EdWorkingPaper No. 22-568

Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program

Matthew A. Kraft
Brown University

John A. List
University of Chicago,

Jeffrey A. Livingston
Bentley University

Sally Sadoff
University of California
San Diego

A substantial body of experimental evidence demonstrates that in-person tutoring programs can have large impacts on K-12 student achievement. However, such programs typically are costly and constrained by a limited local supply of tutors. In partnership with CovEducation (CovEd), we conduct a pilot program that has potential to ease both of these concerns. We conduct an experiment where volunteer tutors from all over the country meet 1-on-1 with middle school students online during the school day. We find that the program produces consistently positive (0.07σ for math and 0.04σ for reading) but statistically insignificant effects on student achievement. While these estimates are notably smaller than those found in many higher-dosage in-person tutoring programs, they are from a significantly lower-cost program that was delivered within the challenging context of the COVID-19 pandemic. We provide evidence that is consistent with a dosage model of tutoring where additional hours result in larger effects.

VERSION: April 2022

Suggested citation: Kraft, Matthew A., John A. List, Jeffrey A. Livingston, and Sally Sadoff. (2022). Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program. (EdWorkingPaper: 22-568). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/b1ch-0g29

Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program

Matthew A. Kraft, John A. List, Jeffrey A. Livingston, and Sally Sadoff*

April, 2022

Abstract

A substantial body of experimental evidence demonstrates that in-person tutoring programs can have large impacts on K-12 student achievement. However, such programs typically are costly and constrained by a limited local supply of tutors. In partnership with CovEducation (CovEd), we conduct a pilot program that has potential to ease both of these concerns. We conduct an experiment where volunteer tutors from all over the country meet 1-on-1 with middle school students online during the school day. We find that the program produces consistently positive (0.07σ for math and 0.04σ for reading) but statistically insignificant effects on student achievement. While these estimates are notably smaller than those found in many higher-dosage in-person tutoring programs, they are from a significantly lower-cost program that was delivered within the challenging context of the COVID-19 pandemic. We provide evidence that is consistent with a dosage model of tutoring where additional hours result in larger effects.

*Sadoff: University of California San Diego, 9500 Gilman Drive #0553, San Diego, CA 92107 (email: ssadoff@ucsd.edu). We thank the Chicago Heights school district and CovEducation (CovEd) for their partnership. For their invaluable contribution to project implementation, we thank Tom Amadio, William Seidelmann, Korinne Johnson, Safiyah Melvin, Shannon Vera and Anthony Sanchez in Chicago Heights; and Catherine Eng, Anna Kim, Suraj Joshi, Dheekshita Kumar, Evelyn Wong and Minaz Zad at CovEd. Yana Gallen provided insightful comments. Charlie Brobst, Uditi Karna, Amani Mryan and Max Ross contributed exceptional research assistance. The research was conducted with approval from the University of Chicago IRB.

1 Introduction

A substantial body of experimental evidence now demonstrates the large impacts that in-person tutoring programs can have on K-12 student achievement (Nickow et al., 2020). This evidence has motivated interest in scaling tutoring across public schools to address COVID-19 learning disruptions and expand equitable access to individualized instruction (Kraft and Falken, 2021). However, these efforts have faced two primary constraints: high program costs and limited local supply of tutors.

Online tutoring provided by volunteers offers a low-cost model with the potential to reach more students in need. Tapping volunteers as tutors substantially reduces program costs given that tutor compensation often comprises a large portion of program budgets. Delivering tutoring online expands the supply of potential tutors, reduces time costs related to commuting, and allows for more flexible work schedules. It also offers the opportunity to better match students with tutors based on interests, expertise, and background.

In this paper, we report results from a randomized field trial of a pilot program to deliver online tutoring to middle school students in the spring of 2021. College students from highly selective universities served as volunteer tutors working 1-on-1 with predominantly low-income students of color twice a week for 30 minutes during the school day. Tutoring focused on building personal relationships with students and supplementing their learning in math and reading.

We find that the pilot program produced consistently positive but statistically insignificant effects on student achievement. The size of our overall intent-to-treat estimates, 0.07 standard deviations (σ) for math and 0.04σ for reading, are roughly a third as large as the pooled effect sizes for middle/high school tutoring programs reported by Nickow et al. (2020). Although our estimates are notably smaller than those found in many higher-dosage in-person tutoring programs, they are from a significantly lower-cost program that was delivered within the challenging context of the COVID-19 pandemic. Finally, we provide evidence that is consistent with a dosage model of tutoring where additional hours result in larger effects.

Our study contributes to a small but growing literature examining the impact of online tutoring, and is among the first to evaluate online tutoring in the U.S context (Carlana and La Ferrara, 2021). It also serves as proof of concept that low-cost online tutoring can be integrated into the regular school day during both remote and in-person learning. The scalability of the model rests critically on the open question of whether there is a broader supply of college students who are willing to volunteer their time as tutors.

2 Online Tutoring Program

To recruit tutors, we partnered with a volunteer mentorship organization, CovEducation (CovEd), founded by undergraduates at Harvard, MIT, and UT Austin in March 2020. CovEd is a non-profit organization run by college volunteers with the aim of recruiting undergraduates to support K-12 students in need of academic and socio-emotional support during the COVID-19 pandemic. By the end of the 2020-2021 school year, CovEd had matched volunteer undergraduates and recent graduates with over 5,000 K-12 students to provide personalized mentoring and tutoring.

A total of 230 tutors participated in the pilot program from across forty-seven different colleges and universities.[1] About three-fourths of our tutors were women; 40% were white, 34% were Asian, 20% were Hispanic and 5% were Black. About 70% were science or engineering majors, with business, economics, and other social sciences comprising the majority of other majors.[2]

Prior to the start of the program, CovEd provided tutors with a three-hour training on pedagogical techniques, relationship building, and educational resources. During the program, CovEd offered weekly peer mentoring sessions to troubleshoot common tutoring obstacles, share best-practices, and build community.

We matched tutors with 6th-8th grade students at Chicago Heights Middle School (CHMS) in Chicago Heights, IL. Ninety-nine percent of CHMS students are from low income backgrounds, almost two-thirds are Hispanic and a little less than one-third are Black. Prior to the pandemic, about a quarter of students were meeting grade-level standards in math and reading.

In February of 2021, CMHS offered students the choice to transition from remote learning to hybrid learning where students would attend school in person on a rotating two/three days a week schedule. Only students who chose to participate in hybrid learning were eligible to participate in the study (58% of students).

Starting in March of 2021, tutors met with their matched student 1-on-1 two days a week for thirty minutes during a daily advisory period. Online tutoring sessions took place when students were attending school in person and while they were learning remotely at home across alternating weeks. The program ran for a 12-week period which included a week off for spring break and several days off for state testing.

During the online Zoom sessions, tutors worked to build personal relationships and

[1] About two-thirds of the tutors came from Texas A&M University, UC San Diego, UT Austin, University of Chicago, and Harvard University.

[2] Other studies in this symposium discussed programs that undergraduates majoring in economics and STEM could support (e.g. Chuan et al., 2022; Gelfer et al., 2022; Fiala et al., 2022).

provide individualized tutoring in math and reading. Tutors were instructed to begin each session with a personal check-in and then inquire about help the students might need with their schoolwork. If students did not request help, tutors were directed to engage in guided instruction to build core skills in reading or math using the online textbooks used by the school (Carnegie Learning/MATHia and My Perspectives). In practice, many tutors reported using supplemental online resources in their tutoring sessions such as Newsela, Quizlet and Khan Academy.

3 Data and Experimental Framework

The aim of the pilot program was to support students' overall wellbeing as well as their academic achievement. Our analyses focus on the effect of tutoring on student achievement because of data constraints rather than a singular focus on academic performance. We measure student achievement in math and reading captured by two distinct assessments: the Illinois Assessment of Readiness (IAR) and the iReady tests. The IAR is the state standardized test administered for school accountability purposes. The iReady is a formative assessment developed by Curriculum Associates used to track student learning progress. Students took the math IAR exams on May 11th and 12th and the reading exams between May 18th and 20th before tutoring concluded. Students took the iReady tests in both subjects between May 26th and May 28th, the last week of the program.

We standardize all assessments to have a mean of 0 and a standard deviation of 1 within grade, subject, and test-type using control-group averages and standard deviations. We complement these achievement measures with administrative data on student characteristics and detailed timestamp records tracked by Zoom, which we cross-checked using time records collected by research assistants hosting the Zoom sessions.

We used a completely randomized experimental design to allocate a total of 560 students to either receive tutoring or a control condition (students in control participated in regular advisory period activities). We randomized within grade-level across three different waves as more students chose hybrid learning and more tutors became available.[3] Students randomized to the treatment condition were instructed by the school that they would receive 1-on-1 tutoring during their advisory period and were provided with instructions for logging into the sessions.

[3]Wave 1 on March 8th (n=268), wave 2 on March 22nd (n=209) and wave 3 on April 5th (n=83). Balance tests in the online appendix reveal no statistically significant differences between treatment and control groups for baseline test scores or student characteristics in our full sample or within individual waves.

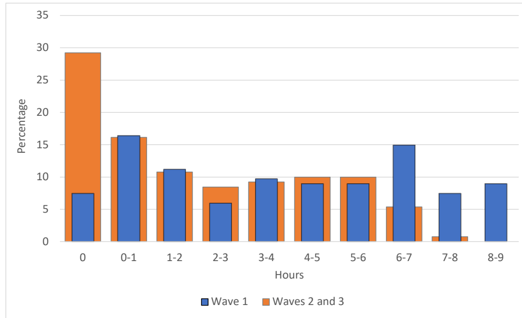

Figure 1: Hours spent with tutor

Figure 1 displays a histogram of the total number of tutoring hours treated students received in wave 1 versus waves 2 and 3. Students could have attended at most 18 total tutoring sessions across 12 weeks, or roughly 9 hours of tutoring. On average, students attended 3.1 hours of tutoring (4.0 in wave 1, 2.2 in waves 2 and 3). 18% percent of students did not attend a single minute of tutoring (7.5% in wave 1, 29% in waves 2 and 3). Low take-up was due in part to poor overall school attendance during this transitional period from fully-remote to hybrid learning. In addition, students randomized into later waves had fewer sessions they could attend.

Following our pre-registered analysis plan (Kraft et al., 2021), we estimate intent-to-treat (ITT) effects of the pilot tutoring program on student achievement from a standard OLS regression model that includes indicators for wave-by-grade randomization blocks and a vector of student-level controls, including quadratic functions of baseline iReady test scores in math and reading as well as indicators for race (Black, Hispanic, other), gender, grade, English as a second language, and treatment status from a previous experiment. We then use the assignment to treatment as an exogenous instrument for ever attending a complete tutoring session (25 minutes or more of total time) to estimate treatment-on-the-treated (TOT) effects in a two-stage least squares (2SLS) framework. Finally, we estimate variable treatment intensity using the total number of tutoring hours a student received in a parallel 2SLS framework. This IV approach identifies the average causal response of an additional hour of tutoring.

4 Results

4.1 Student Achievement

We find that online tutoring by college volunteers had positive but statistically insignificant effects on student achievement. Table 1 reports results from our full sample and

Table 1: Effect of tutoring on student achievement in math and reading

	Pooled tests	Math		Reading	
		iReady	IAR	iReady	IAR
	(1)	(2)	(3)	(4)	(5)
Intent to treat: Full Sample	0.053	0.068	0.077	0.043	0.040
	(0.043)	(0.057)	(0.057)	(0.059)	(0.060)
Intent to treat: Wave 1	0.087	0.128	0.124	0.051	0.034
	(0.062)	(0.079)	(0.086)	(0.083)	(0.084)
Intent to treat: Waves 2 & 3	0.021	-0.006	0.033	0.028	0.061
	(0.059)	(0.084)	(0.078)	(0.086)	(0.088)
Treatment on treated (at least one session): Full Sample	0.066	0.083	0.097	0.053	0.051
	(0.053)	(0.067)	(0.071)	(0.071)	(0.074)
Treatment on treated: Wave 1	0.097	0.140	0.140	0.057	0.038
	(0.069)	(0.083)	(0.093)	(0.089)	(0.091)
Treatment on treated: Waves 2 & 3	0.029	-0.009	0.047	0.038	0.087
	(0.082)	(0.108)	(0.108)	(0.114)	(0.120)
Average causal response (hours with tutor)	0.017	0.020	0.025	0.013	0.013
	(0.013)	(0.017)	(0.018)	(0.018)	(0.019)
Observations	2092	497	545	507	543

Notes: Each row of the panel reports estimates from a separate regression. iReady is a diagnostic exam administered by the school. IAR is the state-administered standardized test. In the average causal response regressions, treatment status is used as an instrument for total hours the student met with their tutor. Standard errors in parentheses. Standard errors are clustered by student in column 1. All regressions control for baseline math and reading scores, gender, race, ethnicity, EAL status, wave-by-grade fixed effects, indicators for treatment group in a previous experiment, and dummy variables indicating imputation of missing baseline test scores (10% of sample) to the sample mean.

separately by wave (pooling waves 2 and 3). Column 1 contains estimates from a stacked model that pools test scores across subjects and test types and reports standard errors clustered at the student level. Pooled ITT and TOT estimates are 0.053σ and 0.66σ, respectively. Columns 2 through 5 report treatment estimates for each individual test separately. ITT estimates across subjects and test types are quite consistent ranging from 0.040σ to 0.077σ, with slightly larger magnitudes in math compared to reading. Corresponding TOT estimates are only modestly larger (0.051σ to 0.097σ) given our take-up rate of 78.4%.

4.2 Dosage

The bottom row of Table 1 reports our 2SLS estimate of the average causal response of receiving an additional hour of tutoring. Here again our estimates are imprecise but suggest that an additional hour of tutoring increased test scores by 0.017σ [First Stage: β=3.2

hours]. Taking this estimate at face value, and imposing a somewhat strong assumption of a linear dosage effect, suggests that students who attended 9 hours of the tutoring program would have benefited from a 0.15σ increase in achievement. We caution against projecting effects of even higher dosage programs that fall outside of the range of common support of our data shown in Figure 1.

Disaggregating our estimates across the early wave 1 randomization group versus the later groups in waves 2 and 3 adds further evidence about possible dosage effects. Estimates for the pooled ITT effects on achievement are 0.087σ for wave 1 where students attended an average of 4.0 hours of tutoring. Students in waves 2 and 3 had fewer opportunities to attend tutoring sessions given their later start dates, averaging 1.2 hours of total tutoring. Our pooled ITT estimate for waves 2 and 3 is correspondingly smaller, 0.021σ. Although we cannot definitely rule out the possibility that this pattern is explained by differences in student or tutor characteristics across waves, we view these results as consistent with a model of positive returns to additional hours of tutoring.

4.3 Cost-Effectiveness

These estimated effects are meaningful in magnitude but notably smaller than those of higher-cost, higher-dosage, in-person tutoring programs with highly-trained tutors. Nickow et al. (2020) report a pooled effect size of 0.16σ from eight randomized control trials of tutoring programs serving grades 6 through 11. Studies of the Match/Saga high-dosage tutoring model delivered across the entire school year find effects as large as 0.37σ (TOT) on 9th and 10th grade achievement using a 2:1 student-to-tutor ratio with full-time, highly trained tutors at a cost of $3,800 per student (Guryan et al., 2021).

At the same time, the online volunteer tutoring program we study is extremely low cost making it far more cost-effective than other models. We spent $7,200 on RA staff to help coordinate tutoring, and $50 for a single Zoom account in which RAs paired students and tutors in breakout sessions.. CovEd operates with a thin budget, spending about $1,200 on their website and recruitment during the spring of 2021. Together, this amounts to a per-treated-student cost of roughly $32. However, this estimate does not reflect the hundreds of hours of uncompensated time that college students volunteered for tutoring or the effort of CovEd, the research team, and district teachers and staff. It also does not reflect the indirect costs the district incurred prior to our program to purchase laptops and online textbook materials.

5 Conclusion

Our study of a pilot program to deliver online tutoring by college volunteers serves as an initial test of the efficacy of low-cost online tutoring models integrated into the school day. It remains an open question whether such a model can be replicated at scale given the extraordinary efforts of undergraduate volunteers from highly selective colleges who served as tutors and operated CovEd. Yet, our approach does align well with the scaling principles that are arising in the science of using science literature (see e.g., List, 2022). Going forward, recruiting college volunteers to serve as long-term tutors may become more challenging than during the initial wave of volunteerism that emerged in the first year of the pandemic when many undergraduates were attending college remotely. If this is the case, scaling online tutoring programs may require expanding the portfolio of volunteers beyond college students or pivoting to higher-cost models that pay competitive wages.

References

Carlana, Michela and Eliana La Ferrara, "Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic," *CEPR Discussion Paper No. DP15761*, 2021.

Chuan, Amanda, John A. List, Anya Samek, and Shreemayi Samujjwala, "Gender Gaps in Parental Investments in Early Childhood," 2022.

Fiala, Lenka, John Eric Humphries, Juanna Schroter Joensen, Uditi Karna, John A List, and Gregory F. Veramendi, "How Early Adolescent Skills and Preferences Shape Economics Education Choices," 2022.

Gelfer, Sacha, Jeffrey A. Livingston, and Sutanuka Roy, "The Effect of Teaching Economics with Classroom Experiments: Estimates from a Within-Subject Experiment," 2022.

Guryan, Jonathan, Jens Ludwig, Monica P Bhatt, Philip J Cook, Jonathan MV Davis, Kenneth Dodge, George Farkas, Roland G Fryer Jr, Susan Mayer, Harold Pollack, and Laurence Steinberg, "Not Too Late: Improving Academic Outcomes Among Adolescents," *National Bureau of Economic Research*, 2021.

Kraft, Matthew A and Grace Falken, "A blueprint for scaling tutoring across public schools," *AERA Open*, 2021, *7.*

_ , John A List, Livingston, Jeffrey A, and Sally Sadoff, "Online Tutoring During COVID-19," *AEA RCT Registry https://doi.org/10.1257/rct.7284-1.0*, June 08 2021.

List, John A, *The voltage effect: How to make good ideas great and great ideas scale*, Currency, 2022.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan, "The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence," *National Bureau of Economic Research*, 2020.

A Appendix: Balance of student characteristics by treatment

We balanced our randomization on each of the variables that appear in the balance tables displayed below. Table A1 reports balance of these characteristic across treatment and control for the entire sample. Table A2 reports balance for students who entered the experiment in wave 1 only. Table A3 reports balance for students who entered the experiment in wave 2 or wave 3 only.

As the p-values reported in column 3 of each table, there are no statistically significant differences between our treatment and control groups for any of these characteristics.

Baseline math and reading scores are 2020 iReady scores standardized by grade. CHECC treatment indicators are controls for participation in an experiment that began with preschool-age children in Chicago Heights, IL, where our experiment was also conducted. The authors of that study are following long-term outcomes for the students who were part of that experiment. To avoid biasing their results, we balanced our randomization on that experiment's treatment groups. Many of the children who were part of that experiment were in 6th, 7th, or 8th grade at the time of our experiment. Students who were not part of the CHECC study are the omitted category in our regressions.

Table A1: Balance of student characteristics by treatment and control: All waves

	Control (1)	Treatment (2)	p-val of diff (3)
Baseline math score	-0.033	-0.031	0.939
	(0.956)	(0.920)	
Baseline reading score	-0.036	-0.084	0.556
	(0.897)	(0.993)	
Female	0.470	0.473	0.848
	(0.500)	(0.500)	
Black	0.314	0.303	0.736
	(0.465)	(0.460)	
Hispanic	0.632	0.636	0.875
	(0.483)	(0.482)	
ELL status	0.270	0.261	0.782
	(0.445)	(0.440)	
CHECC control	0.199	0.182	0.675
	(0.400)	(0.386)	
CHECC treatment 1	0.047	0.057	0.632
	(0.213)	(0.232)	
CHECC treatment 2	0.108	0.095	0.594
	(0.311)	(0.293)	
CHECC treatment 3	0.091	0.095	0.862
	(0.288)	(0.293)	
Observations	296	264	

Table A2: Balance of student characteristics by treatment and control: Wave 1

	Control (1)	Treatment (2)	p-val of diff (3)
baseline math score	-0.013	0.025	0.735
	(0.937)	(0.894)	
baseline reading score	0.001	-0.022	0.842
	(0.892)	(1.006)	
Female	0.463	0.493	0.625
	(0.500)	(0.502)	
Black	0.291	0.284	0.892
	(0.456)	(0.452)	
Hispanic	0.649	0.687	0.517
	(0.479)	(0.466)	
ELL status	0.284	0.276	0.893
	(0.452)	(0.449)	
CHECC control	0.194	0.187	0.877
	(0.397)	(0.391)	
CHECC treatment 1	0.060	0.060	1.000
	(0.238)	(0.238)	
CHECC treatment 2	0.127	0.112	0.695
	(0.334)	(0.316)	
CHECC treatment 3	0.112	0.097	0.691
	(0.316)	(0.297)	
Observations	134	134	

Table A3: Balance of student characteristics by treatment and control: Waves 2 and 3

	Control (1)	Treatment (2)	p-val of diff (3)
baseline math score	-0.049 (0.974)	-0.088 (0.946)	0.835
baseline reading score	-0.067 (0.903)	-0.147 (0.980)	0.531
Female	0.475 (0.501)	0.454 (0.500)	0.832
Black	0.333 (0.473)	0.323 (0.469)	0.739
Hispanic	0.617 (0.488)	0.585 (0.495)	0.699
ELL status	0.259 (0.440)	0.246 (0.432)	0.799
CHECC control	0.204 (0.404)	0.177 (0.383)	0.665
CHECC treatment 1	0.037 (0.189)	0.054 (0.227)	0.477
CHECC treatment 2	0.093 (0.291)	0.077 (0.268)	0.719
CHECC treatment 3	0.074 (0.263)	0.092 (0.291)	0.494
Observations	162	130	