

Abstract: This paper presents the results from a randomized controlled trial of an early elementary reading tutoring program that embeds part-time tutors into the classroom to provide short bursts of 1:1 instruction. Eligible kindergarten students were randomly assigned to receive supplementary tutoring during the 2021-22 school year (N=818). The study occurred in a large Southeastern district serving predominantly Black and Hispanic students. Students assigned to the program were over two times more likely to reach the program's target reading level by the end of kindergarten (70% vs. 32%). The results were largely homogenous across student populations and extended to district-administered assessments. These findings provide promising evidence of an affordable and sustainable approach for delivering personalized reading tutoring at scale.

Outlet: Learning and Instruction

Link: https://doi.org/10.1016/j.learninstruc.2024.102021

Apart but Connected: Online Tutoring, Cognitive Outcomes, and Soft Skills

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

Abstract: We study the Tutoring Online Program (TOP), where tutoring is entirely online and tutors are volunteer university students matched with underprivileged middle school students. We leverage random assignment to estimate effects during and after the pandemic (2020 and 2022), investigating channels of impact. Three hours of individual tutoring per week increased math performance by 0.22 SD in 2020 and 0.20 SD in 2022. Higher dosage yielded stronger effects, while group tutoring smaller effects. TOP enhanced students' aspirations, socioemotional skills, and psychological well-being, but only during school closures. We also estimate the impact of TOP on tutors, finding an increase in empathy.

Outlet: American Economic Review

Link: https://www.aeaweb.org/articles?id=10.1257/aer.20240401

Effects of Virtual Tutoring on Young Readers: Results from a Randomized Controlled Trial

Carly Robinson, Stanford University; Cynthia Pollard, Stanford University; Sarah Novicoff, Stanford University; Sara White, Vanderbilt University; Susanna Loeb, Stanford University

Abstract: In-person tutoring has been shown to improve academic achievement. Fewer studies have examined the impact of virtual tutoring and have focused on older students. We present findings from the first randomized controlled trial of virtual tutoring for young children. Students in grades K–2 were assigned to 1:1 tutoring, 2:1 tutoring, or a control group. Virtual tutoring increased early literacy skills by 0.05-SD for all students and 0.08-SD for a sample excluding English learners and students with disabilities (i.e., students not eligible for additional support services). One-on-one tutoring tended to produce larger gains, especially for students initially scoring well below benchmark (0.15-SD). Effects are smaller than typically seen from in-person early literacy tutoring programs but still positive and statistically significant.

Outlet: Education Evaluation and Policy Analysis

Link: https://doi.org/10.3102/01623737241288845

Online tutoring works: Experimental evidence from a program with vulnerable children

Claudia Hupkau, CUNEF Universidad; Antonio Roldan-Mones, ESADE; Lucas Gortazar, ESADE

Abstract: We provide evidence from a randomized controlled trial on the effectiveness of a novel, 100-percent online math tutoring program, targeted at secondary school students from highly disadvantaged neighborhoods. The intensive, eight-week-long program was delivered in groups of two students during after-school hours, mostly by qualified math teachers. The intervention significantly increased standardized test scores (+0.26 SD) and end-of-year math grades (+0.49 SD), while reducing the probability of repeating the school year. The intervention also raised aspirations, as well as self-reported effort at school. The two-on-one design allows us to significantly reduce costs and improve scalability, while showing similar results as one-on-one tutoring programs.

Outlet: Journal of Public Economics

Link: https://doi.org/10.1016/j.jpubeco.2024.105082

Sustained Effects of Small-Group Instruction in Mathematics

Henning Finseraas, NTNU, Ole Henning Nyhus, NTNU Samfunnsforskning, Kari Vea Salvanes, Institutt for Samfunnsforskning, Astrid Marie Jorde Sandsør, University of Oslo

Abstract: Recent research suggests that using additional teachers to provide small-group instruction or tutoring substantially improves student learning. However, treatment effects on test scores can fade over time, and less is known about the lasting effects of such interventions. We leverage data from a Norwegian large-scale field experiment to examine the effects of small-group instruction in mathematics for students aged 7-9. This intervention shares many features with other high-impact tutoring programs, with some notable exceptions:

instruction time was kept fixed, it had a lower dosage, and it targeted students of all ability levels. The latter allows us to assess fadeout across the ability distribution. Previous research on this intervention finds positive short-run effects. This paper shows that about 60% of the effect persists 3.5 years later. The effect size and degree of fadeout are surprisingly similar across the ability distribution. The study demonstrates that small-group instruction in mathematics successfully targets student performance and that effects can be sustained over time.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-931

The Effects of In-School Virtual Tutoring on Student Reading Development: Evidence from a Short-Cycle Randomized Controlled Trial

Douglas Ready, Teachers College, Columbia University; Sierra McCormick, Teachers College, Columbia University; Rebecca Shmoys, Teachers College, Columbia University

Abstract: Over the past several years, empirical studies have reported positive impacts of one-to-one and small-group tutoring on student academic development. This literature, however, also points to major implementation challenges, including staffing costs and low student participation rates. This paper describes a 12-week cluster randomized controlled trial that examined the efficacy of a virtual literacy-focused tutoring intervention that sought to address limitations in prior models and implementations by providing live tutoring services during the school day. Intent-to-Treat models indicate that first through fourth-grade students assigned to the program outperformed their control-group peers by roughly 0.05 SDs. However, only 20% of students received the recommended tutoring dosage. Given the variability in usage rates, we also leveraged Treatment-on-the-Treated approaches. These models suggest that students who completed 10 or more tutoring sessions experienced a reading advantage of 0.08 SDs, while those who completed 20 or more sessions—the recommended dosage—experienced a 0.26 SD developmental advantage. Further research should explore how implementation conditions, school characteristics, and student needs influence uptake rates and program effectiveness.

Outlet: Journal of Education for Students Placed at Risk

Link: https://www.tandfonline.com/doi/abs/10.1080/10824669.2026.2615787

The Impact of High-Impact Tutoring on Student Attendance: Evidence from a State Initiative

Monica G. Lee, Stanford University; Susanna Loeb, Stanford University; Carly D. Robinson, Stanford University

Abstract: Student absenteeism, which skyrocketed during and after the COVID-19 pandemic, has negative consequences for student engagement and achievement. This study examines the impact of the High-Impact Tutoring (HIT) Initiative, implemented by the Office of the State Superintendent of Education in Washington

DC, on reducing absenteeism. The HIT initiative was designed to mitigate learning loss by providing additional academic supports with a focus on students affected by the pandemic's disruptions. Leveraging detailed daily school attendance and tutoring session data, we employ a within-student approach with student and date fixed effects to isolate the causal effect of having a scheduled tutoring session on daily school attendance. We find that the likelihood of being absent decreases by 1.2 percentage points on days when students have a scheduled tutoring session; this translates to a 7.0% reduction in absenteeism. These effects are most pronounced among middle school students and those with extreme absenteeism in the prior year, with reductions of 13.7% and 7.0%, respectively. Furthermore, key features of high-impact tutoring, such as in-school delivery and smaller tutor-to-student ratios, amplify the effect. These findings underscore the dual benefits of high-impact tutoring for both academic and engagement outcomes, highlighting its potential as a scalable strategy to addressing chronic absenteeism and promoting equitable access to supportive educational environments.

Outlet: AERA Open

Link: https://edworkingpapers.com/ai25-1107

Impacts of Saga Blended Learning on Math Achievement After COVID-19

Laura Meyer, Mathematica; Lily Fesler, Mathematica; Anna Gu, Mathematica; Caroline Lauver, Mathematica

Abstract: We conducted an analysis of Saga Education's (Saga) high-dosage blended learning tutoring in three school districts across the United States during the 2021–2022 school year, when schools were still struggling with the COVID-19 pandemic and its repercussions for student learning. These blended models alternate student tutoring in small groups with individual adaptive online math practice. Our evaluation explored the implementation of these models, the impacts of Saga, and the relationships between implementation factors and impacts. For the impact evaluation, we compared outcomes for Saga students to a group of "non-Saga" students who did not participate in Saga but who were otherwise similar to Saga students. This brief summarizes findings and provides recommendations for action based on those findings.

Outlet: Mathematica

Link:
https://www.mathematica.org/publications/impacts-of-blended-learning-tutoring-models-on-math-achievement-after-covid-19

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

Lily Fesler, Mathematica; Anna Gu, Mathematica; Greg Chojnacki, Mathematica

Abstract: Air Tutors is an online tutoring organization that provides one-on-one and small group tutoring for K-12 students. Air Tutors uses live, online tutoring and an online platform that incorporates video conferencing and interactive whiteboards; recruits high-quality paid tutors with tutoring experience and engaging online

personalities; and engages students and their families via text messages, phone calls, and emails. This study aims to provide evidence of the impact of tutoring on student math achievement in a randomized controlled trial, as well as on students' self-reported math confidence, sense of belonging, and relationships with their tutors via student surveys. Specifically, it examines the impact of participating in Air Tutors math tutoring sessions on 4th-, 5th-, and 6th-grade students' math achievement, as well as changes in student math confidence. This report is one in a series of six reports on math tutoring programs. The goal of this report series is to inform the tutoring field more broadly and support the provision of high-quality tutoring to as many students in the priority communities as possible. [This report was prepared with Air Tutors.]

Outlet: Mathematica

Link: https://eric.ed.gov/?id=ED628638

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

Matthew Kraft, Brown University; Danielle Sanderson Edwards, Old Dominion University; Marisa Cannata, Vanderbilt University

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

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-1030

Realizing the Promise of High Dosage Tutoring at Scale Preliminary Evidence for the Field

Monica P. Bhatt, Terence Chau, Barbara Condliffe, Rebecca Davis, Jean Grossman, Jonathan Guryan, Jens Ludwig, Matteo Magnaricotte, Shira Mattera, Fatemeh Momeni, Philip Oreopolous, and Greg Stoddard

Abstract: This technical report outlines preliminary results from the Personalized Learning Initiative showing that high dosage tutoring can be scaled and can work – even when delivered in the aftermath of the pandemic and in diverse academic settings. Students who participated in tutoring during the school day saw large and positive gains on end-of-year test scores, at least in math; the results for reading are not yet conclusive. The impact on math scores is equivalent to about two-thirds of a year of learning, which would be enough to totally undo the effects of the pandemic for the average student. The impact on reading scores are inconclusive so far, but data collection is ongoing. These results provide hints that high dosage tutoring delivered in school can be scaled successfully.

Outlet: Education Lab brief

Link: Realizing the Promise of High Dosage Tutoring at Scale: Preliminary Evidence for the Field

Beyond the One-Teacher Model: Experimental Evidence on Using Embedded Paraprofessionals as Personalized Instructors

Elizabeth Huffaker, University of Florida; Monica G. Lee, Stanford University; Helen Zhou, Harvard University; Carly D. Robinson, Stanford University; and Susanna Loeb, Stanford University

Abstract: Using embedded paraprofessionals to provide personalized instruction is a promising model for differentiating instruction within the classroom. This study examines two randomized controlled trials of paraprofessional-led tutoring in early-grade math and literacy. However, intent-to-treat (ITT) analyses revealed no overall achievement impacts for either program. We then explore two mechanisms that have surfaced in the tutoring literature as central efficacy moderators—dosage and tailoring—as plausible explanations to these results. While dosage was low for both programs, we estimate significant benefits from treatment assignment at higher-dosage campuses in numeracy (i.e., up to 0.28 SD at 80% progression) but no effect at any level of observed dosage on literacy. Curricular analysis revealed the literacy program's rigid structure may have impeded adaptation to student proficiency while student skill did not predict differences in numeracy program impacts. Supplemented by tutor survey data, these findings suggest that successful implementation of para-tutoring may depend on role prioritization, instructional coordination, and the use of student data to provide responsive instruction.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1326

Recruitment, Training, & Support: Impacts of A District-Partnership Tutoring Model to Support Elementary Literacy Skills

Catherine Asher and Robin Jacob

Abstract: This study presents findings from a multi-site student-level stratified randomized evaluation of Ed Corps, a tutoring program that provides tutor recruitment and training as well as implementation support for

high-impact tutoring. We find that relative to a business-as-usual control, students in grades K-3 assigned to receive Ed Corps-supported tutoring score 0.15 SD higher on their end-of-year DIBELS Composite scores and 0.18 SD higher on a researcher developed index aligned to the specific literacy skills each student struggled with at the beginning of the year. While we find positive (though not significant) impacts for a wide variety of student and school subgroups, there is some evidence that boys may benefit more than girls, potentially due to different experiences in Tier 1 instruction.

Outlet: Youth Policy Lab, University of Michigan

Link: https://youthpolicylab.umich.edu/sites/ypl/files/2026-01/NCEC%20working%20paper.pdf

Investing in Human Capital During Wartime: Experimental Evidence from Ukraine

Lelys Dinarte-Diaz, World Bank; James Gresham, World Bank; Renata Lemos, World Bank; Harry A Patrinos, University of Arkansas; Rony Rodriguez-Ramirez, Harvard University

Abstract: This paper provides insights into human capital investments during wartime by presenting evidence from three experiments of an online tutoring program for Ukrainian students amid Russia's invasion of Ukraine. Conducted between early 2023 and mid-2024, the experiments reached nearly 10,000 students across all regions of Ukraine. The program offered three hours per week of small-group tutoring in math and Ukrainian language over six weeks, combining academic instruction with psychosocial support. Results show that the program led to substantial improvements in learning—up to 0.49 standard deviations in math and 0.40 standard deviations in Ukrainian language—and consistent reductions in stress—up to 0.12 standard deviations. We observe high take-up-and-engagement rates and identify four mechanisms as drivers of impact: structured peer interactions, improved attitudes toward learning, enhanced socio-emotional skills, and increased student investments. A complementary experiment using information nudges to increase parental engagement highlights challenges in promoting parental investments in students' education in a conflict setting. The program was cost-effective, with benefit-to-cost ratios between 16.6 and 31, and scalable given its reliance on existing educational infrastructure and teaching capacity.

Outlet: EdWorkingPapers

Link: https://renatalemos.com/wp-content/uploads/2026/03/ukraine_03222026.pdf

Online Tutoring, School Performance, and School-to-Work Transitions: Evidence from a Randomized Controlled Trial

Silke Anger, IAB Nürnberg; Bernhard Christoph, IAB Nürnberg; Agata Galkiewicz, University Potsdam; Shushanik Margaryan, University Potsdam; Malte Sandner, IAB Nürnberg

Abstract: Tutoring programs for low-performing students, delivered in-person or online, effectively enhance school performance, yet their medium- and longer-term impacts on labor market outcomes remain less understood. To address this gap, we conduct a randomized controlled trial with 839 secondary school students in Germany to examine the effects of an online tutoring program for low-performing students on academic performance and school-to-work transitions. The online tutoring program had a non-significant intention-to-treat

effect of 0.06 standard deviations on math grades six months after program start. However, among students who had not received other tutoring services prior to the intervention, the program significantly improved math grades by 0.14 standard deviations. Moreover, students in non-academic school tracks experienced smoother school-to-work transitions, with vocational training take-up 18 months later being 5 percentage points higher—an effect that was even larger (12 percentage points) among those without prior tutoring. Overall, the results indicate that tutoring can generate lasting benefits for low-performing students that extend beyond school performance.

Outlet: European Economic Review

Link: https://docs.iza.org/dp18307.pdf

Coursemojo Efficacy Study: Grade 6, 2024-2025 NWEA MAP & STAAR

Rachel Schechter; Michaela Gulemetova; Colin Ackerman; Laura Janakiefski; LXD Research

Abstract: This quasi-experimental study investigated the impact of Coursemojo, an AI-powered literacy platform, on sixth-grade English Language Arts achievement in a large Texas district during the 2024-2025 academic year. The study compared 541 students using Coursemojo across two schools with 2,786 comparison students across 11 schools. Students using Coursemojo significantly outperformed comparison students on the Texas STAAR assessment, with 70% achieving passing levels compared to 60% in the comparison group. The platform demonstrated particularly strong effects for historically underserved populations: special education students showed substantial gains on NWEA MAP assessments (effect sizes of 0.21-0.36), while economically disadvantaged students significantly outperformed peers on STAAR by 13.8 points and showed higher proportions of students passing and meeting grade level expectations. Dosage analysis revealed threshold effects, suggesting that usage for at least 12 weeks of implementation and 1.7 days per week of engagement is important to see meaningful improvements. Educator feedback was universally positive, with 100% teachers expressing high satisfaction and 78% of students reporting that the platform helped their learning. The study provides evidence that Coursemojo's AI-powered instruction can enhance educational equity when properly implemented with comprehensive support systems.

Outlet: ERIC

Link: https://eric.ed.gov/?id=ED677099

Coursemojo Efficacy Study: Grade 6, 2024-2025, Aimsweb & TCAP

Rachel Schechter; Michaela Gulemetova; Colin Ackerman; Laura Janakiefski; LXD Research

Abstract: This quasi-experimental study examined the impact of Coursemojo's AI-powered teaching platform on sixth-grade English Language Arts achievement in a large, diverse Tennessee school district during the 2024-2025 school year. The study involved 2,203 students across twelve schools, with 1,017 students using Coursemojo and 1,186 serving as comparison students. Results demonstrated significant positive effects across multiple assessment measures. Coursemojo students outperformed comparison students by 8 points on the state standardized test (TCAP) ([beta] = 7.461, p < 0.001, Hedges' g = 0.12). The platform also showed

evidence of supporting more equitable outcomes, reducing achievement gaps for students with disabilities by two-thirds (from 16 to 5 points) and for economically disadvantaged students by more than half. Dosage analysis revealed that more sustained implementation of 2.5+ days per week on the weeks the program was used was critical for maximizing benefits. Survey data from 851 students and 22 teachers, along with interviews with 4 educational leaders, provided convergent evidence of positive learning impacts and effective instructional integration. Student surveys revealed that 39.7% rated Coursemojo as helping them learn "quite a bit" or "very much," while 80.0% provided high effort ratings, with persistence scores averaging 4.03 out of 5. These findings suggest that AI-powered literacy support can enhance middle school English Language Arts instruction while advancing equity outcomes for historically underserved populations.

Outlet: ERIC

Link: https://eric.ed.gov/?id=ED676253

Air Reading: A Randomized Evaluation of a Virtual Tutoring Model in Louisiana and Texas Schools

Amanda J. Neitzel; Nathan Storey; and Xue Wang; Johns Hopkins University

Abstract: This study employed a randomized controlled trial (RCT) to estimate the causal impact of Air Reading, a structured virtual tutoring program, on elementary students' literacy outcomes. Randomization occurred at the student level within schools and grade cohorts, ensuring that treatment and control students were directly comparable at baseline. The sample included students in grades 1–4 randomized to either the Air Reading program (n = 174) or the business-as-usual control condition (n = 203). Two primary types of data were collected to assess program impacts: student achievement outcomes and program usage indicators. On average, treatment students attended 55 sessions, with 56% reaching the high-dosage threshold of 56 or more sessions, translating to an average of 27.1 hours of tutoring per student. Across two districts in a full year implementation, Air Reading produced consistently positive impacts on early literacy outcomes. Students assigned to receive tutoring significantly outperformed their control group peers, with an effect size of +0.29 SD, equivalent to 2.8 additional months of learning. Subgroup analyses showed that the benefits of Air Reading were broadly distributed across student groups. Students who reached the high-dosage threshold of 56 or more sessions showed significantly greater achievement gains compared with their control peers, whereas low-dosage students demonstrated smaller, though still significant, achievement gains compared with control peers. Taken together, these findings indicate that Air Reading can produce meaningful improvements in early literacy, particularly when implemented across a full school year. At the same time, the absence of strong subgroup variation suggests that the program's effects are not narrowly concentrated but instead benefit a broad range of students.

Outlet: JScholarship

Link: https://jscholarship.library.jhu.edu/handle/1774.2/71611

Personalized Learning Initiative interim Report: Findings from 2023-2024

Monica P. Bhatt, University of Chicago Education Lab; Terence Chau, University of Chicago Education Lab; Barbara Condliffe, MDRC; Rebecca Davis, University of Delaware; Jean Grossman, Princeton University; Jonathan Guryan, University of Chicago Education Lab; Jens Ludwig, University of Chicago Education Lab; ; Matteo Magnaricotte, University of Chicago Education Lab; Shira Mattera, MDRC; Philip Oreopolous, University of Toronto; and Greg Stoddard, University of Chicago Education Lab

Abstract:

1. Tutoring - both high dosage tutoring and sustainable high dosage tutoring - is effective overall.
    - Pooled analyses show the effect of participating in tutoring is statistically significant and ranges from 0.06-0.09 SD, or approximately 1-2 months of additional learning. These overall effects mask considerable variability across sites.
2. Tutoring impacts seem robust across a variety of models.
    - Lower cost models ($1200 per student) are just as effective as higher cost models ($2000 per student).
    - Virtual tutoring seems just as effective as in person tutoring in PLI sites.
3. More tutoring minutes correlate with greater learning gains.
    - But, minutes of tutoring provided are much lower than past tutoring studies (corresponding to smaller gains in student learning).

Outlet: University of Chicago Education Lab

Link:
https://educationlab.uchicago.edu/resources/personalized-learning-initiative-interim-report-findings-from-2023-24/

EFFECTS OF PROGRAM CHARACTERISTICS

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

Carly Robinson, Stanford University; Biraj Bisht, UC Irvine; Susanna Loeb, Stanford University

Abstract: Billions of dollars are invested in opt-in educational resources to support struggling students. Yet, there is no guarantee these students will use these resources. We report results from a school system's implementation of on-demand tutoring. The take up was low. At baseline, only 19% of students ever accessed the platform and low-performing students were even less likely to log in. We conducted a randomized controlled trial (N=4,763) testing behaviorally-informed messages directed at students and/or their parents to increase participation. Communications to students alone had no impact, whereas those to parents and students together increased usage by 46%. Nonetheless, take-up remained low, highlighting that opt-in resources may increase—instead of reduce—inequality. Without targeted outreach, opt-in educational resources are unlikely to reach many students who could benefit.

Outlet: Educational Researcher

Link: https://doi.org/10.3102/0013189X251331518

Can Technology Facilitate Scale? Evidence from a Randomized Evaluation of High-Dosage Tutoring

Monica P. Bhatt, University of Chicago; Jonathan Guryan, Northwestern University; Salman Khan, University of Chicago; Michael LaForest-Tucker, U.S. Air Force Academy; Bhavya Mishra, University of Chicago

Abstract: High-dosage tutoring is an effective way to improve student learning (Nickow et al., 2024; Guryan et al., 2023). Finding ways to deliver high-dosage tutoring at large scale remains a challenge. Two primary challenges to scaling are cost and staffing. One possible solution is to reduce costs by substituting some tutor time with computer-assisted learning (CAL) technology. The question is: Does doing so compromise effectiveness? This paper provides evidence from a randomized controlled trial (RCT) of approximately 4,000 students in two large school districts in 2018- 19 and 2019-20. The RCT tested the effectiveness of an in-school math tutoring program where students worked in groups of four, with two students working with an in-person tutor while the other two worked on CAL, alternating every other day. The tutoring model had per-pupil costs approximately 30 percent lower than the 2- to-1 tutoring model studied in Guryan et al. (2023). We find gains in students' math standardized test scores of 0.23 standard deviations for participating students, which are almost as large as the effect sizes of the 2-to-1 tutoring model reported in Guryan et al. (2023). These findings suggest strategic use of technology may be a way to increase the scalability of HDT.

Outlet: National Bureau of Economic Research

Link: https://www.nber.org/papers/w32510

Answering the call: How changes to the salience of job characteristics affects college students' decisions

Carly Robinson, Stanford University; Katharine Meyer, Brookings Institution; Chasity Bailey-Fakhoury, Grand Valley State University; Amirpasha Zandieh, University of Colorado Boulder; Susanna Loeb, Stanford University

Abstract: College students make job decisions without complete information. As a result, they may rely on misleading heuristics ("interesting jobs pay badly") and pursue options misaligned with their goals. We test whether highlighting job characteristics changes decision making. We find increasing the salience of a job's monetary benefits increases the likelihood college students apply by 196%. In contrast, emphasizing prosocial, career, or social benefits has no effect, despite students identifying these benefits as primary motivators for applying. The study highlights the detrimental incongruencies in students' decision making alongside a simple strategy for recruiting college students to jobs that offer enriching experiences.

Outlet: Social Science Research

Link: https://edworkingpapers.com/ai24-956

Reading Ready: An Evaluation of a High-Dosage Tutoring Program for Emergent Readers

Shirin Hashim, Hollyhock Foundation; Erin Croke, CUNY; Katie Pace Miles, CUNY

Abstract: This study presents the first within-program, within-tutor experimental evidence comparing the impact of in-person versus remote tutoring. Based on results from an early literacy tutoring initiative delivered by university students over Summer 2023, we find no statistically significant differences in students' literacy outcomes by instructional modality. However, students receiving in-person tutoring exhibited higher attendance rates and tutors reported closer relationships with their in-person students. Notably, we find substantial variation in students' outcomes due to differences among tutors, while these effects do not vary by modality. These findings suggest that while differences between in-person and remote tutoring may exist, the advantages of having a proficient tutor greatly outweigh these disparities. The study underscores the efficacy of remote tutoring, particularly when geographical constraints are a factor, and highlights the necessity of including interpersonal skills in tutor training, ensuring consistent attendance and program fidelity, and identifying and retaining highly effective tutors to maximize student learning.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1176

The Impact of Tutor Gender Match on Girls' STEM Interest, Engagement, and Performance

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

Abstract: Gender disparities in STEM persist despite girls performing as well as boys academically, suggesting girls may benefit from role models who shape their perceptions of STEM. We examine whether female math tutors influence girls' STEM interest, attendance, and performance. We randomly assigned 422 ninth-grade students taking Algebra 1 to a same-gender or opposite-gender tutor. Girls assigned to female tutors reported higher STEM interest (.73 SD) and were more likely to pass the course with a C– or better (3.9 percentage points) than those with male tutors. We found no impact on attendance. Effects were stronger for students working with tutors in-person rather than virtually. We provide the first experimental evidence that female tutors can boost girls' STEM self-concept and academic outcomes.

Outlet: American Educational Research Journal

Link: https://journals.sagepub.com/doi/abs/10.3102/00028312261417332

The Trade-off between Quality and Quantity: Evidence from a Field Experiment on Tutoring

Rohen Shah, Yale University

Abstract: High-dosage tutoring has the potential to substantially raise adolescent academic achievement. However, at scale, schools may not have the financial ability to deliver small-group tutoring frequently. In this paper, I test the relative importance of group size (quality) versus tutoring frequency (quantity). I evaluate the impact of an in-school math tutoring program in a middle school in the Midwestern United States. Students are randomized to either 1) control, 2) receive tutoring twice a week in 2-student groups, or 3) receive tutoring

three times a week in 3-student groups. Importantly, the total cost per student is the same in both treatment conditions. I find that the 2-student group tutoring led to a significant improvement in math skills (0.23 SD), whereas the equal-cost, more frequent tutoring in the 3-student groups did not lead to a significant improvement in math skills.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/sites/default/files/ai26-1379.pdf

Improving Student Learning with Hybrid Human-AI Tutoring: A Three-Study Quasi-Experimental Investigation

Danielle R Thomas, Carnegie Mellon University; Jionghao Lin, Carnegie Mellon University; Erin Gatz, Carnegie Mellon University; Ashish Gurung, Carnegie Mellon University; Shivang Gupta, Carnegie Mellon University: Kole Norberg, Carnegie Mellon University; Stephen E. Fancsali, Carnegie Mellon University; Vincent Aleven, Carnegie Mellon University; Lee Branstetter, Carnegie Mellon University; Emma Brunskill, Carnegie Mellon University; Kenneth R. Koedinger, Carnegie Mellon University.

Abstract: Artificial intelligence (AI) applications to support human tutoring have potential to significantly improve learning outcomes, but engagement issues persist, especially among students from low-income backgrounds. We introduce an AI-assisted tutoring model that combines human and AI tutoring and hypothesize this synergy will have positive impacts on learning processes. To investigate this hypothesis, we conduct a three-study quasi-experiment across three urban and low-income middle schools: 1) 125 students in a Pennsylvania school; 2) 385 students (50% Latinx) in a California school, and 3) 75 students (100% Black) in a Pennsylvania charter school, all implementing analogous tutoring models. We compare learning analytics of students engaged in human-AI tutoring compared to students using math software only. We find human-AI tutoring has positive effects, particularly in student's proficiency and usage, with evidence suggesting lower achieving students may benefit more compared to higher achieving students. We illustrate the use of quasi-experimental methods adapted to the particulars of different schools and data-availability contexts so as to achieve the rapid data-driven iteration needed to guide an inspired creation into effective innovation. Future work focuses on improving the tutor dashboard and optimizing tutor-student ratios, while maintaining annual costs per student of approximately $700 annually

Outlet: Association for Computing Machinery

Link: https://doi.org/10.1145/3636555.3636896

Creating Coherence: Does Instructional Alignment Affect the Impact of Tutoring?

Cara Jackson, The Center for Outcomes Based Contracting at SEF; Ayman Shakeel, Abt Global

Abstract: This study examines the impact of using instructionally aligned literacy tutoring with students in kindergarten through third grade under a Response to Intervention framework. We conducted a randomized controlled trial to evaluate the impact on literacy assessment scores for 296 students in four schools in a large suburban school district in the southeastern United States. Students in the treatment group received tutoring where strategies and materials were aligned with core instruction, while those in the control group received

tutoring that used supplemental strategies and materials that were distinct from core instruction. We find that students in the treatment group score an average of 0.12 standard deviations higher than the students in the control group. Exploratory analyses reveal that instructional alignment appears to have a greater impact on boys and lower-performing students. Additional exploratory analyses suggest the treatment effect is stronger when delivered in groups of four and by tutors who do not hold a master's degree.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1332

NOVEL METHODS

Educator Attention: How computational tools can systematically identify the distribution of a key resource for students

Qingyang Zhang, Asana; Rose E. Wang, Stanford University; Ana Ribeiro, Stanford University; Susanna Loeb, Stanford University; Dorottya Demszky, Stanford University

Abstract: Educator attention is critical for student success, yet how educators distribute their attention across students remains poorly understood due to data and methodological constraints. This study presents the first large-scale computational analysis of educator attention patterns, leveraging over 1 million educator utterances from virtual group tutoring sessions linked to detailed student demographic and academic achievement data. Using natural language processing techniques, we systematically examine the recipient and nature of educator attention. Our findings reveal that educators often provide more attention to lower-achieving students. However, disparities emerge across demographic lines, particularly by gender. Girls tend to receive less attention when paired with boys, even when they are the lower achieving student in the group. Lower-achieving female students in mixed-gender pairs receive significantly less attention than their higher-achieving male peers, while lower-achieving male students receive significantly and substantially more attention than their higher-achieving female peers. We also find some differences by race and English learner (EL) status, with low-achieving Black students receiving additional attention only when paired with another Black student but not when paired with a non-Black peer. In contrast, higher-achieving EL students receive disproportionately more attention than their lower-achieving EL peers. This work highlights how large-scale interaction data and computational methods can uncover subtle but meaningful disparities in teaching practices, providing empirical insights to inform more equitable and effective educational strategies.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1144

Three Paradoxes to Reconcile to Promote Safe, Fair, and Trustworthy AI in Education

Rachel Slama, RAND Corporation; Amalia Toutziardi, Massachusetts Institute of Technology; Justin Reich, Massachusetts Institute of Technology

Abstract: Incorporating recordings of teacher-student conversations into the training of LLMs has the potential to improve AI tools. Although AI developers are encouraged to put "humans in the loop" of their AI safety protocols, educators do not typically drive the data collection or design and development processes underpinning new technologies. To gather insight into privacy concerns, the adequacy of safety procedures, and potential benefits of recording and aggregating data at scale to inform more intelligent tutors, we interviewed a pilot sample of teachers and administrators using a scenario-based, semi-structured interview protocol. Our preliminary findings reveal three "paradoxes" for the field to resolve to promote safe, fair, and trustworthy AI. We conclude with recommendations for education stakeholders to reconcile these paradoxes and advance the science of learning.

Outlet: ACM

Link: https://dl.acm.org/doi/10.1145/3657604.3664658

Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise

Rose E Wang, Stanford University; Ana Ribeiro, Stanford University; Carly Robinson, Stanford University; Susanna Loeb, Stanford University; Dorottya Demszky, Stanford University

Abstract: Generative AI, particularly Language Models (LMs), has the potential to transform real-world domains with societal impact, particularly where access to experts is limited. For example, in education, training novice educators with expert guidance is important for effectiveness but expensive, creating significant barriers to improving education quality at scale. This challenge disproportionately hurts students from under-served communities, who stand to gain the most from high-quality education and are most likely to be taught by inexperienced educators. We introduce Tutor CoPilot, a novel Human-AI approach that leverages a model of expert thinking to provide expert-like guidance to tutors as they tutor. This study presents the first randomized controlled trial of a Human-AI system in live tutoring, involving 900 tutors and 1,800 K-12 students from historically under-served communities. Following a preregistered analysis plan, we find that students working on mathematics with tutors randomly assigned to have access to Tutor CoPilot are 4 percentage points (p.p.) more likely to master topics ($p<0.01$). Notably, students of lower-rated tutors experienced the greatest benefit, improving mastery by 9 p.p. relative to the control group. We find that Tutor CoPilot costs only $20 per-tutor annually, based on the tutors' usage during the study. We analyze 550,000+ messages using classifiers to identify pedagogical strategies, and find that tutors with access to Tutor CoPilot are more likely to use strategies that foster student understanding (e.g., asking guiding questions) and less likely to give away the answer to the student, aligning with high-quality teaching practices. Tutor interviews qualitatively highlight how Tutor CoPilot's guidance helps them to respond to student needs, though tutors flag common issues in Tutor CoPilot, such as generating suggestions that are not grade-level appropriate. Altogether, our study of Tutor CoPilot demonstrates how Human-AI systems can scale expertise in real-world domains, bridge gaps in skills and create a future where high-quality education is accessible to all students.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai24-1054

Understanding Disruptions to Virtual Learning: Causes of and Variation in Lost Instructional Time

Xander Beberman, Stanford University; Sarah Novicoff, Public Policy Institute of California; Ana Trindade Ribeiro, Stanford University; Carly Robinson, Stanford University; and Susanna Loeb, Stanford University

Abstract: Virtual instruction has boomed after the COVID-19 pandemic, including the use of virtual environments within in-person schools. But, research has provided little evidence about student experiences on these virtual platforms, nor how to improve the use of these platforms. Through natural language processing techniques, this study examines over 26,000 virtual tutoring sessions that took place within in-person schools to identify the frequency of disruptions during virtual learning. We find that 81 percent of allocated instructional time was undisrupted. Technology problems consumed approximately nine percent of planned instructional time, while student disruptions consumed another seven percent. We find substantial variation across schools, with fewer disruptions when tutoring occurred in dedicated environments such as libraries or tutoring-specific classrooms.

Outlet: EdWorkingPapers

Link: https://edworkingpapers.com/ai25-1239

The Power of Personalized Attention: Comparing Pedagogical Approaches in Small Group and One-on-One Early Literacy Tutoring

Hsiaolin Hsieh, Stanford University; David Gormley, Stanford University; Carly D. Robinson, Stanford University; Susanna Loeb, Stanford University

Abstract: Tutoring has played a significant role in pandemic-related learning recovery, supporting student learning and engagement. This paper follows up on a recent randomized controlled trial (RCT) estimating that one-on-one virtual early literacy tutoring was nearly twice as effective as two-on-one tutoring for improving student learning. To better understand this gap, we analyze transcripts from 16,629 tutoring sessions from this RCT—which included over 3.7 million tutor utterances—using natural language processing and machine learning techniques. We explore how tutors allocate attention across content instruction, relationship building, and classroom management between one-on-one and two-on-one formats. While tutors dedicate similar time to content instruction and relationship building across both formats, students receiving one-on-one tutoring receive more attention and personalized support. To improve the effectiveness of two-on-one tutoring, it may be beneficial to equip tutors with strategies that engage multiple students simultaneously, thereby reducing downtime and minimizing the potential for disengagement

Outlet: Education Sciences

Link: https://www.mdpi.com/3685182