NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

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

Susanna Loeb
Stanford University

Sarah Novicoff
Stanford University

Cynthia Pollard
Stanford University

Carly Robinson
Stanford University

Sara White
Vanderbilt University

October 2023

studentsupportaccelerator.org

Abstract

This study is the first randomized controlled trial of a virtual early literacy tutoring program. OnYourMark Education provides tutoring grounded in the science of reading and focused on foundational literacy skills (e.g., phonics, phonological awareness, reading fluency). During the 2022-23 school year, OnYourMark partnered with 12 schools in a large charter-management organization in the southern United States to provide virtual tutoring in school to kindergarten, first, and second grade students. The program included four sessions per week for 20 minutes per session from September through May. We randomly assigned students to one-on-one tutoring (N=510), two-on-one tutoring (N=570), or a business-as-usual control group (N=1,005). We find that students assigned to OnYourMark tutoring scored approximately 0.08 SD higher on end-of-year early literacy tests than students in the business-as-usual control group, with lower-performing students (0.18 SD) and first graders (0.19 SD) assigned to 1:1 tutoring benefiting most. These positive findings from a virtual program with young readers provides initial evidence that virtual tutoring could be a promising option, especially in contexts with barriers to implementing in-person early literacy tutoring.

We are grateful for the research partnership with OnYourMark Tutoring and the partnering charter organization as well as the Overdeck Family Foundation who provided funding for this study.

Introduction

Learning to read presents a challenge for many students in the United States. In the 30 years that the National Assessment of Educational Progress has administered a reading exam, no more than 37 percent of students have scored at the proficient level or above. In other words, for at least 30 years, more than 60 percent of U.S. students have struggled to learn to read (NAEP Reading, 2022).

An unusually strong research base provides evidence that intensive tutoring can improve academic achievement generally (Dietrichson et al., 2017; Nickow et al., 2020) and in early literacy specifically (Elbaum et al., 2000; Gersten et al., 2020; Neitzel et al., 2022) with effect sizes ranging from 0.24 to 0.41 standard deviations (SDs). The effectiveness of tutoring programs varies widely. Programs with certain program features tend to be more effective than others. For example, while a wide range of tutors can be advance student learning, teachers tend to produce larger gains when they are tutors than do paraprofessionals or volunteers (Nickow et al., 2020), though volunteers can still have positive and meaningful effects (Ritter et al., 2009). The same patterns hold true in the early literacy context, with teachers being more effective tutors than others (Gersten et al., 2020; Slavin et al., 2011). Similarly one-on-one tutoring tends to produce larger achievement gains than programs in which students work with their tutor in small groups, and this pattern holds true both generally (Nickow et al., 2020) and in early literacy contexts, in particular (Gersten et al., 2020; Neitzel et al., 2022).

Though researchers have established the efficacy of in-person tutoring, rigorous causal research of online tutoring programs is sparse and focuses only on older students. A randomized controlled trial of online video-call-based tutoring delivered by volunteer university students to students in grades six through eight produced an effect size of 0.26 SD on multiple subjects (Carlana & Ferrara, 2021). A similar video-call-based tutoring intervention with teachers in Spain produced an effect size of 0.26 SD in math (Gortazar et al., 2023). In the United States context, a randomized control trial of online tutoring for middle school students by volunteers improved math test scores by 0.07 SD and reading test scores by 0.04 SD (Kraft et al., 2022).

Research on online tutoring for elementary students is even more sparse, perhaps due to concerns about whether younger children can sustain attention and build connections online. We identified two randomized controlled trials of online tutoring for elementary students in developed countries. For one online tutoring intervention implemented by teachers of fifth graders in the United States and Canada, research showed a 0.41 SD effect size on math, though the measure used was not a standardized test and covered only targeted content, thus potentially overstating the effectiveness of the program on other, broader measures (Roschelle et al., 2020). In another study, in which tutors from Sri Lanka and India supported year 6 students (ages 10-11) from England in math, the intervention produced a non-statistically-significant 0.002 SD change in math test scores and a 0.01 SD change in English (Torgerson et al., 2016).

In this brief, we present the first rigorous causal evidence of the efficacy of online literacy tutoring for young students. Specifically, we use a randomized controlled trial to examine the effects of OnYourMark, a fully virtual tutoring program for students in kindergarten through second grade. We

leverage multiple treatment arms to test the extent to which the ratio of students to tutors affects program efficacy.

In our primary analyses, we find that students assigned to receive OnYourMark tutoring performed 0.08 SD higher on their end-of-year DIBELS composite score than students assigned to the business-as-usual control group. Examining effects by baseline reading performance and by grade reveals that students with lower beginning-of-year reading scores and first graders benefited most (0.14 SD, $p<.05$; 0.12 SD, $p<.10$, respectively). We find that 1:1 tutoring was more effective (0.101 SD, $p<.05$) than 2:1 tutoring (0.041, $p>.10$) relative to the business-as-usual (BAU) control group. At the same time, our examination of effects on DIBELS subtests provides evidence that the effect of group size may depend on the grade of students and literacy skills targeted. For example, 1:1 tutoring may be particularly effective for first graders developing decoding skills.

A High-Impact Tutoring Model Delivered Virtually

The COVID-19 pandemic and subsequent federal funding prompted the creation and expansion of many tutoring companies promising to help address unprecedented reductions in learning (Robinson & Loeb, 2021). Companies offering completely virtual tutoring emerged especially quickly, given the restrictions on in-person schooling. Their models varied widely in features and experiences, from on-demand, chat-based platforms to comprehensive models aligned with best practices of high-impact tutoring (Robinson & Loeb, 2021).

OnYourMark Education began in 2021 offering virtual, full-service early literacy tutoring grounded in the science of reading. The evidence-based features of OnYourMark's program include:

- Small-group (1:1 or 2:1) sessions embedded into the school day;
- A consistent tutor assignment for each student;
- Initial tutor training and ongoing professional development focused on content knowledge, building relationships, and effective delivery of the intervention;
- High dosage and frequency of sessions (20-minute sessions, four times per week);
- A high-quality curriculum informed by the science of reading with a focus on phonics, phonological awareness, and fluency; and
- Regular progress monitoring and data use to target specific early literacy skills via two-week lesson sprints.

In its initial launch in fall 2021, OnYourMark served 58 students in one school and by spring 2022 had expanded to serve 180 students across two states in two schools. By the fall 2022, OnYourMark had partnered with districts in two states to serve 1,400 kindergarten, first, and second grade students.

Methods

Study Details

During the 2022-23 school year, a charter management organization in Texas partnered with OnYourMark to provide early literacy tutoring to kindergarteners, first graders, and second graders in 12 of its schools. The program included four sessions for students per week for 20 minutes per session. Students were in their classrooms while their tutors were online. Tutoring rolled out in September and continued through May.

In each of the 12 schools, building-level staff selected approximately 10 students who would be guaranteed to receive tutoring. These 121 students are excluded from the study sample because their enrollment in tutoring was not determined by random assignment. Schools then selected double the number of students as they had remaining seats to potentially receive tutoring. For example, if a school was allotted 96 tutoring seats, it selected ten students guaranteed to receive tutoring and then indicated eligibility and availability for another 172 students to fill the remaining 86 tutoring seats.

We assigned the eligible students into pairs and then randomly assigned the pairs of students into either treatment or control conditions. Among the student pairs in the treatment condition, we randomly assigned half to receive 1:1 tutoring and the other half to receive 2:1 tutoring. For example, in a school with 130 available spots, 10 slots would be filled with students selected by their school for need, leaving 120 slots for the study. The school provided a list of 240 students, and we randomly assigned 120 to the BAU control condition, and then 60 to 1:1 tutoring and 60 to 2:1 tutoring. When the sample size was uneven, the 1:1 and 2:1 conditions could have slightly different numbers.

Our final study sample consists of 2,085 students in which 510 students were assigned to 1:1 tutoring, 570 students were assigned to 2:1 tutoring and the remaining 1,005 students were assigned to the business-as-usual condition and placed on a randomly ordered waitlist; even if those students moved off of the waitlist, we still consider them control students for our analysis.

A considerable number of students randomized into study conditions were identified as multilingual learners and/or students with disabilities, and so received other special services. In some cases, students' specialized instructional needs did not preclude them from participating in OnYourMark tutoring. However, requirements for more intensive support and/or associated scheduling changes often resulted in students with disabilities and multilingual learners being withdrawn from tutoring. As such, there may be systematic differences between treatment and control groups in how multilingual learners and students with disabilities were removed from groups. Given potential selection biases associated with these differences, we present our main findings on the effects of tutoring and group size for both the full sample of students randomized, and then for the sample excluding multilingual learners and students with disabilities.

Data

Our primary outcome is performance on the end-of-year Dynamic Indicators of Basic Literacy Skills (DIBELS) 8th edition composite score. DIBELS is a widely-used and extensively validated set of measures and procedures to assess the acquisition of literacy skills (Smolkowski & Cummings, 2016). Additionally, DIBELS is well aligned to the early literacy skills targeted by the OnYourMark intervention. Classroom teachers assessed students on DIBELS at the beginning (BOY), middle (MOY) and end-of-year (EOY).

The DIBELS composite score is made up of a series of subtests that are typically 60-second, individually administered assessments that measure specific literacy subskills (e.g., letter sounds, decoding, reading fluency). The set of subtests administered and included in the composite score varies across grades.

As an additional measure of reading performance, we use EOY scores on the NWEA MAP Reading Fluency assessment. The MAP Reading Fluency assessment is a 20-minute online adaptive assessment designed for universal screening and progress monitoring of literacy skills for students in grades PK-5. MAP Reading Fluency has a strong focus on reading fluency, comprehension, and foundational reading skills (NWEA, 2023).

We also use student administrative data provided by the districts, including student demographics[1] (identifiers for gender, race/ethnicity, low-income economically marginalized, multilingual learner, student with disabilities), as well as school attendance rates. Finally, we collected tutoring session-level attendance and instructor data captured by the online tutoring platform.

Analysis

We preregistered our study design, hypotheses, and analytic plan on the Open Science Framework prior to conducting the primary analysis (see: https://osf.io/pq4g6). We use the following model to evaluate the impact of receiving tutoring on student outcomes:

$$Y_{ijk} = \alpha + \beta_1 Treatment_j + \gamma X_i + \omega_k + \epsilon_{ij}$$

where α = intercept; Y_{ijk} is the outcome of interest for student i with tutor j in school by grade k; $Treatment_j$ is the indicator for student assignment to receive OYM tutoring, X_i = vector of student-level covariates, including demographics and Fall 2022 BOY DIBELS scores; ω_k = fixed effect for strata (school x grade), ϵ_{ij} = a residual clustered at the pair level to account for the nesting of students within tutor groups. We used a similar regression model to evaluate the impact of receiving 1:1 tutoring and 2:1 tutoring relative to the control group.

[1] In the absence of information on how students/families self-identify, we use language describing demographic characteristics of participants that is most inclusive of individuals with different identities and experiences (APA, 2021). We sometimes use language to describe participants that diverge from the category labels used in administrative data.

Additionally, we conducted exploratory analyses that examined heterogeneity of the effect of OnYourMark tutoring for students by baseline reading performance on DIBELS and by grade. Specifically, we compared results for students who scored above and below the 50th percentile on the BOY DIBELS assessment, and for kindergartners, first, and second grade students.

Results

Descriptive Statistics and Balance Check

Table 1 displays descriptive statistics of the sample of students in each of the study conditions and tests for differences. The top panel confirms that students in the full sample were similar across study conditions in both demographic characteristics and on their baseline DIBELS scores. The bottom panel shows that when multilingual learners and students with disabilities are excluded from the analytic sample of students, the demographics of the students remain similar across conditions. We find no evidence of differences across groups at traditional levels of statistical significance.

Table 2 displays the results of attrition analyses predicting withdrawal status, missingness of outcome data, and being switched from the waitlist control group to treatment. The most frequent reasons for withdrawal included transferring out of the school (N=68) and accommodations for separate support needs for multilingual learners and students with disabilities (N=58).

The first column of Table 2 shows that treatment students (compared to control students) were more likely to be withdrawn, as were multilingual learners, students with disabilities, and, to a lesser extent, male students (compared to female students). The tutoring administrators were more likely to record withdrawals for students enrolled in tutoring than for students who were in the control group, at least in part because tutoring conflicted with the other services the withdrawn students were receiving while control students could receive those additional services as part of BAU. We run analyses that do not include multilingual learners and students with disabilities because of this differential withdrawal.

Columns 2 and 3 of Table 2 show that missingness of end-of-year DIBELS and MAP scores are similar across the treatment and control groups, though Asian American, Black and Latina/o/x students were slightly more likely to be missing EOY MAP scores. Column 4 shows that students switched from the waitlisted control group into tutoring had slightly lower beginning-of-year DIBELS scores than students overall. These switched students remain in our control group, and thus our experimental estimate provides a lower bound of the effect as some students in the control group did receive OnYourMark tutoring despite their assignment.

Table 1. Sample Descriptive Statistics and Balance Test

	1:1 Tutoring		2:1 Tutoring		BAU Control		
	Mean	N	Mean	N	Mean	N	*p-value*
A. Sample of Randomized Students							
Student Demographics							
Female	.49	510	.49	570	.51	1,005	0.798
Asian American	>.01	510	.01	570	.01	1,005	0.437
Black	.25	510	.23	570	.26	1,005	0.468
Latina/o/x	.66	510	.67	570	.66	1,005	0.869
Multiracial	.03	510	.03	570	.01	1,005	0.075
White	.04	510	.05	570	.03	1,005	0.378
MLL	.32	478	.32	532	.31	929	0.936
SWD	.32	478	.25	532	.43	929	0.661
LIEM	.83	478	.82	532	.84	929	0.666
Grade							
Kindergarten	.29	510	.28	570	.30	1,005	0.843
First	.39	510	.41	570	.41	1,005	0.751
Second	.32	510	.32	570	.29	1,005	0.401
BOY DIBELS Comp.	307.80	509	306.57	570	305.05		
	(33.70)		(31.21)		(31.62)	1.004	0.274
B. Sample of Randomized Students Excluding Multilingual Learners and Students with Disabilities							
Student Demographics							
Female	.55	331	.50	367	.53	656	0.47
Asian American	.01	331	.01	367	.01	656	0.87
Black	.35	331	.32	367	.36	656	0.48
Latina/o/x	.51	331	.50	367	.50	656	0.91
Multiracial	.05	331	.04	367	.03	656	0.18
White	.05	331	.07	367	.04	656	0.20
LIEM	.77	331	.77	367	.80	656	0.55
Grade							
Kindergarten	.29	331	.27	367	.32	656	0.29
First	.40	331	.43	367	.41	656	0.72
Second	.31	331	.30	367	.27	656	0.47
BOY DIBELS Comp.	310.67	330	308.99	367	305.94	656	0.08
	(35.54)		(30.62)		(32.42)		

Note: MLL= Multilingual learners; SWD= Students with disabilities. LIEM=Low-income economically marginalized. *P-value* is from an F-statistic.

Table 2. Attrition Analysis

	Withdrawn	Missing EOY DIBELS	Missing EOY MAP	Switched
	(1)	(2)	(3)	(4)
Treatment	0.067***	-0.002	0.000	–
	(0.008)	(0.008)	(0.004)	
BOY DIBELS Comp.	-0.010+	-0.004	-0.002	-0.018*
	(0.005)	(0.004)	(0.002)	(0.007)
Female	-0.017*	0.002	-0.004	0.043**
	(0.008)	(0.008)	(0.004)	(0.015)
Asian American	0.046	-0.030	0.006*	0.185
	(0.061)	(0.019)	(0.003)	(0.122)
Black	-0.002	-0.026	0.013*	0.016
	(0.021)	(0.022)	(0.006)	(0.043)
Latina/o/x	-0.006	-0.024	0.006+	0.012
	(0.020)	(0.021)	(0.004)	(0.042)
Multiracial	0.004	-0.027	0.000	-0.053
	(0.036)	(0.025)	(0.003)	(0.054)
MLL	0.033**	-0.011	0.000	0.010
	(0.011)	(0.009)	(0.004)	(0.018)
SWD	0.124***	0.030	0.012	0.001
	(0.033)	(0.024)	(0.012)	(0.033)
Constant	-0.001	0.061**	-0.000	0.087*
	(0.020)	(0.021)	(0.004)	(0.041)
School-Grade FE	Yes	Yes	Yes	Yes
R^2	0.100	0.211	0.034	0.047
Observations	1937	1937	1937	1937

Notes: MLL= Multilingual learners; SWD= Students with disabilities. “Switched” indicates that a student was switched from the randomly ordered, waitlisted control group to the treatment group. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

DIBELS Results

Table 3 presents our primary results. As shown, for the overall sample, with no controls, OnYourMark tutoring performed 0.097 SD (p<.05) higher on their end-of-year DIBELS scores than students assigned to the BAU control group. We present a number of alternative models, all showing positive effects of OnYourMark. When controlling for baseline reading performance and student demographic characteristics, students assigned to receive OnYourMark tutoring performed 0.053 SD (p<.10) higher on their end-of-year DIBELS scores than students assigned to the BAU control group.

For our preferred sample excluding multilingual learners and students with disabilities for the reasons discussed above, we find that students assigned to OnYourMark tutoring performed 0.075 SD higher (p<.05) on the end-of-year DIBELS assessment. Figure 1 shows that this estimated effect translates to OnYourMark students performing 1.57 percentile points higher than students assigned to the BAU control group.

Table 3. The Effect of Tutoring on Standardized End-of-Year DIBELS Composite Scores

	(1)	(2)	(3)	(4)	(5)
Tutoring	0.097*	0.053+	0.059+	0.070*	0.075*
	(0.044)	(0.029)	(0.030)	(0.034)	(0.036)
School-Grade FE	Yes	Yes	Yes	Yes	Yes
DIBELS Control	No	Yes	Yes	Yes	Yes
Student Demos	No	Yes	Yes	Yes	Yes
Sample					
Includes MLL	Yes	Yes	Yes	No	No
Includes SWD	Yes	Yes	No	Yes	No
R^2	0.108	0.578	0.574	0.580	0.578
Control Mean	0.092	0.072	0.095	0.130	0.151
Observations	1869	1867	1765	1238	1163

Notes. SWD= Students with disabilities, MLL= Multilingual learners. Student-level controls include dummy variables for female, Black, Latina/o/x, MLL, SWD, and low-income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Figure 1. Average DIBELS Composite Percentile by Condition Assignment

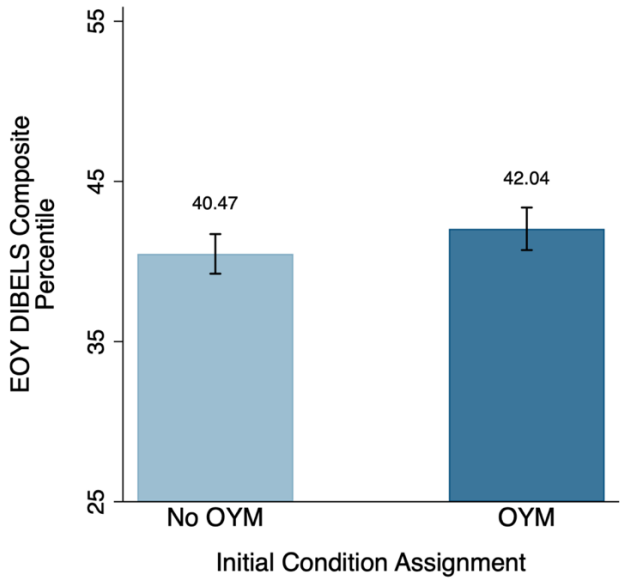

Table 4 displays results of analyses estimating the effects of OnYourMark tutoring on end-of-year MAP Reading Fluency scores. We find no effect on MAP scores for this full sample of students.

Table 4. Effect of Tutoring on Standardized MAP Reading Scores

	(1)	(2)	(3)	(4)	(5)
OYM Tutoring	0.027	-0.006	0.009	0.014	0.038
	(0.039)	(0.030)	(0.029)	(0.040)	(0.039)
School-Grade FE	Yes	Yes	Yes	Yes	Yes
DIBELS Control	No	Yes	Yes	Yes	Yes
Student Demos	No	Yes	Yes	Yes	Yes
Sample					
Includes SWD	Yes	Yes	No	Yes	No
Includes MLL	Yes	Yes	Yes	No	No
R^2	0.073	0.374	0.375	0.366	0.377
Control Mean	-0.156	-0.171	0.095	-0.076	-0.046
Observations	1928	1926	1819	1279	1201

Notes. SWD= Students with disabilities, MLL= Multilingual learners. Student-level controls include dummy variables for female, Black, Latina/o/x, MLL, SWD, and low-income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Group Size

Table 5 provides estimates for analyses estimating the effect of 1:1 tutoring and 2:1 tutoring relative to the control group on both EOY DIBELS and EOY MAP scores. Across the board estimates of the effects of 1:1 tutoring, which is the model that OnYourMark uses most often, are higher than estimates of the effects of 2:1 tutoring. In our final model controlling for baseline reading achievement and student demographics, we find that students assigned to 1:1 tutoring performed 0.112 SD higher than students assigned to the BAU control group, which is statistically significant ($p<.05$). The estimated benefit of 2:1 tutoring relative to the BAU control group was smaller (0.041 SD) than the benefit from 1:1 tutoring, and this estimated effect was not statistically significant.

Examining results on MAP Reading Fluency scores, we find that in the model controlling for baseline reading performance and student demographics, the effect of tutoring relative to the control group was greater for students assigned to 1:1 than to 2:1 (0.052 vs 0.025), though neither of these estimated effects is statistically significant.

Table 5. Effect of 1:1 and 2:1 Tutoring on Reading Achievement

	DIBELS Composite Score					MAP Reading Fluency Score				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1:1 Tutoring	0.121*	0.062+	0.070+	0.101*	0.112*	0.030	-0.013	0.003	0.032	0.052
	(0.056)	(0.034)	(0.036)	(0.044)	(0.045)	(0.050)	(0.038)	(0.036)	(0.050)	(0.050)
2:1 Tutoring	0.076	0.045	0.049	0.041	0.041	0.024	0.001	0.015	-0.003	0.025
	(0.051)	(0.039)	(0.040)	(0.044)	(0.046)	(0.046)	(0.039)	(0.038)	(0.049)	(0.047)
School-Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DIBELS Control	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Student Demos	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Sample										
Includes SWD	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	No
Includes MLL	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
R^2	0.108	0.578	0.574	0.580	0.578	0.073	0.374	0.375	0.366	0.377
Control Mean	-0.005	0.019	0.036	0.060	0.076	-0.183	-0.165	-0.149	-0.090	-0.084
Observations	1869	1867	1765	1238	1163	1928	1926	1819	1279	1201

Notes. MLL= Multilingual learners, SWD= Students with disabilities. Student-level controls include dummy variables for female, Black, Latina/o/x, MLL, SWD, and low-income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Heterogeneity Analysis

Baseline Reading Performance

We tested whether the effect of OnYourMark tutoring differed for students who performed in the bottom and top half of BOY DIBELS scores, and found that the estimates were larger for initially lower-scoring students. Panel A in Table 6 displays effects on DIBELS and MAP scores for students in the bottom half of the distribution and Panel B shows the effects for students in the top half. Overall, we find that the effect of tutoring was statistically significant and stronger for students with lower BOY DIBELS scores. This trend is consistent for tutoring overall, as well as for students assigned to 1:1 and 2:1 groups.

These analyses reveal that students with lower scores at beginning experienced positive and statistically significant effects from 2:1 tutoring as well as from 1:1 tutoring. The positive effect of 1:1 tutoring is significant both for students in the bottom half of baseline reading performance and for students in the top half.

Table 6. Heterogeneity Analyses by Baseline Reading Performance

	Students At/Below 50th Percentile at Baseline				Students Above 50th Percentile at Baseline			
	DIBELS		MAP		DIBELS		MAP	
A. Effect of Tutoring Overall vs. BAU Control								
Tutoring Overall	0.149*	0.142*	0.061	0.041	0.073	0.061	0.041	0.038
	(0.065)	(0.054)	(0.064)	(0.054)	(0.064)	(0.047)	(0.064)	(0.060)
B. 1:1 and 2:1 Tutoring vs. BAU Control								
1:1 Tutoring	0.166+	0.177*	0.071	0.050	0.138+	0.103+	0.089	0.068
	(0.086)	(0.079)	(0.085)	(0.075)	(0.079)	(0.057)	(0.079)	(0.072)
2:1 Tutoring	0.133+	0.109+	0.051	0.032	0.010	0.020	-0.004	0.011
	(0.076)	(0.058)	(0.077)	(0.066)	(0.079)	(0.062)	(0.072)	(0.071)
School-Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DIBELS Control	No	Yes	No	Yes	No	Yes	No	Yes
Student Demos	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.175	0.363	0.117	0.274	0.204	0.479	0.154	0.269
Control Mean	-0.585	-0.581	-0.534	-0.524	0.574	0.581	0.259	0.260
Observations	515	515	527	527	648	648	674	674

Notes. Coefficients in the top panel are from models comparing assignment to any tutoring to the BAU control. Coefficients in the second panel are from models comparing each tutoring model (1:1, 2:1) to the BAU control. All models exclude MLLs and SWDs. Student-level controls include dummy variables for female, Black, Latina/o/x, and low-income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Effects by Grade

Table 7 provides the estimates of the differences in the effect of OnYourMark tutoring for kindergarten, first, and second grade students. It reveals that on DIBELS scores, the effect of OnYourMark tutoring was significant and strongest for first graders, followed by kindergarteners. The estimated effects for second grade were smaller in magnitude and not significant.

For first graders, we see positive effects on MAP scores overall, and for assignment to both 1:1 and 2:1 groups, suggesting that the MAP Reading Fluency assessment may be more sensitive to the OnYourMark intervention in first grade compared to in kindergarten and second grade.

Table 7. Effect of Tutoring Group by Grade

	Kindergarten		1st Grade		2nd Grade	
	DIBELS	MAP	DIBELS	MAP	DIBELS	MAP
A. Effect of Tutoring Overall vs. BAU Control						
Tutoring Overall	0.077	-0.005	0.089+	0.117*	0.031	-0.046
	(0.066)	(0.080)	(0.048)	(0.053)	(0.075)	(0.064)
B. 1:1 and 2:1 Tutoring vs. BAU Control						
1:1 Tutoring	0.105	0.031	0.186**	0.134+	0.008	-0.037
	(0.084)	(0.101)	(0.068)	(0.074)	(0.080)	(0.083)
2:1 Tutoring	0.047	-0.042	0.014	0.104+	0.056	-0.054
	(0.092)	(0.102)	(0.055)	(0.060)	(0.098)	(0.084)
R^2	0.473	0.308	0.653	0.437	0.676	0.425
Control Mean	0.043	-0.015	0.077	-0.050	0.122	-0.195
Observations	347	350	504	506	312	345

Notes. Coefficients in the top panel are from models comparing assignment to any tutoring to the BAU control. Coefficients in the second panel are from models comparing each tutoring model (1:1, 2:1) to the BAU control. All models exclude multilingual learners and students with disabilities. All model control for school-grade fixed effects, and baseline reading performance. Student-level demographic controls include dummy variables for female, Black, Latina/o/x, and low-income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

DIBELS Subtests

Tables 8, 9, and 10 display results from analyses estimating the impact of assignment to OYM on DIBELS subtests for kindergartners, first, and second graders, respectively. DIBELS subtests measure discrete literacy skills and as such, scores provide more actionable information than composite scores, which have the potential to obscure gains and needs on specific literacy skills. Subtests are typically one-minute tests administered in order of skill development. The developers of DIBELS provide guidance on discontinuing DIBELS testing when a student does not answer any items on a given subtest correctly (University of Oregon, 2023). In such cases, they do not want students to be given subsequent subtests in order to save time and avoid student frustration. While scores on discontinued subtests should be recorded as 0, in practice some administrators do not record a score. As students age, the sub-tests that they are administered changes. For first graders, DIBELS begins by testing letter sounds and word recognition and only administers letter names and phonemic awareness to students who struggle on the initial assessments. As a result, missingness in letter names and phonemic awareness is endogenous to performance and we exclude those measures.

Our analyses show that OYM tutoring was most effective in improving kindergarteners' mastery of letter sounds. These changes in raw scores translate to an additional 3.5 letter sounds recognized per minute. First graders' letter sound mastery and decoding skills also improved from OnYourMark tutoring. For second graders, we do not see significant effects of tutoring on any subtests.

Table 8. The Effect of OnYourMark on Kindergartners' DIBELS Subtest Scores

	Letter Names	Phon. Awareness	Letter Sounds	Word Recognition	Decoding
A. Effect of Tutoring Overall vs. BAU Control					
Tutoring Overall	0.170	2.575	3.500*	-0.786	1.020
	(1.584)	(1.580)	(1.380)	(-0. .982)	(.729)
B. 1:1 and 2:1 Tutoring vs. BAU Control					
1:1 Tutoring	1.992	2.300	3.713*	-1.305	1.337
	(1.866)	(1.961)	(1.812)	(1.191)	(0.964)
2:1 Tutoring	-1.732	2.861	3.277	-0.245	0.689
	(1.929)	(1.985)	(1.94)	(1.363)	(0.918)
CR^2	0.336	0.278	0.413	0.423	0.360
Control Mean	45.561	26.236	37.442	14.808	37.967
Observations	347	347	347	347	347

Notes. Estimates are from raw scores. Coefficients in the top panel are from models comparing assignment to any tutoring to the BAU control. Coefficients in the second panel are from models comparing each tutoring model (1:1, 2:1) to the BAU control. All models exclude MLLs and SWDs. All model control for school-grade fixed effects, and baseline reading performance. Student-level demographic controls include dummy variables for female, Black, Latina/o/x, and low income economically marginalized. Standard errors clustered at the student pair level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 9. The Effect of OnYourMark on First Graders' DIBELS Subtests Scores

	Letter Sounds	Word Recognition	Decoding	Reading Accuracy	Reading Fluency	Fluency Error Rate	Spelling
A. Effect of Tutoring Overall vs. BAU Control							
Tutoring	4.016	1.341	1.717	2.228	0.783	-0.574+	0.226
Overall	(2.692)	(1.344)	(1.110)	(1.625)	(2.004)	(0.323)	(1.021)
B. 1:1 and 2:1 Tutoring vs. BAU Control							
1:1 Tutoring	7.863+	2.453	3.884*	0.611	3.145	-0.633***	0.702
	(4.109)	(1.719)	(1.532)	(1.384)	(2.909)	(0.010)	(1.29)
2:1 Tutoring	0.971	0.451	0.000	3.030+	-1.085	-0.528	-0.138
	(2.843)	(1.524)	(1.203)	(1.585)	(2.244)	(0.371)	(1.257)
R^2	0.398	0.541	0.366	0.333	0.599	0.253	0.381
Control Mean	66.578	33.085	17.630	88.724	58.056	3.194	41.209
Observations	504	504	504	504	504	504	475

Table 10. The Effect of OnYourMark on Second Graders' Graders' DIBELS Subtests Scores

	Reading Accuracy	Reading Fluency	Reading Compre-hension	Maze Correct Responses	Maze Incorrect Responses	Spelling
A. Effect of Tutoring Overall vs. BAU Control						
Tutoring	-0.077	1.593	3.388	-0.089	0.457	-0.087
Overall	(1.093)	(2.772)	(4.411)	(0.544)	(1.338)	(1.377)
B. 1:1 and 2:1 Tutoring vs. BAU Control						
1:1 Tutoring	-0.341	-0.130	1.124	0.527	1.382	-0.354
	(1.217)	(1.547)	(1.297)	(3.034)	(5.509)	(0.669)
2:1 Tutoring	-0.207	0.288	-1.326	2.704	5.475	0.190
	(1.513)	(1.615)	(1.341)	(3.779)	(5.059)	(0.712)
R^2	0.641	0.659	0.238	0.132	0.168	0.562
Control Mean	90.600	70.646	71.458	9.069	8.754	50.211
Observations	312	312	312	312	312	283

Translated to additional learning, the positive effects on students' specific reading skills can be interpreted as the proportion of the control group's average gains from beginning to end of year. For example, kindergartners assigned to tutoring gained an additional 13.4% of the control group's average gains in letter sound mastery from beginning to end of year, or what translates to an additional about 24 days of school based upon the CMO's 178-day school year[2]. First graders assigned to 1:1

[2] A previous version of this report included extended school year days. This version is adjusted to include only the traditional school year.

tutoring gained an additional 28.4% of the BAU control group's average gains in decoding from beginning to end of year, or an additional about 51 days of school.

Conclusion

This study presents the first rigorous evaluation of early literacy tutoring delivered completely virtually. We find positive and statistically significant effects from OnYourMark's model, even in just its second year of operation and while it expanded to serve more than seven times the number of students from the previous school year.

At the same time, we note that the positive effects produced by this virtual model are more modest than similar early literacy tutoring programs delivered in person (e.g., Cortes et al., 2023). With limited comparisons of virtual tutoring models, it is difficult to determine whether some of the difference in effect sizes may be due to differences in core programmatic features (e.g., virtual vs. online delivery) or a host of other differences related to program implementation, differences in measures, etc. We also note that features of the study design may lend themselves to more conservative estimates. Specifically, excluded from our analyses are the 121 students selected by their schools guaranteed to receive tutoring. These students are likely to be struggling students who would particularly benefit from tutoring, and thus our estimates may represent the lower bound of effects we might expect had they been included. Future evaluations of both virtual and in-person early literacy tutoring models will support stronger inferences about the relative effectiveness of virtual tutoring.

Our overall findings examining the effects of 1:1 tutoring and 2:1 tutoring are generally in line with the research base on the relative benefits of one-one-one instruction in small-group early literacy interventions (Nickow et al., 2020). However, our exploratory analyses of effects by grade and on discrete early literacy skills suggest that the effect of group size may depend on the grade of students and the skills being targeted.

The OnYourMark program represents a model for high impact tutoring with the potential to address some of the challenges associated with implementing high-quality, relationship-based personalized instruction at scale (e.g., Groom-Thomas et al, 2023), especially for contexts where the supply of in-person tutors is particularly constrained. Additional evaluations of the OnYourMark program across time and contexts, and of other innovative early literacy tutoring programs, will allow us to better understand how to increase equity in access to the most effective high-quality tutoring.

References

American Psychological Association. (2021). Inclusive language guidelines. https://www.apa.org/about/apa/equitydiversity-inclusion/language-guidelines.pdf

Carlana, M., & Ferrara, E. L. (2021). Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai21-350

Cortes, K., Kortecamp, K., Loeb, S., & Robinson, C. D. (2023). *A Scalable Approach to High-Impact Tutoring for Young Readers: Results of a Randomized Controlled Trial.* https://studentsupportaccelerator.org/sites/default/files/Scalable%20Approach%20to%20High-Impact%20Tutoring.pdf

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619. https://doi.org/10.1037/0022-0663.92.4.605

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-Analysis of the Impact of Reading Interventions for Students in the Primary Grades. *Journal of Research on Educational Effectiveness*, *13*(2), 401–427. https://doi.org/10.1080/19345747.2019.1689591

Groom-Thomas, L., Leung, C., Loeb, S., Pollard, C., Waymack, N., & White, S. (2023). Challenges and Solutions: Scaling Tutoring Programs. Inter-American Development Bank. http://dx.doi.org/10.18235/0005070

Gortazar, L., Hupkau, C., & Roldan, A. (2023). Online tutoring works: Experimental evidence from a program with vulnerable children. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai23-743

Kraft, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022). Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program. *AEA Papers and Proceedings*, *112*, 614–618. https://doi.org/10.1257/pandp.20221038

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools. *Reading Research Quarterly*, *57*(1), 149–179. https://doi.org/10.1002/rrq.379

Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://www.edworkingpapers.com/ai20-267

NWEA (2023, October). *MAP Reading Fluency.* Retrieved October, 2023 from https://www.nwea.org/map-reading-fluency/

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The Effectiveness of Volunteer Tutoring Programs for Elementary and Middle School Students: A Meta-Analysis. *Review of Educational Research*, *79*(1), 3–38. https://doi.org/10.3102/0034654308325690

Robinson, C. D., & Loeb, S. (2021). High-impact tutoring: State of the research and priorities for future learning.

Roschelle, J., Cheng, B. H., Hodkowski, N., Neisler, J., & Haldar, L. (2020). *Evaluation of an online tutoring program in elementary mathematics*. Digital Promise. https://files.eric.ed.gov/fulltext/ED604743.pdf

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1–26. https://doi.org/10.1016/j.edurev.2010.07.002

Torgerson, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., Jefferson, L., Mitchell, N., & Torgerson, D. (2016). *Affordable online maths tuition: Evaluation report and executive summary*. Education Endowment Foundation. https://durham-repository.worktribe.com/output/1606609

University of Oregon (2023). 8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Administration and Scoring Guide, 2023 Edition. Eugene, OR: University of Oregon. Available: https://dibels.uoregon.edu