THE UNIVERSITY OF CHICAGO
EDUCATION LAB
Urban Labs

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

Education Insights, Real-World Impact

MARCH 2024

Realizing the Promise of High Dosage Tutoring at Scale

Preliminary Evidence for the Field

Personalized Learning Initiative Research Team[1]

Contents

Click to go directly to each page.

Photo courtesy of Jean Lachat, Saga Education.

ONE

Introduction

The global pandemic was a once-a-century **public health crisis** that left us with a once-a-century **public education crisis**.

Thanks to the federal government's Operation Warp Speed, the public health crisis is now mostly behind us (Schulkin, 2021). But despite the $189.5 billion that the federal government sent to schools through the Elementary and Secondary School Emergency Relief (ESSER) Fund, the public education crisis remains, and without sustained intensive intervention, it may worsen.[2]

Why haven't we made more progress in overcoming pandemic-induced learning loss? What lessons does this suggest for how we might make more progress moving forward?

In this paper, we report on initial lessons learned from our work on the Personalized Learning Initiative (PLI), partnering with four large education agencies around the country over the past several years. The goal of the PLI, an initiative led by the University of Chicago Education Lab in collaboration with MDRC and researchers from the University of Toronto and Stanford University, is to understand whether and how we can scale the benefits of high dosage tutoring such that more students might benefit.

We present findings here from the 2022-23 academic year with four education agencies around the country that we partnered with or tried to partner with: Chicago Public Schools (CPS), Illinois; Fulton County Schools, Georgia; the New Mexico Public Education Department; and a mid-sized urban school district in California. The lessons are of interest partly because each agency tried to follow the recommendation of US Secretary of Education Miguel Cardona about how to deploy their ESSER dollars to overcome pandemic learning loss by using the strategy that's long been recognized as the best way to teach anyone anything: high dosage tutoring (Nickow et al., 2024).

Secretary Cardona steered districts and states towards tutoring because it is one of the few educational interventions shown to be capable of generating learning gains large enough to overcome the amount of learning lost during the pandemic. A pre-pandemic study by our University of Chicago-based research team, in partnership with CPS and Saga Education, looked at tutoring that consisted of (1) delivery during the school day to ensure students participate; (2) a high number of scheduled contact hours where students were scheduled into a credit-bearing course for 50 minutes per day, every day, throughout the entirety of the school year; (3) small student-tutor ratios (2:1); and (4) delivery by trained recent college graduates or mid-career switchers willing to work in exchange for a public-service stipend to help hold costs down and make such an intensive intervention cost-feasible.

LEARN MORE

To learn more, contact **Sadie Stockdale Jefferson, PhD**, Executive Director of the University of Chicago Education Lab (ssjefferson@uchicago.edu).

Photo courtesy of Saga Education.

The data show this type of tutoring can double or even triple what students learn in a year (Guryan et al., 2023).

Whether this type of tutoring can be scaled in the aftermath of the pandemic is an open question. Many schools have struggled to deal with the logistics of re-opening and have experienced declining enrollment, chronic absenteeism, and worsening mental health crises among young people. Exacerbating matters is the labor shortage felt by industries across the economy, including schools. Education agencies attempted to scale up tutoring in this environment more quickly and at a larger scale than had previously been attempted. Districts often were adapting tutoring program parameters to fit their local context but without any guidance about which program design features are essential (versus unnecessary) to promote student learning.

We found that the two sites we tried to partner with that focused on out-of-school tutoring – New Mexico and a mid-sized California district – failed to get high levels of student participation. The California district tried to integrate tutoring into its after-school program, which faltered due to low student participation in the after-school program itself. New Mexico tried to get students to do virtual tutoring at home, but recruitment was difficult. New Mexico has subsequently pivoted to incorporating tutoring into the school day and we will be assessing the efficacy of that model in the coming year.

The results are more encouraging for the two districts that used ESSER funding to incorporate tutoring during the school day from the start – Chicago Public Schools and Fulton County Schools:

- **Results from the 2022-23 academic year** suggest that tutoring can be scaled and can work, even in the aftermath of the pandemic. Students who participated in tutoring 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** is still too noisy to know how big the effect is so far.

These findings suggest some answers as to why ESSER funding did not do more to overcome pandemic learning loss.

First, one key to ensuring the successful delivery of high dosage tutoring seems to be incorporating it directly into the school day rather than trying to get students to do tutoring after school or at home.

That's a lesson that many districts and states took some time to learn. (Other shared features of seemingly successful tutoring programs is the use of a trained and supported tutor and a structured curriculum aligned with core instruction at the school rather than just being used as 'homework help').

Second, even in the districts that did incorporate tutoring into the school day, the funding available from ESSER to help students is not nearly enough to serve all the students affected by the pandemic.

The ESSER funding was only enough to increase K-12 spending by 6% per year (Guryan & Ludwig, 2023). Chicago and Fulton County have helped thousands of students, but there are tens of thousands more who could benefit from tutoring.

The lesson is that ESSER funding can and has generated positive gains for students.

Some adjustments to current policies and practices could help expand that impact, including getting districts and states to focus more on the types of tutoring programs that are most helpful to students and expanding funding to serve more of the students who have been harmed by the pandemic. Without adequately remediating pandemic learning loss among the 50 million American children who were of school age during the pandemic, the result may be lifelong scars in terms of long-term outcomes like earnings, with aggregate losses that Kane et al. (2022) estimate could be as large as $900 billion.

TWO

Classroom teaching and the pandemic[3]

High rates of absenteeism and the imperfect substitution of remote schooling for in-person instruction have contributed to large learning losses, exacerbating longstanding racial and socioeconomic disparities in educational achievement. **But the challenge is not merely one of short-term learning losses.**

Because education is intrinsically cumulative, there is the real possibility that pandemic-induced school disruptions may set a whole generation of students off track for the rest of their lives.

To see the problem, reflect on your own schooling experiences. Schools are organized into grades from K through 12. Within those grades, students are usually taught in classes of, say, 20 to 35 (depending on the district, school, grade, subject, etc.), usually by a single teacher. Teachers are encouraged to teach students grade-level content: that is, what their students are tested on and what the teachers are evaluated on. For perhaps well-intentioned reasons of not wanting some students to be stigmatized or give up on school altogether, the vast majority of students get promoted to the next grade, whether or not they have mastered grade-level skills.

The frequent result: a teacher standing in front of a classroom, trying to teach grade-level content to students whose academic levels vary enormously.

Even before the pandemic, the average fifth-grade class, for instance, contained some students working at a third-grade level and some working at an eighth-grade level.[4]

For decades, dealing with this variation in student needs has been regularly reported by teachers as one of the hardest parts of teaching (Guryan et al., 2023).

Or, as well-known education scholar Steve Raudenbush frequently notes, “Dealing with heterogeneity is the central problem of education.”[5] Some indications are that teachers wind up targeting instruction toward something like the 60th percentile of the distribution (Bloom, 1984). That means students who are far from grade level, particularly far below grade level, may benefit less from grade-level instruction – so-called “academic mismatch.”

Put differently, the students who are behind (and need the benefits of schooling the most) may benefit the least from classroom instruction.[6]

That wide range of instructional needs within each classroom has only gotten wider since the pandemic began, particularly because the learning impact of the pandemic fell disproportionately on the most disadvantaged students (Lewis et al., 2022). So, the problem of academic mismatch has grown, while the benefits of regular classroom instruction for these students over the rest of their schooling careers may be attenuated.

This problem takes on urgency because of some indications that there are clear developmental milestones in school that may be particularly important. For example, students who can’t read at grade level by third grade are four times less likely to graduate high school.[7] Students who haven’t passed Algebra I by the end of 9th grade are five times less likely to graduate.[8]

It would seem that something needs to be done.

THREE

The promise of tutoring

There is one educational intervention capable of accelerating learning enough to overcome pandemic learning loss, an intervention that's been around since at least the fifteenth century at Oxford University: **high dosage tutoring.**

This intervention involves one instructor working with one or two students at a time for several hours per week. (One could think of this practice as extreme class-size reduction.) High dosage tutoring helps address what teachers report in surveys to be the two most difficult challenges of classroom teaching: variability in students' academic levels (and hence their needs); and, perhaps relatedly, classroom management.[9]

Modern social science has confirmed the wisdom of the Oxford dons hundreds of years ago. A series of demonstration projects in the 1980s found that compared to regular classroom instruction, students tutored one-to-one spent almost 40 percent more time on-task. Students in tutoring learned fully two standard deviations (SDs) more than their peers in traditional classroom settings (Bloom, 1984). That's larger than the test score gap between rich and poor (Reardon, 2011; Loveless, 2012) and would be enough to move an average student to the 95th percentile.

We also see large gains from tutoring outside of controlled lab conditions in real-world school settings. A review of more than 90 randomized controlled trials (RCTs) of smaller-scale tutoring programs showed an average effect of 0.37 SDs (Nickow et al., 2020).[10]

Bloom (1984) noted that the enormous impacts of tutoring suggest the challenge of education is not a pedagogical one, but rather an economic one: **How do we scale tutoring and hold down its cost and labor requirements to make it feasible nationally? How do we deliver Oxford-style tutoring at American public school prices?**

One breakthrough insight by the non-profit Saga Education is that tutoring is a sufficiently different task from regular classroom instruction. Unlike teaching, tutors do not need extensive pedagogical training to differentiate instruction across thirty students or experience in classroom management. Instead, with only one or two students and a structured curriculum, tutors can be hired to do a much narrower task and set of responsibilities, which means the staffing requirements have fewer barriers to entry, thereby expanding the pool of possible individuals who can be tutors beyond certified teachers. Put differently, we can scale up individualized instruction without doing so on the backs of already overburdened and hard-working, professionally trained school teachers.

High Dosage Tutoring

A series of demonstration projects in the 1980s found that compared to regular classroom instruction:

Students tutored one-to-one spent almost **40 percent more time on-task**.

Students in tutoring learned fully **two standard deviations (SDs) more than their peers** in traditional classroom settings (Bloom, 1984).

Photo courtesy of Saga Education.

Photo courtesy of Saga Education.

Saga Education's human resources model sought recent college graduates or mid-career switchers who are willing to tutor for one year as a public service for a modest living stipend ($20,000 to $30,000). This model lowered costs and made it more feasible to deliver intensive, personalized instruction at such high dosages in a way that is less cost prohibitive.

Our first RCT of Saga tutoring with CPS involved 2,633 ninth- and tenth-grade students in low-performing schools in economically under-resourced areas on the south and west sides of Chicago. We found that two-on-one tutoring for 45-50 minutes a day in school every day increased math test scores by 0.16 SDs and reduced math-course failures by 49 percent (Guryan et al., 2023). A replication RCT in the 2014–2015 academic year with 2,710 ninth and tenth graders found even larger impacts, with test score gains of 0.37 SDs and grade impacts comparable to the first study. When the studies were pooled together, the overall effect on math test scores was 0.28SDs and seems to largely persist – equal to 0.23 SD in 11th grade (Guryan et al., 2023).

A separate analysis compared potential learning loss policy solutions, including high dosage tutoring (Kraft & Falken, 2021).[11] Overall, the learning gains from high dosage tutoring are much closer to offsetting the average learning loss experienced during the pandemic than other potential policy measures are. High dosage tutoring is plausibly the intervention most up to the task of meeting the scale of our current learning loss challenge. As one education expert put it, tutoring sessions are "the best learning conditions we can devise" (Bloom, 1984, p. 4).[12]

Summary of Results from Education Lab Randomized Controlled Trials (RCTs) of Saga Tutoring

First Education Lab RCT of Saga tutoring with CPS

Two-on-one tutoring for 45-50 minutes a day in school every day:

increased math test scores by **0.16 SDs**

reduced math-course failures by **49%**

(Guryan et al., 2023)

2,633 ninth- and tenth-grade students in low-performing schools in economically under-resourced areas on the south and west sides of Chicago.

Replication RCT in the 2014–2015 academic year

Grade impacts were comparable to the first study.

increased math test scores by **0.37 SDs**

2,710 ninth and tenth graders

Pooled RCT studies

overall effect on math test scores is **0.28 SDs**

equal to **0.23 SD** in 11th grade

(Guryan et al., 2023)

FOUR

Tutoring in a Post-Pandemic World

The evidence shows that high dosage tutoring *can* generate remarkably large gains in student learning, large enough to overcome pandemic learning loss for most students with just a year or two of intervention.

However, success is not automatic; there seem to be ways to deliver tutoring that increase versus decrease the value to students.

While we recognize our findings come from just four site partners, these four data points nonetheless paint a fairly suggestive picture: The two that tried to deliver tutoring outside of school failed to effectively deliver tutoring at all to any meaningful number of students. In contrast, the two districts that incorporated tutoring into the school day show signs in these initial data of learning gains in math; these preliminary data are too noisy to know how large the effect is at this point for reading or for reading and math scores pooled together.

OUT-OF-SCHOOL TUTORING

Chicago is a large metropolitan area with lots of amenities that appeal to recent college graduates. When Saga tried to hire tutors for our initial tutoring RCTs with CPS, there was no shortage of excellent applicants.

In contrast, the New Mexico Public Education Department (NMPED) had a very different read on their own local labor market conditions. Besides the general labor shortages seen on the heels of the pandemic, NMPED made the judgment that recruiting lots of tutors to work in person in a large, rural state would be difficult, so it decided to incorporate virtual rather than in-person tutoring. The decision to adopt virtual tutoring made it possible for virtual tutors to meet with students after school, at night, and over the weekend to avoid having to give up some time during the school day for tutoring.

In the end, 1.5 percent of students–or 527 students–signed up for the evening and weekend sessions out of an estimated 34,262 eligible students statewide after considerable recruitment efforts on behalf of the state and study team (McCormick et al., 2023). Statewide coverage was not equitably distributed: rural students were less likely to sign up than students in urban or suburban settings. Of the approximately 500 students who signed up, only 326, or 62 percent, actually participated in tutoring. The limited scope and scale of this impact led NMPED to use these data to redesign their high dosage tutoring model. In this current year, SY23-24, NMPED has deployed a virtual high dosage tutoring model that takes place in a monitored classroom during the school day and is able to reach hundreds of students much more consistently.

A second district we partnered with, a large urban district in California, had what it thought was a robust after-school program that it believed was serving thousands of students already. This program seemed to create a natural delivery vehicle for high dosage tutoring that could avoid operational disruptions from carving out time for tutoring during the school day. However, attendance was much lower than anticipated in this after-school program, perhaps in part because it was launched after the school year was already underway, and perhaps after many parents had already found alternate solutions to aftercare.

Photo courtesy of Saga Education.

IN-SCHOOL TUTORING

In contrast to the failed attempts to deliver high dosage tutoring out of school, two partner districts–Chicago Public Schools and Fulton County Schools–delivered tutoring during the school day.[13] We worked with them to randomize a total of over 2,000 students to either high dosage tutoring or a business-as-usual control during the 2022-23 academic year. The tutoring was successful in the most basic sense in that thousands of students received tutoring. Our interim analysis of student performance also seems to show that this tutoring increased learning, but the ranges of the potential impacts (the confidence intervals) are larger than one might like. With those caveats in mind, there are signs of large gains in math, with imprecise estimates for reading and large (but somewhat imprecise) gains in test scores for reading and math pooled together.

In Fulton County, our research team worked with the district to stand up tutoring in math or reading in grades 3-8 and math for grade nine. Across 17 study schools, we randomly assigned 736 students to high dosage tutoring and 770 to "business as usual." Four vendors implemented tutoring during the school day: New Generation and Applerouth (both in-person and virtual tutoring), The Tutor Shop (in-person tutoring only), and FEV (virtual tutoring only). Vendor-provided tutoring started in mid-October at a few schools, with most of the schools starting tutoring in January 2023. Thus, note our pooled impact estimates are not for a full year of tutoring.

TABLE I:
Balance Table for Analysis Sample, Fulton County, SY22-23

COVARIATE	ASSIGNED TO TREATMENT N = 556	ASSIGNED TO CONTROL N = 607	*P*-VALUE	N
Age	11.52	11.67	0.777	1163
% Male	49.3%	48.9%	0.799	1163
% Black	91.6%	91.9%	0.63	1163
% Hispanic	9.9%	10.2%	0.234	1163
% Receiving Free/Reduced Lunch	84.7%	77.1%	0.772	1163
% English as a Second Language	2.5%	2.5%	0.74	1163
% Diverse Learner	2.3%	3.3%	0.93	1163
% of Days Attended (Prior Year)	1.3%	0.8%	0.491	1163
Core GPA	0.93	0.94	0.862	972
i-Ready Reading Score (BOY SY23)	3.21	3.22	0.417	900
i-Ready Math Score (BOY SY23)	548.76	552.66	0.5	1162
Milestones ELA Score (EOY SY22)	454.54	457.51	0.17	1162
Milestones Math Score (EOY SY22)	502.19	508.36	0.324	614
F-Test - Baseline Cov.	N/A	N/A	0.605	637

Note: Reported p-values test the difference in means for the treatment and control groups. To conduct the t-test, no imputation was carried out and the number of observations vary reflecting availability of the variable. In the final row, we test the joint hypothesis of overall differences in baseline characteristics between the treatment and the control group. To test the joint hypothesis, we regress a treatment indicator on baseline covariates (missing values imputed with average values within their school, year, and grade among the study sample), corresponding missingness indicators, and randomization block fixed effects and calculate the resulting F-statistic from this regression. To avoid distributional assumptions, we then randomly re-assign the treatment indicator within randomization blocks (fixing randomization rate within the block) and estimate the corresponding F-statistic and associated p-value from each placebo draw. We repeat this process 10,000 times. In the distribution of 10,000 placebo treatments, we see where the originally calculated F-statistic lies, and report the rank, i.e., the original F-statistic has a rank of ~6,050 among placebos. We use EOY SY21-22 test scores at baseline.

Of the 1,506 randomized students in Fulton County, we currently have test scores for 1,163 students who comprise our current analysis sample. Students are predominantly Black (92%) with some Hispanic students (10%) represented. Approximately 81% of students receive free or reduced-price lunch, and most students have strong attendance and an average core-subject GPA of 3.21.

The treatment and control groups are statistically balanced as indicated by an overall F-test on all the baseline covariate characteristics. However, adjusting for randomization block fixed effects, students assigned to treatment exhibited slightly higher prior-year math scores on the Milestones assessment. In Fulton County, 82% of those who were assigned to the high dosage tutoring group participated in tutoring for at least one session.

Among the subset of students for whom we have dosage data (N=541), conditional on participating in at least one session, treatment students received 11.08 sessions over the course of the year.[14] There is also a large amount of control crossover in the business-as-usual group of students - 35% of those who were assigned to the control group also participated in tutoring for at least one session. The difference in dosage between program and business-as-usual groups on average in Fulton County is approximately 5.52 sessions per student. Note that the relationship between tutoring sessions attended and student learning gains is difficult to determine with high confidence given the issues noted with the quality of the attendance data during the 2022-23 school year.

TABLE II:
Take-up and Dosage, Fulton County, SY22-23, by Treatment Status

TREATMENT STATUS	HIGH DOSAGE TUTORING	CONTROL
N	541	570
Take-up Rate	81.5% (441)	35.1% (200)
Mean Dosage (# Sessions, Conditional on Take-up)	11.08	9.91
Mean Dosage (# Sessions, All Assigned Students)	9.0	3.48

Note: Reported N reflects the number of students for whom both participation and outcome data are available. Take-up rate reflects what percent of each randomization arm participated in at least one tutoring session in the subject to which they were assigned.

Photo courtesy of Jean Lachat, Saga Education.

In Chicago, our research team randomized 548 students in grades K-11 across 13 CPS study schools into either high dosage tutoring or business as usual. All students were offered tutoring at least three times per week for 30-minute sessions via the CPS Tutor Corps—an initiative where the CPS central office hired and managed tutors who were trained to deliver tutoring in reading by Amplify and in math by Saga Education. Of the 548 CPS students who were randomized to tutoring, we were able to access primary outcomes for 429 students.

As in Fulton County, the F-test that tests whether the covariates of the two groups are jointly equal to each other indicates that the two samples are balanced overall. However, as can be expected, some of the individual covariates differed. For example, students randomly assigned to the treatment (high dosage tutoring) group were more likely to be eligible for free or reduced-price lunches at baseline, less likely to be classified as diverse learners (students with Individualized Education Programs), and had a slightly lower GPA than their counterparts who were assigned to the business-as-usual group.

TABLE III:

Balance Table for Analysis Sample, Chicago, SY22-23

COVARIATE	ASSIGNED TO TREATMENT N = 243	ASSIGNED TO CONTROL N = 186	*P*-VALUE	N
Age	10.05	9.87	0.326	427
% Male	45.5%	53.5%	0.537	427
% Black	52.3%	55.9%	0.662	427
% Hispanic	44.8%	42.5%	0.871	427
% Homeless	13.9%	7.6%	0.212	403
% Receiving Free/Reduced Lunch	94.0%	86.4%	0.047**	411
% English as a Second Language	22.1%	22.1%	0.254	403
% Diverse Learner	3.9%	7.0%	0.073*	403
Number of Disciplinary Incidents	0.1	0.08	0.752	397
Overall GPA	3.24	3.32	0.058*	385
Standardized Math IAR Test Score	-0.23	-0.04	0.344	221
Standardized Math STAR Test Score	-0.24	0	0.363	220
Standardized Reading DIBELS Test Score	0.06	-0.01	0.539	109
Standardized Reading IAR Test Score	0.03	-0.02	0.663	229
Standardized Reading STAR Test Score	0.13	0.01	0.574	221
RI F-Test - Baseline Cov.			0.933	1163

Note: Reported p-values test the difference in means for the treatment and control groups. To conduct the t-test, no imputation was carried out and the number of observations vary reflecting availability of the variable. In the final row, we test the joint hypothesis of overall differences in baseline characteristics between the treatment and the control group. To test the joint hypothesis, we regress a treatment indicator on baseline covariates (missing values imputed with average values within their school, year, and grade among the study sample), corresponding missingness indicators, and randomization block fixed effects and calculate the resulting F-statistic from this regression.

Among the 429 students in our Chicago analysis group, 67% of those were assigned to the high dosage tutoring group. Conditional on participating in at least one session, they received 27.3 sessions over the year.[15] We see some control crossover, though, as well; among the business-as-usual group, 18% participated in at least one tutoring session. The difference in dosage between the treatment and business-as-usual groups in Chicago is 13.8 sessions on average per student.

The difference in dosage between the treatment and business-as-usual groups in Chicago is **13.8 sessions on average per student**.

TABLE IV:

Take-up and Dosage, Chicago, SY22-23, by Treatment Status

TREATMENT STATUS	HIGH DOSAGE TUTORING	CONTROL	TOTAL
N	243	186	429
Take-up Rate	67.5% (164)	17.7% (33)	45.9%
Mean Dosage (# Sessions, Conditional on Take-up)	27.26	26	27.05
Mean Dosage (# Sessions, All Assigned Students)	18.40	4.61	12.42

Note: *This table only includes students for whom outcome data is available. Take-up rate reflects what percent of each randomization arm participated in at least one tutoring session in the subject to which they were assigned.*

With baseline balance and no statistically significant differential attrition, we determine the intention-to-treat (ITT) effect of high dosage tutoring by regressing student test scores against an indicator for assignment to tutoring (controlling for randomization fixed effects and other baseline covariates to improve statistical precision).[16] We also used the random assignment status variable as an instrumental variable to estimate the effect of getting at least one session of tutoring on those who are tutored.[17]

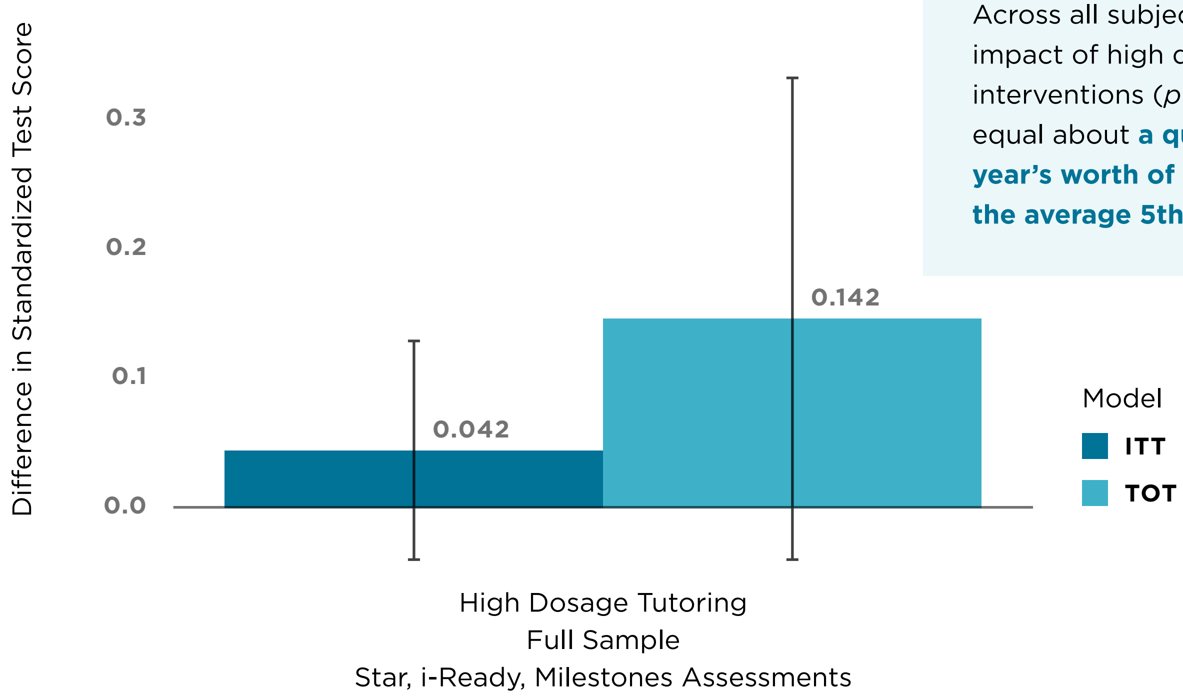
Among our pooled sample, 76% of students assigned to high dosage tutoring participated in at least one session; of those students that participated in at least one session, they received an average of 17 high dosage tutoring sessions.[18] We calculate pooled estimates using standard methodology for a meta-analysis.[19]

Overall, the effects for high dosage tutoring interventions pooled across both sites and both subject areas (reading and math together) is 0.04 SD for the intent-to-treat estimates (p = 0.31) and 0.14 SD for the effects of receiving tutoring (the LATE) treatment-on-the-treated estimates (see **FIGURE I**).

Photo courtesy of DGL Images, Adobe.

FIGURE I:

Pooled Effects of High Dosage Tutoring on Student Learning in Chicago and Fulton County, SY22-23



Across all subjects and sites, the impact of high dosage tutoring interventions (p=0.14) would equal about **a quarter of a year's worth of learning for the average 5th grader**.

Photo courtesy of Kenny Eliason, Unsplash.

How to think about these estimates depends on the purpose for which they are intended. For research purposes, the usual focus is on statistical significance; that is, whether the 95% confidence intervals around the estimates contain zero versus can rule out a null impact. In this case, for both subjects pooled together, the confidence interval suggests the effect could be as large as over 0.3 SD but at the same time could be zero, or even slightly negative. Science is appropriately cautious about overturning a null hypothesis based on a single noisy estimate.

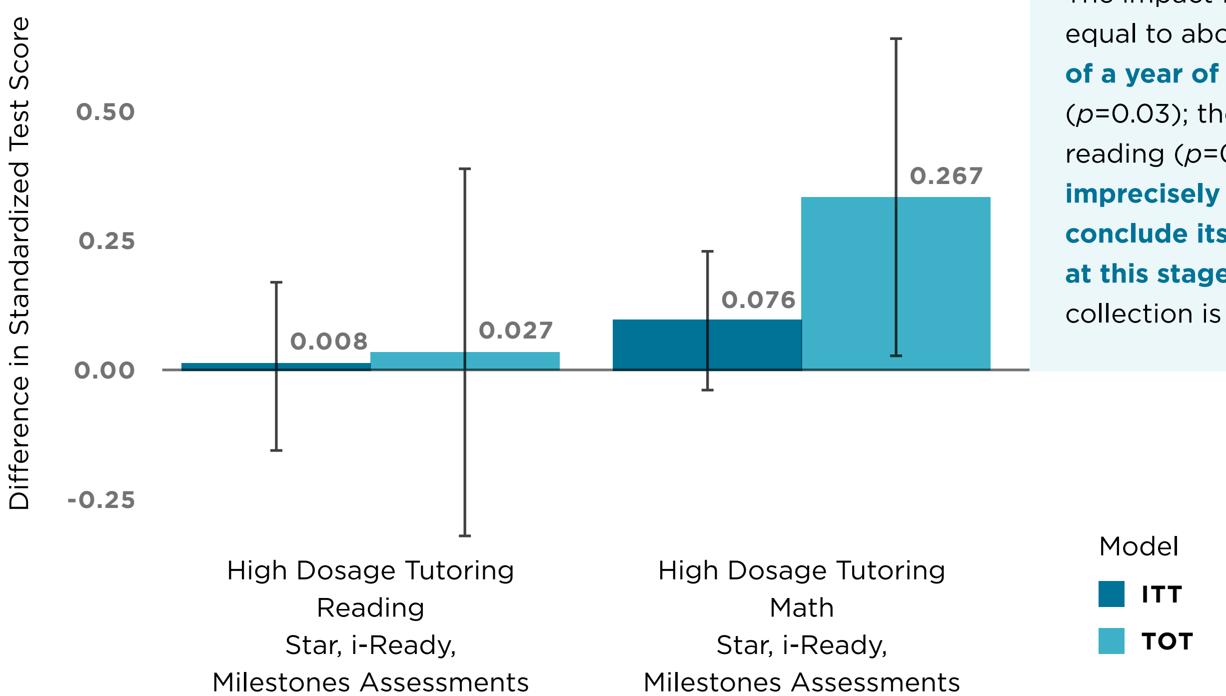
But for policy purposes, it is important to realize that an estimate that is not quite statistically significant is not the same as a 'true zero' (see, for example, Ziliak & McCloskey, 2008; Manski, 2019; Imbens, 2021). Most of the impact values captured by the 95% confidence intervals are above zero, while the benefit-cost analysis in Guryan et al. (2023) suggests that impacts of this magnitude (0.14 SD is close to what was reported from RCT 1 in Guryan et al.) generate somewhere between $2 and $4 in benefits to each dollar expended. From a policymaker's perspective, these results imply that, more likely than not, this ESSER-funded tutoring generated meaningful and material benefits to students and society as a whole.

From a policymaker's perspective, these results imply that, more likely than not, this **ESSER-funded tutoring generated meaningful and material benefits to students and society as a whole**.

The results for reading alone are smaller and quite imprecisely estimated; the 95% confidence interval suggests the results could range from substantially negative (-0.25 SD) to substantially positive (+0.25 SD) (see **FIGURE II**).

FIGURE II:

Pooled Effects of High Dosage Tutoring on Student Learning in Chicago and Fulton County, SY22-23, by Tutoring Subject



The impact for math is equal to about **two-thirds of a year of learning** (p=0.03); the impact for reading (p=0.85) is **too imprecisely estimated to conclude its magnitude at this stage**–but data collection is ongoing.

However, in math, the ITT effect is 0.08 SD (p=0.15) and the effect of participating in tutoring (the LATE) is 0.27 SD (p=0.03). Despite the less-than-ideal delivery setting of the post-pandemic period and Fulton County's midyear start, the estimated TOT effects are remarkably similar to what we saw with our findings from our two pre-pandemic RCTs of Saga-delivered tutoring with more than 2,600 CPS students (TOT = 0.28 SD) (see Guryan et al., 2023). Yet, CPS was no longer using tutors hired directly by Saga but instead in-housed the tutoring provision. Fulton County was also not using Saga but instead a variety of different vendors. And, the 'dosage' delivered in Chicago and Fulton County in the 2022-23 academic year is lower on average than what Saga delivered in Chicago many years ago. Our results provide hints that high dosage tutoring can be scaled successfully.

The site-specific results mirror our pooled results. Despite the differences in the tutoring program design and delivery across sites and recognizing that the site-specific estimates are even less precisely estimated than the pooled results, the pattern of results seems qualitatively similar in Fulton County and Chicago.

Our results provide hints that **high dosage tutoring can be scaled successfully**.

FIVE

Conclusion

While these results represent only a portion of our expected study sample and are subject to change, **we think it is important to share them now**. This urgency stems from the fact that school districts and states nationwide are in the midst of strategizing for the 2024-2025 school year, coinciding with the expiration of federal relief funds in September 2024.

The key lesson is tutoring can work; ESSER dollars have translated into student learning gains, particularly in districts that deployed their funding to deliver tutoring that was (1) delivered during the school day at a consistent, scheduled time; (2) delivered by trained and supported tutors; and (3) makes use of a structured curriculum that is aligned with the school's instruction.

Absent a long list of alternative approaches for remediating pandemic learning loss, high dosage tutoring remains one of the more promising strategies for solving this generational challenge.

Put differently, we can scale up individualized instruction without doing so on the backs of already overburdened and hard-working, professionally trained school teachers.

Photo courtesy of Saga Education.

The Personalized Learning Initiative receives substantial support from Accelerate – The National Collaborative for Accelerated Learning, which is supported by Citadel founder and CEO Kenneth C. Griffin and Griffin Catalyst; Arnold Ventures; the Bill & Melinda Gates Foundation; the Overdeck Family Foundation; and the Walton Family Foundation. Additional funding comes from the AbbVie Foundation, Crown Family Philanthropies, Citadel founder and CEO Kenneth C. Griffin and Griffin Catalyst, IMC Chicago Charitable Foundation, and Vivo Foundation.

Endnotes

1. 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

2. While new research suggests these efforts bolstered learning in math and reading beyond traditional gains, it is clear that the magnitude of these gains pale in comparison to what is needed.

3. The next two sections draw on Guryan and Ludwig (2023).

4. The majority of students are one to two years behind (Peters et al., 2017).

5. Private communication, Jens Ludwig with Steve Raudenbush.

6. While the measurement issues are subtle, there is some indication that the variance of student learning increases as children progress through school, which, if true, would be consistent with the idea that students who are behind benefit less from classroom instruction (Cascio & Staiger, 2012; Nielsen, 2023).

7. 16 percent of students who are not at grade-level reading proficiency in third grade do not go on to graduate high school, compared to only 4 percent of students who are proficient (Hernandez, 2011).

8. 80 percent of students who do not pass algebra do not go on to graduate high school, compared to only 15 percent of students who do pass algebra (Schachter, 2013).

9. Even before the pandemic, the average fifth-grade class contained some students working at a third-grade level and some working at an eighth grade-level. The majority of students are one to two years behind (see Peters et al., 2017).

10. The review, which covered tutoring programs ranging in dosage from 1–2 days per week to every day of the week, found that the more time students spent in tutoring, the better. In-school programs were also nearly twice as effective as after-school programs. However, paraprofessional tutoring programs generated effect sizes nearly as large (0.4 standard deviations) as professional teachers (0.5 standard deviations), indicating that who performs the tutoring is not as critical as might have been expected (Nickow et al., 2020).

11. Kraft and Falken (2021) includes an excellent discussion of the measured impact of alternate policies such as class-size reduction, additional school hours, additional school days, and summer school. All these alternatives show relatively lower impact than HIT.

12. As we discuss below, our research to date suggests that tutoring is effective when it's done in schools at a ratio of two students per full-time, dedicated adult tutor; meets daily; and follows a set curriculum. Whether tutoring might be equally effective at higher student ratios, or with part-time or peer tutors, or when face-to-face instruction is supplemented with computer time—these remain open questions.

13. We focus in this paper on the contrast between students who were assigned to high dosage tutoring versus control. There is another treatment condition in development – lower-cost, more sustainable high dosage tutoring that often incorporates technology – that is being piloted in some site partners but not yet at the same scale. Results for that treatment arm will be the focus of future work.

14. To guard against inconsistent data entry in the district-provided tracking system, the PLI research team administered a survey to each school's tutoring coordinator which asked whether a student ever attended at least one tutoring session. This survey complemented the available participation data received from the district. This survey data suggests that our dosage rates underestimate the actual dosage students received as highlighted by the difference in scheduled sessions reported by tutors in the implementation research surveys.

15. Participation and dosage data is available for all CPS students and relies on manual data entry by tutors, which may suffer from some measurement error.

16. In addition to randomization block indicators, we control for gender, race, age, learning disability, English learner status, free/reduced lunch status, homelessness status, grade level, GPA, number of attending days, latest reading and math test scores available and (where available) number of misconduct incidents to account for any remaining differences between treatment and control students.

Endnotes (cont.)

17. Because there is some control cross-over, this should be interpreted as a local average treatment effect (LATE).

18. Seventeen high dosage tutoring sessions is the average across both sites using the weights used to estimate pooled TOT impacts.

19. The research team will calculate pooled estimates using student-level data across study sites and years after we generate and post a pre-analysis plan that will govern how we pool estimates across sites and site-years. This pre-analysis plan will be posted in Spring 2024, prior to receiving end-of-year data for SY23-24 in June. In the estimates presented here, we calculate pooled estimates of high dosage tutoring versus BAU across the two sites by weighting each site according to the precision of their estimate (i.e., the inverse of the estimated variance of the parameters of interest). Hence, with two sites, if the estimate of impact at site 1 has a standard error of 1, it will get a weight of 80% when the other site has a standard error of 2 (the second site will correspondingly get a weight of 20%). This methodology is the standard methodology for meta-analysis.

References

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16. https://www.jstor.org/stable/1175554?origin=JSTOR-pdf

Cascio, E., & Staiger, D. (2012). Knowledge, Tests, and Fadeout in Education Interventions. *National Bureau of Economic Research*, Working Paper 18038. https://doi.org/10.3386/w18038

Fahle, E., Kane, T. J., Reardon, S. F., Staiger, D. (2024).The First Year of Pandemic Recovery: A District-Level Analysis. *Center for Education Policy Research, Harvard University*. https://educationrecoveryscorecard.org/wp-content/uploads/2024/01/ERS-Report-Final-1.31.pdf

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer, R. G., Mayer, S., Pollack, H., Steinberg, L., & Stoddard, G. (2023). Not Too Late: Improving Academic Outcomes among Adolescents. *American Economic Review*, 113 (3), 738-65. https://doi.org/10.1257/aer.20210434

Guryan, J., & Ludwig, J. (2023). Overcoming Pandemic-Induced Learning Loss. *Building a More Resilient US Economy*. https://www.economicstrategygroup.org/wp-content/uploads/2023/11/Guryan_Ludwig_2023_Chapter.pdf

Hernandez, D. (2011). Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation. *Annie E. Casey Foundation*. https://www.aecf.org/resources/double-jeopardy

Imbens, G. W. (2021). Statistical Significance, p-Values, and the Reporting of Uncertainty. *Journal of Economic Perspectives*, 35 (3), 157-74. https://aeaweb.org/articles?id=10.1257/jep.35.3.157

Kane, T. J., Doty, E., Patterson, T., & Staiger, D. O. (2022). What Do Changes in State Test Scores Imply for Later Life Outcomes? *Center for Education Policy Research, Harvard University*. https://educationrecoveryscorecard.org/wp-content/uploads/2022/11/Long-Term-Outcomes_11.18.pdf

Kraft, M., & Falken, G. (2021). A Blueprint for Scaling Tutoring and Mentoring across Public Schools. *AERA Open 7*. https://doi.org/10.1177/23328584211042858

Lewis, K., Kuhfeld, M., Langi M., Peters, S., & Fahle, E. (2022). The Widening Achievement Divide during COVID-19. *Center for School and Student Progress Research Report*. https://files.eric.ed.gov/fulltext/ED627139.pdf

Loveless, T. (2012). How Well Are American Students Learning? *Brown Center Report on American Education*. https://www.brookings.edu/wp-content/uploads/2016/06/0216_brown_education_loveless.pdf

Manski, C. F. (2019). Communicating Uncertainty in Policy Analysis. *Proceedings of the National Academy of Sciences*, 116, 7634–41. https://doi.org/10.1073/pnas.1722389115

McCormick, R., Woo, J., Steiner, B., & Grossman, J. (2023). Tutoring Lessons from New Mexico. *MDRC*. https://www.mdrc.org/work/publications/tutoring-lessons-new-mexico

References (cont.)

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *National Bureau of Economic Research*, Working Paper no. 27476. https://www.nber.org/papers/w27476

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *American Educational Research Journal*, 61(1), 74-107. https://doi.org/10.3102/00028312231208687

Nielsen, E. (2023). The Variance of Achievement Increases During Childhood. *Federal Reserve Board of Governors*. Working paper.

Peters, Scott J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2017). Should Millions of Students Take a Gap Year? Large Numbers of Students Start the School Year above Grade Level. *Gifted Child Quarterly* 61(3), 229–38. https://doi.org/10.1177/0016986217701834

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, 1(1), 91-116.

Schachter, R. (2013). Solving Our Algebra Problem: Getting All Students through Algebra I to Improve Graduation Rates. *District Administration* 49(5), 43–46.

Shulkin, D. (2021). What Health Care Can Learn from Operation Warp Speed. *NEJM Catalyst Innovations in Care Delivery*, 2(1). https://catalyst.nejm.org/doi/full/10.1056/CAT.21.0001

Ziliak, S. T., & McCloskey, D. N. (2008). The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. *University of Michigan Press*. https://doi.org/10.3998/mpub.186351