Quarterly
Research Note

Accelerate
The National Collaborative for Accelerated Learning

Introduction

Welcome to the second issue of the Quarterly Research Note (QRN), a research brief that reflects Accelerate's approach to continuous learning through ongoing inquiry into what educational interventions work, for which students, and under what conditions. In this issue, we profile Accelerate's latest research report, "Contextualizing the Impact of Tutoring on Student Learning: Efficiency, Cost Effectiveness, and the Known Unknowns," in which we introduce two novel metrics for evaluating the time efficiency and cost-effectiveness of high-dosage tutoring programs. These new metrics make it possible for future research to assess the return on investment of various tutoring programs and educational interventions, with the ultimate goal of helping schools make evidence-informed decisions.

In this issue's Research Roundup, we focus on recent studies which examine the impact of tutoring and personalized learning initiatives on student learning outcomes. We profile studies conducted by Accelerate's research partners at the University of Chicago Education Lab and the Center for Education Policy Research (CEPR) at Harvard University, and present evidence on two Accelerate grantees: Once, which provided short-bursts of targeted early literacy tutoring; and Intervene K-12, which provided math tutoring to grade 9 English Language Learner (ELL) students.

In Looking Ahead, we present Accelerate's research agenda, which calls for significantly more evidence to identify the tutoring providers that meaningfully improve student achievement for students in different grade levels, with different educational needs, and in different schooling environments, as well as additional evidence to determine which features of tutoring program design are associated with improvements in student achievement. Accelerate's research agenda aims to ensure that the students who most need academic support are receiving the most efficient and cost-effective educational interventions.

Inside this Quarterly Research Note:

1. Accelerate's Report on Tutoring Evidence, Efficiency, and Cost Effectiveness

2. Research Roundup
   a. Personalized Learning Initiative (University of Chicago Education Lab)
   b. Road to Recovery (Harvard Center for Education Policy Research)
   c. Once
   d. Intervene K-12

3. Looking Ahead: Accelerate's Research Agenda

New Evidence and Novel Metrics on Tutoring Impacts

In May 2024, Accelerate released a research report titled "Contextualizing the Impact of Tutoring on Student Learning: Efficiency, Cost Effectiveness, and the Known Unknowns."

In this report, we:

i. Summarize existing evidence on tutoring program impact from Nickow et al. (2023), a recent meta-analysis of tutoring program impacts;

ii. Construct a new measure of the efficiency of tutoring program impact, which we define as the hours of tutoring necessary to improve student learning by one month, and compare the efficiency of tutoring program impact across select tutoring providers based on evidence from well-designed randomized control trials (RCTs);

iii. Propose an approach to measure the cost effectiveness of tutoring program impact, which we define as the additional months of student learning produced at a cost of $1,000 per pupil; and

iv. Lay out a research agenda calling for significantly more evidence on the specific tutoring programs that improve student learning, for which students, in what educational contexts, at what fiscal cost, and the programmatic features most associated with student achievement gains.

In this report, we show that evidence confirms that high-dosage tutoring remains the most effective academic intervention, outperforming other educational interventions such as reduced class sizes, summer school, and extended school day/year. Among the 89 RCTs profiled in Nickow et al. (2023), small-scale studies – those with smaller sample sizes – tend to have greater effects on student learning, on average, than large-scale studies. Tutoring programs with teacher-led tutoring have larger effects than tutoring provided by non-professionals and parents, and tutoring providers that offer more intensive dosage (at least three sessions per week of tutoring) have greater impacts on student learning than providers with less intensive dosage (fewer than three sessions per week). Yet, we underscore the fact that the evidentiary base is less robust than initially expected; just 15 of the 89 RCTs included in Nickow et al. (2023) would meet the highest standards of evidence on program impact (i.e., ESSA Tier 1 evidence standards based on sample size criteria - namely, that the study includes a minimum of 350 study participants).

We then reanalyzed tutoring program impact data from 14 high-quality RCTs

**Tutoring Efficiency: Hours of tutoring to improve student learning by one month**

$$\text{Tutoring Efficiency} = \frac{\text{Hours of intended dosage}}{\text{\# of additional months of learning gained by tutoring}}$$

Cost Effectiveness: Number of additional months of learning gained by investing $1,000 per pupil

$$\text{Cost Effectiveness} = \frac{\text{Hours of tutoring for \$1,000 per pupil}}{\text{Tutoring Efficiency}}$$

of 12 tutoring providers that met the evidence bar researchers, educators, and policymakers should demand all providers meet. We define a new approach for calculating the **efficiency** of tutoring program impact, and show that, among this select sample of 12 tutoring providers, math tutoring more efficiently improves student learning (and with less variability) than literacy tutoring (please see Figure 1). We then offer guidance to policymakers and researchers on the collection of valid programmatic cost data necessary to calculate our proposed measure of the **cost-effectiveness** of tutoring programs.

Figure 1. Distribution of Math and Literacy Efficiency Estimates, By Tutoring Program
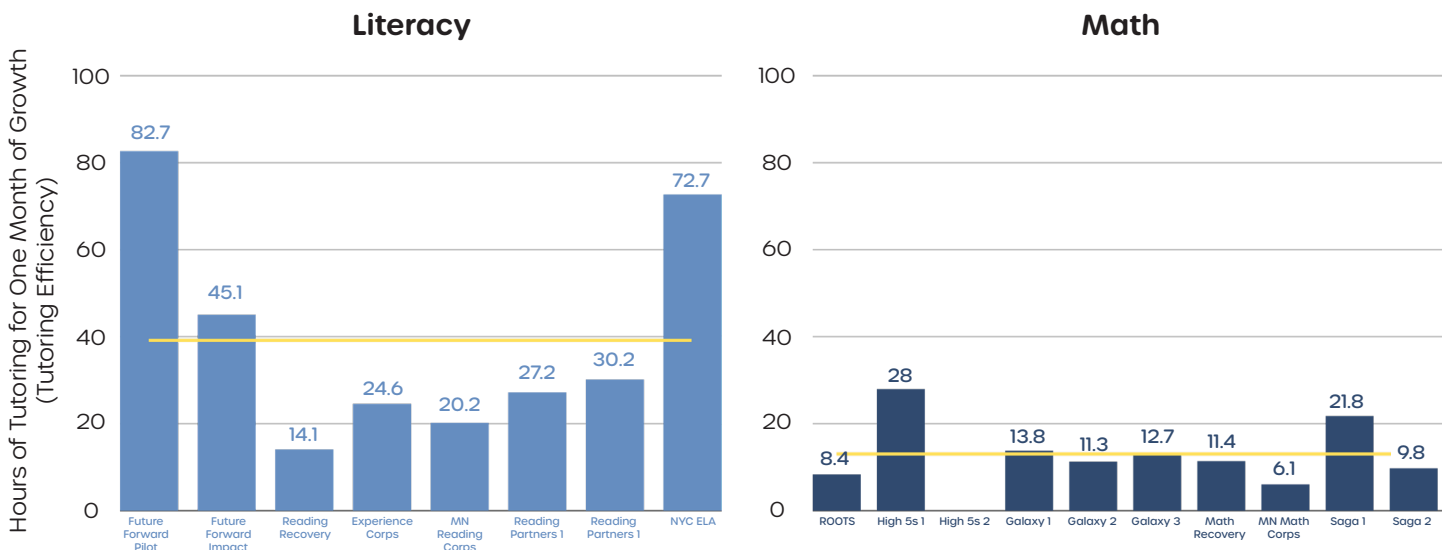
The aim of this report is to provide policymakers, school leaders, and researchers with an approach to compare (and select) tutoring programs that most efficiently improve student learning at the lowest cost, and to support ongoing decision-making among educational leaders investing scarce resources to improve student learning. We recommend that:

- **Researchers** should focus on large-sample evaluations to produce rigorous evidence on tutoring effectiveness so that education leaders can make the most cost-effective decisions when selecting tutoring providers.
- **Tutoring providers** should gather data on all students they serve, not just those who receive the full dosage, and should provide more transparency around the cost of their programs.
- **Federal agencies** should lead high-quality research efforts and enact policies that incentivize tutoring providers to meet rigorous standards.
- **States** should identify and curate a list of vetted, evidence-based high-dosage tutoring programs that meet a standard of efficiency and mandate cost reporting for vetted tutoring providers that make it onto state approved lists.
- **School districts** should build tutoring into their budgets for the long term and rely on evidence to select only the most effective and efficient programs.

We welcome readers' thoughts and reactions to this report as we continue to engage researchers, policymakers, and school leaders in this important conversation. Please find the feedback form here.

Research Roundup

In this Research Roundup, we describe new and emerging findings on tutoring impact from Accelerate's portfolio of grantees and research partners. These new findings aim to inform ongoing efforts to identify effective tutoring models and practices, the educational settings that shape their implementation, and, ultimately, their impact on student learning. We first profile two studies of high-dosage tutoring and personalized learning initiatives conducted by two of Accelerate's research partners: the University of Chicago Education Lab; and the Center for Education Policy Research (CEPR) at Harvard University. Then, we present evidence of program impact on two of Accelerate's Call to Effective Action (CEA) grantees: Once, which provided short-bursts of targeted early literacy tutoring; and Intervene K-12, which provided math tutoring to 9th grade English Language Learner (ELL) students who had experienced interruptions in their education.

Personalized Learning Initiative: Evidence on High-Dosage Tutoring from the 2022-23 School Year

In March 2024, the University of Chicago Education Lab and MDRC released preliminary results from the Personalized Learning Initiative (PLI). The data provides impact findings from over 2,000 students who were randomly assigned to receive tutoring during the 2022-23 school year in two (of four) partner sites: Chicago Public Schools and Fulton County Schools. Outcomes from these 2,000 students comprise approximately 10 percent of the total PLI study sample to date (N=20,000+ and counting), so should be considered preliminary and are subject to change. The PLI research team found positive but statistically insignificant effects thus far on high-dosage tutoring when combining K-5 literacy and 6-12 math efforts. Encouragingly, the research team found large, positive, and statistically significant effects on student learning in middle school math; the results for reading scores are not yet conclusive.

STUDY SNAPSHOT | Personalized Learning Initiative

PUBLISHED: March 2024
RESEARCH TEAM: University of Chicago Education Lab
STUDY PERIOD: 2023-23 school year
RESEARCH METHOD: Randomized Control Trial (RCT)

STUDENT GROUP(S) STUDIED: Over 2,000 students in Chicago Public Schools and Fulton County Schools - approximately 10% of the total PLI study sample to date - who received high-dosage tutoring during the school day

STUDY QUESTIONS:
- Do local adaptations of tutoring work on average?
- Which adaptations work for which students, in which contexts?
- How can we maximize student learning for as many students as possible?

KEY FINDINGS:
- Large, positive, and statistically significant effects on student learning in middle school math; the results for reading scores are not yet conclusive.
- Positive but statistically insignificant effects thus far on high dosage tutoring when combining K-5 literacy and 6-12 math impacts.

KEY TAKEAWAYS:
- When comparing the results from this study to the results from a study of a "nights and weekend" program in New Mexico and an afterschool program in a California partner district, the evidence indicates that delivering tutoring during the school day is critical to widespread student participation and realizing large gains in student learning.
- These results are among the first indications that in-school high-dosage tutoring, done at scale, can effectively counteract pandemic-era learning setbacks.

Note: These are preliminary results and are subject to change.

Importantly, these results provide a contrast to two sites that tried to stand up high-dosage tutoring programs outside of the school day – a "nights and weekend" program in New Mexico and an afterschool program in a California partner district – in which the research team found insufficient participation or "takeup" to conduct a rigorous impact study (i.e., an RCT) despite intensive recruitment efforts. The key lesson from these four sites is that delivering tutoring during the school day – as opposed to ad-hoc, "on demand," or after school – seems critical to widespread participation and therefore realizing large gains in student learning.

These results are among the first indications that in-school high-dosage tutoring, done at scale, can effectively counteract pandemic-era learning setbacks. In addition to the impact study, the PLI has implementation and cost study components to survey tutors and school coordinators about their experiences. This data will help districts understand the costs of various tutoring program designs, including how to attract a tutor workforce. The ultimate goal of the PLI is to understand not just whether local adaptations of tutoring work on average, but, using novel machine learning techniques, understand which adaptations work for which students, in which contexts, with a key focus on maximizing student learning for as many students as possible. The research team looks forward to sharing results in the coming year.

For more information on the PLI study, contact Monica Bhatt, Senior Research Director of the University of Chicago Education Lab.

Road to Recovery: Evidence on Academic Recovery Strategies from the 2022-23 School Year

In June 2024, Accelerate's research partner at the Center for Education Policy Research (CEPR) at Harvard University, along with their co-authors at the American Institutes for Research (AIR), released a new report examining academic recovery interventions – including tutoring programs, push-in and pull-out small group instruction programs (e.g., "interventionists"), after-school programs, digital learning programs, an extended school year, double-dose classes, and an expert teacher intervention – implemented across eight school districts during the 2022-23 school year, the third year of ESSER funding. These eight districts are all participating in the Road to Recovery research project, a partnership that began in 2021 with the goals of getting districts timely feedback about the impacts of their recovery interventions and sharing this information with the field.

In total, the report examined 10 interventions that provided supplemental instruction for all or a subset of grades K-8 in math and 12 interventions that provided supplemental instruction for all or a subset of grades 1-8 in reading. For most interventions in the study, evidence indicates that they served fewer than the intended number of students, and/or participating students attended or received fewer hours of the intervention than planned. The study relied on quasi-experimental designs (QED) to estimate the impact of each intervention on students' spring 2023 math and reading achievement on NWEA MAP Growth assessments, including Value Added Models (VAMs) and/or Regression Discontinuity Designs (RDD). Of the 5 math and 8 reading tutoring and small group instruction programs for which the study team could estimate achievement impacts across the eight districts, only 2 math programs (that respectively served students in grades 5-7 and grades 3, 5, and 8) and 2 reading programs (that respectively served students in grades 4-5 and grades 4-8) had significant positive effects (ranging from +0.06 SD to +0.33 SD).

The study further finds that these significant impacts are concentrated among tutoring and small group instruction programs that served relatively few students and provided students 30 or more

hours of additional instruction per year. For other types of interventions that provided supplemental instruction time to students (e.g., extended school year, after-school, digital learning programs, and double-dose classes) researchers estimated impacts of 3 of these interventions on math and 4 of these interventions on reading; the study failed to detect significant effects of any of these interventions.

The researchers then examined the impact of being assigned to an "expert" classroom teacher in math or reading in one district and finds that these students gained more in the respective subject (+0.06 SD in math and +0.11 SD in reading) than their peers in the same grade and school.

The study findings suggest that school systems generally struggled to deliver interventions that were both effective at increasing student achievement and delivered at scale. To equip districts with information about the efficiency and scalability of an intervention, future research on academic interventions should estimate not only the average effect of an intervention on treated students, but also the intervention's effect per hour of treatment and its total effect – its average effect multiplied by the number of students it served.

Thus, this study's guidance - that future research should provide evidence on the efficiency and scalability of an intervention's impact - is consistent with guidance that Accelerate has provided in its recent report (profiled in Part 1, above); namely, that evidence on the impact of educational interventions -

> **STUDY SNAPSHOT | Road to Recovery**
>
> **PUBLISHED:** June 2024
> **RESEARCH TEAM:** Center for Education Policy Research (CEPR) at Harvard University, American Institutes for Research (AIR)
> **STUDY PERIOD:** 2023-23 school year
> **RESEARCH METHOD:** Quasi-Experimental Designs (QED), including Value Added Models (VAMs) and Regression Discontinuity Designs (RDD)
>
> **STUDENT GROUP(S) STUDIED:** Students across eight school districts
>
> **STUDY QUESTIONS:**
> * Which academic intervention(s) have the greatest impact on grades K-8 math achievement and grades 1-8 reading achievement?
>
> **KEY FINDINGS:**
> * Most interventions in the study served fewer than the intended number of students, and participating students received fewer hours of the intervention than planned.
> * Of the 5 math and 8 reading tutoring and small group instruction programs for which the study team could estimate achievement impacts, 2 math programs and 2 reading programs had significant positive effects on student achievement.
> * The most significant impacts resulted from tutoring and small group programs that served few students and provided 30 or more hours of additional instruction per year.
>
> **KEY TAKEAWAYS:**
> * School systems generally struggled to deliver interventions that were both effective and scaled to the student population. This evidence suggests that future research should estimate not just the average effect of an intervention, but also the intervention's effect per hour and its total effect (average effect multiplied by the number of students served).

such as high-dosage tutoring and other personalized learning strategies - should focus on efficiency and cost effectiveness so as to inform educational leaders and policymakers on those educational investments with the greatest return, both in terms of the time and dollars invested.

Once: Short-Bursts of Daily Early Literacy Instruction Led by Paraprofessionals

Once, an Accelerate-funded grantee, partnered with a large, urban school district on the East Coast to provide high-impact early literacy tutoring to 105 kindergarten and first grade students in 13 schools during the 2022-23 school year. The Once tutoring model aims to improve early elementary reading fluency, and includes the following program design features: (i) daily early literacy instruction for 15 minutes/day; (ii) one-on-one tutoring during the school day provided by a school-based paraprofessional; (iii) highly scripted tutoring sessions; and (iv) tutors receive ongoing coaching and

professional development support from Once's instructional coaches. Among eligible students - those scoring below grade-level benchmarks in their early literacy skills - the National Student Support Accelerator (NSSA) at Stanford University (the external evaluator) randomly assigned eligible students (at the classroom level) into either the treatment group (105 students) or control group (199 students). Students in the treatment group received one-on-one tutoring from school-based paraprofessionals for 15 minutes every day during the school day between November 2022 and June 2023; students in the control group continued to receive mainstream classroom instruction or another activity during the hours Once tutoring took place for the students in the treatment group.

Findings from NSSA's study indicate that students receiving Once tutoring did not realize substantively (or statistically) significant improvements in early literacy, as measured by the DIBELS assessment, compared to students who did not receive Once tutoring. Though the improvement in DIBELS scores was great-

er among male students (compared to their female peers) and students scoring well below their peers on the baseline (pre-treatment) DIBELS assessment, neither of these changes were statistically significant. Notably, students received, on average, just 42 15-minute tutoring sessions, far less than Once's intended dosage of 140 15-minute sessions. This suggests the need to address challenges with program implementation so as to more closely match Once's intended program dosage with the actual dosage of tutoring that students receive. In Accelerate's QRN 1.1, we discuss how the successful implementation of tutoring requires that students attend and participate in the intended number of tutoring sessions as defined by a given tutoring program's model.

Intervene K-12: Online Small-Group Math Instruction for English Language Learners

Intervene K-12, an Accelerate-funded grantee, partnered with Hartford Public Schools (HPS) during the 2022-23 school year to examine the efficacy of its online, small-group math tutoring program for grade 9 students. The study focused on Spanish-speaking English Language Learners (ELLs) who had interruptions in their education and were not on track to meet Algebra I standards. Intervene K-12's tutoring model offers a comprehensive intervention system that combines advances in learning science, intervention best practices, and smart technology to drive measurable improvement in

student outcomes. Intervene K-12 partnered with researchers from the Center for Research and Reform in Education (CREE) at Johns Hopkins University (JHU) to study the impact of the online tutoring model, which was designed to provide tutoring twice per week in 35-minute sessions.

Unlike the study of Once described above, students were not randomly assigned to receive Intervene K-12 tutoring. Instead, the researchers relied on a propensity-score matching design (or what some call a quasi-experimental design, or QED) to test the impact of tutoring on student math achievement. Specifically, JHU researchers constructed a matched comparison group of students who were most similar to students who received Intervene K-12 tutoring based on observable characteristics including gender, race, economic disadvantage, and special education status. This led to a study sample of 34 treated students and 53 matched comparison students, of which the matched comparison group was statistically equivalent to the treated students on all characteristics except for gender and economic disadvantage - the treatment group contained 65 percent female students and 62 percent economically disadvantaged students; the matched comparison group contained 43 percent female students and 81 percent economically disadvantaged students.

The study compared select math outcomes across treatment and matched comparison students - including end-of-course (EOC) and cumulative final Algebra I exam scores and spring 2023 PSAT mathematics scores.Tutored students outgained matched comparison students on the EOC Algebra I exam by 7 points (0.39 SD) and on the spring 2023 PSAT math assessment by 5 points (0.07 SD); yet, matched comparison students outperformed tutored students on the final Algebra assessment by nearly 6 points (0.27 SD). Notably, none of these differences in outcomes were statistically significant due to small sample sizes. And since students were not randomly assigned to tutoring, these results provide initial - though, far from definitive - evidence on the impact of Intervene K-12's online tutoring for early high school students. In terms of program dosage, tutored students received, on average, approximately 10 hours of Intervene K-12 tutoring, yet with considerable variation in dosage - some students received 18

STUDY SNAPSHOT | Intervene K-12

PUBLISHED: December 2023
RESEARCH TEAM: Center for Research and Reform in Education (CREE) at Johns Hopkins University
STUDY PERIOD: 2023-23 school year
RESEARCH METHOD: Propensity Score Matching Design

STUDENT GROUP(S) STUDIED: 87 Spanish-speaking English Language Learners in ninth grade who had interruptions in their education and were not on track to meet Algebra I standards (35 received tutoring; 53 non-tutored comparison students)

STUDY QUESTIONS:
• What is the impact of the Intervene K-12 tutoring model (35-minute math sessions twice a week, using advancements in smart technology, learning science, and intervention best practices) on 9th grade math achievement?

KEY FINDINGS:
• Students who received tutoring outgained the control group on the End-of-Course (EOC) Algebra I exam and on the spring 2023 PSAT.
• Students in the control group who received no tutoring outperformed students on the final Algebra assessment.
• On average, students received approximately 10 hours of Intervene K-12 tutoring, although dosages varied greatly (some students received 18 hours, some students received less than two hours).

KEY TAKEAWAYS:
• Due to the small sample size, none of these findings are statistically significant. More evidence is needed to understand whether the model impacts student math achievement.

hours of tutoring, and other students received less than two hours of tutoring. Thus, more evidence is needed to understand whether this online tutoring model impacts student math achievement, and the role of dosage in explaining any differences in outcomes among students receiving tutoring.

Looking Ahead: Accelerate's Research Agenda

Despite the extent of existing research on the impact of tutoring, much remains unknown. There is a need for more research into the specific tutoring programs that improve student learning, for which students, in what educational contexts, at what costs, and the programmatic features most associated with student achievement gains. In particular, based on our review of the existing landscape of evidence on tutoring program impact, researchers and policymakers alike would benefit from more studies that address the following:

Understudied Student Groups

- Focus on programs that serve students older than grade 2, especially literacy programs for students in grades 6-12. It is vital to ensure that tutoring programs are available (and accessible) to support students in the upper grades and that these programs are subjected to the same rigorous expectations (and evaluation) as those that provide tutoring to younger students.

- Focus on results for key student groups, such as special education students with individualized education plans (IEPs) and multilingual learners. Serving students who may need tutoring services the most means making sure tutoring programs work for those particular students.

Study Design

- Meet What Works Clearinghouse and ESSA Tier 1 evidence standards with a minimum of 350 study participants. More studies with larger sample sizes will provide a more complete picture of the impact of tutoring when done at scale, and greater opportunity to disaggregate impacts by student characteristics to explore potentially differential impacts for different groups of students. Larger study samples will also provide more precise estimates of tutoring impacts and reduce uncertainty around the expected impact of tutoring on student learning.

- Incorporate multiple treatments (e.g., 3-armed RCTs with large samples) that enable a direct test of program design features such as tutor type, session length, program length, tutor-student ratio, dosage, and tutor training intensity. Directly testing program features, while also continuing to assess the average impact of tutoring programs more generally, will help the field understand which program design features are most essential to improving student learning.

Study Outcomes

- Focus on policy relevant outcomes such as reading comprehension and/or end-of-year state exams. Demonstrating the impact of tutoring on proximal outcomes provides insight into how specific features of tutoring curriculum might translate into changes in discrete skills. Yet, the extent to which improvements in discrete skills translates to more general knowledge and more policy-relevant outcomes that impact students' life outcomes is an area requiring significantly more research attention.

- Include longer term outcomes to assess the extent to which the impact of tutoring persists (or fades away). A better understanding of how tutoring impacts the long term learning trajectory of students will clarify both the long term cost-effectiveness of tutoring and whether new improvement strategies are needed (such as booster doses of tutoring).

- Link improvements in foundational skills to performance on state-mandated end-of-year exams (e.g., grade 3 reading assessments). Given the renewed focus on developing early elementary students' foundational literacy skills, it is becoming increasingly important to understand whether (and to what extent) improvements in foundational skills translate into broader measures of student performance.

Data Collection and Replication

- Collect more precise, valid, and itemized data on program costs (both to the school/district and to society at large). Understanding program cost and being able to compare costs across different tutoring programs and with other education interventions is a core component of understanding the cost-effectiveness of tutoring and, ultimately, the ability to scale effective tutoring programs and practices.

Summary of Accelerate's Research Agenda	
Understudied Student Groups	• **Grade 3+** | Focus on programs that serve students older than 2nd grade, especially literacy programs for students in grades 6-12 • **Specific student populations** | Including students with IEPs, multilingual students
Study Design	• **Rigorous evidence** | Meet What Works Clearinghouse and ESSA Tier 1 evidence standards (>350 students) • **Multiple treatments** | Enable direct test of program design features
Study Outcomes	• **Policy-relevant outcomes** | Including reading comprehension and/or EOY state exams • **Longer-term impact** | Assess the extent to which impact of tutoring persists • **Longitudinal approach** | Link improvements in foundational skills to performance on EOY state exams
Data Collection and Replication	• **Cost data** | Collect more precise and itemized data on program costs (both to the school/district and to society at large) • **Replication trials** | Enables insight into the generalizability and reliability of a program's impact across different schooling settings

- Engage in ongoing replication of impact studies of programs that have already completed RCTs. Understanding the generalizability and reliability of a program, as well as its range of potential outcomes, will improve the use and usefulness of tutoring metrics (e.g., tutoring efficiency and cost effectiveness) and will support more informed and nuanced decision-making. Ongoing replication also prevents wholesale judgment of a program based on a single program impact estimate derived from just one experimental evaluation in a particular research setting.

We welcome readers to share with Accelerate research studies that examine the design, implementation, and/or impact of tutoring programs and personalized learning initiatives. Please contact Matthew Steinberg, Accelerate's Managing Director of Research and Evaluation, with any research studies you wish to share for potential inclusion in a future issue of the Quarterly Research Note.

Accelerate
The National Collaborative for Accelerated Learning