NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Year Two Results Assessing the Effects of a Scalable Approach to High-Impact Tutoring for Young Readers

Kalena Cortes
Texas A&M University

Karen Kortecamp
George Washington University

Susanna Loeb
Stanford University

Carly D. Robinson
Stanford University

December 2023

studentsupportaccelerator.org

Year Two Results Assessing the Effects of a Scalable Approach to High-Impact Tutoring for Young Readers

Kalena Cortes
Texas A&M University

Karen Kortecamp
George Washington University

Susanna Loeb
Stanford University

Carly D. Robinson
Stanford University

Abstract

This research report presents the results from the second year of a randomized controlled trial of an early elementary reading tutoring program that has been designed to be affordable at scale. During the 2021-22 school year, over eight hundred kindergarten students in a large Southeastern school district were randomly assigned to receive supplementary tutoring with the Chapter One program. The program continued during the 2022-23 school year, while the children attended first grade. The program embeds part-time tutors into the classroom to provide short bursts of instruction to individual students each week over the course of the school year. The consistent presence of the tutors allows them to build strong relationships with students and meet students' individual needs at the moment they might most benefit from personalized instruction. The program focuses more time on students with the lowest literacy skills.

We find that students who participated in Chapter One's program increased their early literacy skills on both program-collected and district-collected measures. For example, students receiving Chapter One tutoring were nine percentage points less likely to be considered at risk in early literacy than the control group students (45 percent vs. 54 percent) on the district's winter assessment of early literacy. The positive findings at the end of the second year of implementation continue to provide promising evidence of an affordable and sustainable approach for delivering one-on-one personalized literacy tutoring at scale.

Introduction

The ability to read is highly determinative of outcomes later in life, predicting high school test scores (Sparks et al., 2014) and graduation rates (Hernandez, 2012). However, students nationwide struggle with this skill. Only a third of fourth graders performed at or above proficient in reading on the most recent National Assessment of Educational Progress (US Department of Education, 2022).

High-impact tutoring offers a potent strategy for improving early literacy skills in young students. Meta-analyses have shown tutoring to be highly effective in increasing general academic achievement (Dietrichson et al., 2017; Nickow et al., 2020) and to be effective in early literacy specifically with effect sizes ranging from 0.24 to 0.41 SD (Elbaum et al., 2000; Gersten et al., 2020; Neitzel et al., 2022). Reading tutoring interventions that provide students with one-on-one, personalized reading instruction consistently demonstrate the largest improvements in reading achievement (Cavanaugh et al., 2004; Gersten et al., 2020; Neitzel et al., 2022; Slavin et al., 2011; Wanzek et al., 2018; Wanzek et al., 2016). Specific programs may differ in delivery or approaches, but most effective reading tutoring programs involve students meeting for 20-60-minute sessions several times a week with a consistent educator and use evidence-based reading curricula (Galuschka et al., 2014; Wanzek et al., 2018; Wanzek et al., 2016). These features align with the definition of "high-impact" tutoring, which involves substantial time each week spent in required tutoring; sustained and strong relationships between students and their tutors; close monitoring of student knowledge and skills; alignment with school curriculum; and oversight of tutors to assure quality interactions (Robinson & Loeb, 2021).

High-impact tutoring programs drive the large effect sizes cited in the literature, but they can be hard to scale and require substantial resources to implement (Groom-Thomas, et al., 2023). Successful tutoring programs often require dedicated tutoring blocks within the school schedule and cost, at a minimum, over $1000 per student (e.g., Guryan et al., 2021; Sirinides et al., 2018). Given the large expected effect sizes, high-impact tutoring is quite cost-effective at improving student learning outcomes (Guryan et al., 2021). However, the urgent and growing demand for high-impact tutoring programs to build children's reading skills (U.S. Office of the Press Secretary, 2022) and common implementation issues (Carbonari et al., 2022), may prompt district leaders to search for even lower-cost programs that fit within existing school schedules.

In a prior report, we presented results from a randomized controlled trial of an early elementary reading tutoring program designed to be affordable at scale. During the 2021-22 school year, over eight hundred kindergarten students in a large Southeastern school district were randomly assigned to receive supplementary tutoring with the Chapter One program. The program embeds part-time tutors into the classroom to provide short bursts of instruction to individual students each week over the course of the school year. The consistent presence of the tutors allows them to build strong relationships with students and meet students' individual needs at the moment they might most benefit from personalized instruction. We found that students who participated in Chapter One's program were more likely to reach the target reading stage by the end of kindergarten and showed higher performance on a measure of oral reading fluency collected by the program, as well as three

tenths of a standard deviation higher on the district reading assessment (Cortes et. al., 2023).

Students in the program who remained at the same schools continued with the program during their first-grade year. The tutors focused most of their time on the students in the treatment group with lower reading skills but all of the students in the program continued to receive at least some tutoring. While differential attrition of higher achieving students in the treatment group for the program-collected outcomes resulted in a need to adjust for baseline scores in the analyses for those outcomes, we continue to find clear positive effects of the Chapter One program on student outcomes, including a 35 percent of a standard deviation increase in oral reading fluency and a 16 percent reduction in being classified as at risk on the district literacy exam given in the winter.

The positive findings from the second year of implementation provided promising evidence of an affordable and sustainable approach for delivering one-on-one personalized reading tutoring at scale.

Leveraging close relationships and technology to support early readers

This evaluation of the Chapter One program is among the first to provide evidence that early elementary students can benefit from frequent, short bursts of reading instruction from consistent tutors embedded in the classroom. The program leverages technology and the close relationship tutors build with their students to personalize instruction, dosage, and session length to meet the individual needs of each child to develop a strong foundation in phonics and build reading fluency.

Chapter One uses a "push-in" model that provides districts with part-time tutors, or Early Literacy Interventionists (ELIs), who meet with students one-on-one in the back of the classroom over the course of a school year. One ELI serves multiple classrooms in the school and tutors individual students in 5-7 minute increment sessions during blocks of reading instruction or other opportune moments. At the end of each session, the departing student brings the next student to the ELI to minimize interruptions of classroom instruction.

These short sessions account for young students' short attention spans and allow for each session to focus on a progression of discrete skills (Ehri et al., 2001). Specifically, students progress through stages of phonics development, learning to segment and blend short and long vowel sounds, learning sight words, and learning strategies to fluently read both decodable and noncontrolled texts. The curriculum draws on a strong evidence base on teaching young children to read (Ehri et al., 2001) and is designed to match learning and instruction with a child's developmental level (Vygotsky, 1980). The length of each session and the number of sessions per week vary for each student based on need and rate of progress. For instance, students who are making adequate progress may only meet with their tutor once or twice a week, whereas students who the tutors identify as in need of more support may meet daily for periods of time.

To provide this tailored support, the Chapter One program leverages technology to support instruction, as well as to direct student independent practice. ELIs follow a digital curriculum to conduct each session, which facilitates the assessment and tracking of student performance over time. In addition to

using the technology in one-on-one sessions, students spend 15 minutes each day independently practicing using Chapter One's software on program-provided tablets. All assessments sync in real time with individual student tablets, so that when a student uses the practice software after the one-on-one session, they practice items that are precisely aligned to their most recent tutored instruction. ELIs also regularly meet with teachers, reading coaches, and principals to review online reports of student progress.

The structured curriculum and technological support allow for a wide range of people to serve as ELIs. Some ELIs are former classroom teachers, however, most do not have a teaching certification. All ELIs have earned at least a Bachelor's degree and undergo an extensive series of online training courses with associated assessments that they must pass to proceed in the training plan. ELIs are compensated substantially above minimum wage and receive ongoing support and development.

The program currently costs school districts $375 per student, which includes the ELI, student technology (tablets - Kindle Fires), background check, training time, Chromebook for the ELI, reinforcement materials for the ELI vetted to align with the model, and indirect costs for implementing the program. In implementations that involve over 5,000 students, the district is also asked to fund the cost of district-wide managers, which increases the cost per student to approximately $450. Even in large implementations, this cost is substantially lower than the vast majority of other tutoring programs and does not require districts to coordinate complicated logistical arrangements.

Methods

Study Details

During the 2021-22 and 2022-23 school years, Chapter One partnered with a large school district in the Southeastern US to conduct a randomized controlled trial of the program with early elementary students. Fifty-six percent of students' families in the district qualify for free and reduced priced lunch (FRPL) and 12.6 percent of students are English Learners (ELs).

The district identified 49 kindergarten classes across 13 schools to participate in the evaluation. Tutoring by Chapter One started in early November 2021 and was rolled out to all participating classrooms over the course of the next two months. The first year of the program lasted through the end of the kindergarten school year, in May 2022. Students who remained in the same schools in the district then receive Chapter One tutoring in first grade during the 2022-23 school year.

Our evaluation will explore the effect of receiving Chapter One tutoring in kindergarten and first grade on reading proficiency through early elementary school. In this report, we present the results from the second year of the study, in which we assess the impact of Chapter One tutoring on first grade students' reading development. Specifically, we ask whether receiving Chapter One tutoring in kindergarten increases students' Reading Foundation Stage, oral reading fluency, winter literacy

Table 1. Sample Descriptive Statistics and Balance Test

	Overall			Treatment			Control					
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Diff	SE	
Student Demographics:												
White	0.04		818	0.02		420	0.07		398	-0.05	0.01	***
Black	0.72		818	0.73		420	0.71		398	0.02	0.03	
Hispanic	0.21		818	0.22		420	0.19		398	0.02	0.02	
Other Race	0.03		818	0.03		420	0.03		398	0.00	0.01	
Female	0.47		818	0.50		420	0.44		398	0.07	0.04	+
English-Language Learner	0.28		818	0.31		420	0.25		398	0.06	0.02	*
Special Education	0.11		818	0.11		420	0.11		398	0.00	0.02	
Baseline Achievement												
FLKRS Scaled Score	452.75	93.78	739	445.31	90.63	381	460.66	96.52	358	-15.24	6.59	*
Indicator for Missing	0.10		818	0.09		420	0.10		398	-0.01	0.02	

Notes: FLKRS is a screening instrument, known as the Florida Kindergarten Readiness Screener (FLKRS), that must be administered to all public school kindergarten students within the first 30 days of each school year. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.10$.

score and spring reading score. The first of these (Reading Foundation Stage) measures the specific goal of Chapter One to develop a strong foundation in early phonics. The second (oral reading fluency) is a key determinant of reading achievement, and requires face-to-face administration. Level 9 or 10 is the typical goal for first grade readers. The other two measures are conducted by the district using an online assessment tool that cannot measure oral reading fluency directly and instead measures subskills including word knowledge.

Sample and Randomization

The study consisted of 818 kindergarten students in 13 schools. Panel A of Table 1 provides information on the demographics of the students in the RCT sample. We conducted a student-level randomization stratified by classroom. Specifically, within each kindergarten classroom (N = 49), we randomly assigned 50 percent of the students to the treatment group (i.e., to receive Chapter One tutoring; N = 420) and 50 percent to the control group (i.e., to receive business-as-usual instruction; N = 398).

Data

We collected administrative data from the school district and Chapter One, including data on gender, race/ethnicity, English language learner indicators, and whether students qualify for special education services. As a measure for baseline reading skill, we use the district's administration of the Florida Kindergarten Readiness Screener (FLKRS), which was the Renaissance Star Early Literacy measure in Fall of 2021. The FLKRS was administered to all public school kindergarten students within the first 30 days of each school year. The literacy classifications for the scores are as follows: Early Emergent Reader (300 - 487), Late Emergent Reader (488 - 674), Transitional Reader (675 - 774), and Probable Reader (775 - 900). Chapter One served Early Emergent and Late Emergent readers and both the treatment and controlled groups are largely limited to these students, though a few (20) higher scoring students were included.

The goal of this study is to measure the effects of the Chapter One program on students' early literacy development. We have four measures of this outcome, each with advantages and disadvantages. The first measure is the Reading Foundation Stages at the end of their first grade year. Chapter One follows a child's progression through six Reading Foundation Stages. Upon mastering the Reading Foundation Stages, students continue to work with ELIs to practice oral reading and adaptive phonics content. The goal of the program is for students to achieve Reading Foundations Stage 5 by the end of first grade. A student who achieves Reading Stage 4 has a beginning competency in decoding consonant-vowel-consonant (CVC) words, while a student who achieves Reading Stage 5 also has a beginning competency in decoding CVCe words that have a silent e on the end. The main drawback of this measure is that it is collected by the ELIs and could be subject to ELI bias in favor of the students with whom they work most closely.

The program also collects a standardized Oral Reading Fluency (ORF) measure, in addition to the Reading Stages, which, while also collected by ELIs, is less subject to ELI influence. The measure is very similar to DIBELS. To administer the assessment, the ELI begins by explaining that the student is to do

their best reading of the passage aloud. If they are stuck on a word, the ELI would tell them the word so that they could keep reading. After checking that the student understands the directions, the ELI reads the title, then points to the first word and says "Begin". The ELI starts the timer when the student says the first word of the passage. If the student fails to say the first word after 3 seconds, the ELI tells them the word and marks it incorrect and starts the timer. The maximum time for each word is 3 seconds. At the end of 1 minute the ELI notes the last word read by marking it with a bracket. If the stop time falls mid-sentence the ELI allows the student to complete reading the sentence but does not record scores for any words read beyond the stop bracket. The ELI records the words correct per minute and calculates accuracy by dividing the words correct per minute by the total words read and multiplying by 100. ELIs are told that they must follow the same script for every assessment without variation. ORF scores are likely to have floor effects for kindergarten students, but it is a better measure for first grade students who are more likely to have the necessary reading skills. The goal for first grade is for students to reach a certain number of words correct per minute, usually measured by having reached Level 9 (I) or Level 10 (J).

The final two outcome measures are the district early literacy and reading assessments. Many schools in the district give the Star Early Literacy assessment in both the fall and winter of the first grade year and the Star Reading assessment in the spring. However, not all schools give these assessments, and, as a result, we are limited to the students in those schools. The Star Early Literacy assessment takes 10 to 20 minutes, includes 27 items and sets a time limit of 90 seconds per item. It covers alphabetic principle, concept of word, visual discrimination, phonemic awareness, phonics, vocabulary, and early numeracy. The Star Reading assessment takes 15 to 20 minutes, includes 34 items and allows 60 seconds per item for the first 10 items and 120 seconds per item for the remaining items. This assessment is designed to measure reading comprehension skills including identifying main ideas and details, making inferences, understanding sequence, and understanding words and phrases in context. It also measures a student's reading proficiency, fluency, and vocabulary. Because Chapter One focuses on students who begin kindergarten without strong early literacy skills, the early literacy assessment measures skills that are more relevant for the study students than does the reading assessment.

Attrition

Approximately 33 percent of the original sample did not receive tutoring from Chapter One during their first grade year. Table 2 describes this attrition. Not only do we see overall attrition, the attrition differs between the treatment and control groups. Twelve percent of the control group and eight percent of the treatment group, a difference that is statistically significant, left the district. Twenty six percent of the treatment and 22 percent of the control group left the program, largely because they switched schools, but stayed in the district.

The characteristics of students attrited also differed across treatment and control. Overall Black students, Hispanic students, students classified as English learners, and students with higher incoming assessment scores at the beginning of kindergarten were less likely to leave the sample. Differentially, Hispanic students in the treatment group were even less likely to leave than Hispanic students in the control group and students with lower initial reading scores were less likely to leave the treatment

group. This differential attrition by prior kindergarten scores is particularly concerning because of the strong relationship across measures of reading over time.

To address this differential attrition, we adjust for prior scores in our analyses. For the outcome measures that are district wide, differential attrition by treatment status is less of a concern because while higher performing students are less likely to leave the district, we see no detectable difference between treatment and control in the relationship between prior scores and exiting the district.

Table 2: Sample Attrition

	Overall Attrition		District Attrition		Program Attrition	
	Students who left the study for any reason		Students who are no longer enrolled in the district		Students who left the study but remained in the district	
	(1)	(2)	(1)	(2)	(1)	(2)
Treatment	0.00139	0.320*	-0.0406*	-0.0444	0.0420	0.365**
	(0.0360)	(0.169)	(0.0229)	(0.105)	(0.0290)	(0.163)
Black		-0.220*		-0.0215		-0.199*
		(0.121)		(0.0732)		(0.105)
Hispanic		-0.234**		-0.0329		-0.202*
		(0.114)		(0.0804)		(0.108)
Other: Race		0.0354		-0.0419		0.0773
		(0.190)		(0.0798)		(0.166)
Fall K Assessment		-0.0791***		-0.0493***		-0.0298**
		(0.0163)		(0.0125)		(0.0130)
Imputed Assessment		0.355***		0.125*		0.230***
		(0.0816)		(0.0630)		(0.0781)
Black x Treatment		-0.248		0.0268		-0.275*
		(0.173)		(0.107)		(0.159)
Hispanic x Treatment		-0.313*		-0.0181		-0.295*
		(0.172)		(0.110)		(0.155)
Other x Treatment		-0.601**		-0.0425		-0.559**
		(0.249)		(0.113)		(0.222)
Assessment x Treatment		0.0637*		0.0267		0.0370
		(0.0343)		(0.0201)		(0.0308)
Imputed x Treatment		-0.228*		-0.0212		-0.207**
		(0.124)		(0.0979)		(0.0954)
Constant	0.337***	0.521***	0.122***	0.137**	0.215***	0.383***
	(0.0286)	(0.117)	(0.0184)	(0.0666)	(0.0227)	(0.109)

Clustered standard errors at the classroom level are shown in parentheses. Model 2 also includes indicators for gender, English learner status, special education status, and the interactions of these with treatment status. None of the coefficients on these measures are statistically different from zero. *** p<0.01, ** p<0.05, * p<0.1

Analysis

We preregistered our study design, hypotheses, and analytic plan (see: https://www.socialscienceregistry.org/trials/10810/history/169581) on the Social Science Registry prior to conducting the primary analysis. We use the following model to evaluate the difference between the treatment and control groups:

$$Y_{ijk} = \alpha + \beta_1 Treatment_i + \beta_2 R_i + \pi X_i + \varepsilon$$

where Y is the outcome for student i in classroom j in school k; $Treatment_i$ is an indicator whether student i was assigned to Chapter One; R_i is a student's beginning of the year score (included as a control for baseline achievement); X_i is a vector of student characteristics (i.e., indicators for gender, race, English Language Learners, special education) and ε is an error term.

We also conduct exploratory analyses that study the heterogeneity of the treatment effects by pre-intervention characteristics. Specifically, we look at outcomes for students with different reading skills at the beginning of kindergarten, and we explore whether the program differentially impacts English language learners and native English speakers.

Because we have program-collected outcomes only for those students who remained in the same schools, we have to limit our sample to this group and adjust for differences in baseline characteristics. For the district outcomes, we can include students who were not in the participating schools in first grade but remained in the district in schools administering the assessments, though we still cannot include students who left the district. These additional students reduce the starting differences between the treatment and control students but do not eliminate them. As a result, we control for baseline characteristics in our preferred models.

Results

## Reading Foundation Stage Results

We present the results for the Reading Foundations Stages in Table 3. We see in panel two that the treatment group began the year approximately one half stage higher than the control group after having been part of the treatment in kindergarten. By the end of the year, as shown by panel one, the treatment group was a whole stage ahead of the control group. Figure 1 compares the percent of treatment and control group achieving Stages four and five. While 76 and 60 percent of the control group achieved Stage 4 and 5, respectively, 96 and 88 percent of the treatment did. Controlling for initial kindergarten scores, these differences are a little larger, two percentage points for Stage 4 and four percentage points for Stage 5.

Table 3. The Effect of Chapter One on Reading Foundation Stages and Oral Language Fluency

	Reading Stage End of Year		Reading Stage Beginning of Year		ORF End of Year		ORF Beginning of Year	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Treatment	0.913***	1.040***	0.373***	0.562***	0.229**	0.337***	-0.134	0.0346
	(0.118)	(0.105)	(0.109)	(0.0854)	(0.0908)	(0.0748)	(0.0962)	(0.0799)
Female		0.0312		0.0745		0.104		0.00444
		(0.0820)		(0.0849)		(0.0779)		(0.0746)
Black		-0.0751		0.335		0.226		0.179
		(0.273)		(0.264)		(0.212)		(0.182)
Hispanic		-0.0314		0.304		0.160		0.0436
		(0.297)		(0.300)		(0.233)		(0.194)
Other: Race		0.158		0.520		0.586**		0.415
		(0.307)		(0.398)		(0.267)		(0.297)
English Learner		-0.0615		-0.225*		0.0530		-0.157**
		(0.130)		(0.132)		(0.103)		(0.0762)
Special Education		-0.277		-0.151		-0.203		-0.0326
		(0.177)		(0.195)		(0.146)		(0.111)
Fall K Assessment		0.426***		0.660***		0.464***		0.542***
		(0.0405)		(0.0547)		(0.0325)		(0.0296)
Imputed Assessment		-0.0568		0.00703		0.0264		-0.101
		(0.166)		(0.183)		(0.174)		(0.143)
Constant	4.637***	4.643***	3.462***	3.053***	-0.114	-0.453**	0.0745	-0.138
	(0.0988)	(0.277)	(0.110)	(0.291)	(0.0704)	(0.225)	(0.0769)	(0.188)
Observations	540	540	540	540	540	540	534	534
R-squared	0.147	0.303	0.020	0.325	0.013	0.263	0.004	0.357
Control Group Avg.	4.637	4.637	3.462	3.462	-0.114	-0.114	0.074	0.074

Clustered standard errors at the classroom level are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Figure 1. The likelihood of achieving Reading Foundation Stage 4 and 5 by the end of first grade by condition assignment (%)
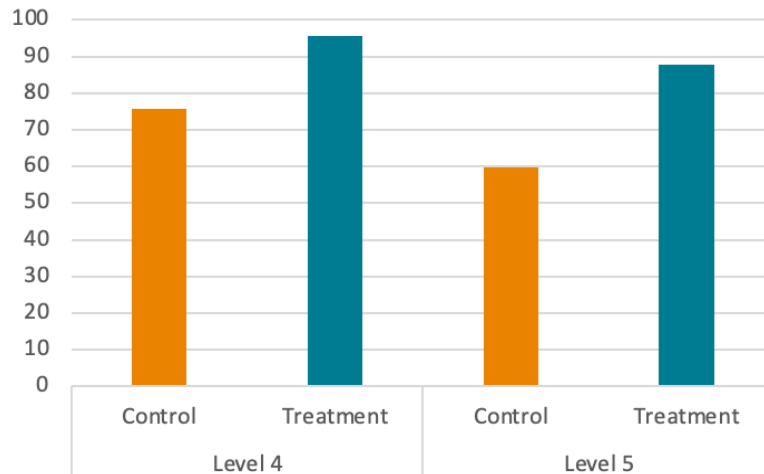
Oral Language Fluency Results

Table 3 also provides the results for the Oral Language Fluency (ORF) assessments, standardized to have a mean of zero and standard deviation of one. At the beginning of the year, the treatment and control students showed no difference in oral language fluency, as shown in the fourth panel. In fact, due to selective attrition of higher scoring treatment students, the students in the treatment sample scored 13 percent (not statistically different from zero) of a standard deviation lower than the control students before adjusting for differences in kindergarten entry score. However, by the end of the school year, as shown in the third panel, students receiving Chapter One tutoring scored significantly higher on the oral fluency assessment than did students in the control group by 35 percent of a standard deviation, controlling for baseline characteristics (24 percent higher without controls). Figure 2 shows the complete distributions of ORF scores for treated and control students both at the beginning of the year and at the end of the year. Approximately 48 percent of the control students reached Level 9 or higher at the end of the school year, while 62 percent of the treated students did–a difference of 14 percentage points gain for treated students, even though they started the year with lower scores. Similarly, approximately 46 percent of the control students reached Level 10 or higher at the end of the school year, while 55 percent of the treated students did–a difference of 9 percentage points.

Figure 2. Oral Reading Fluency Scores at the Beginning and End of the School Year by Treatment Status (%)
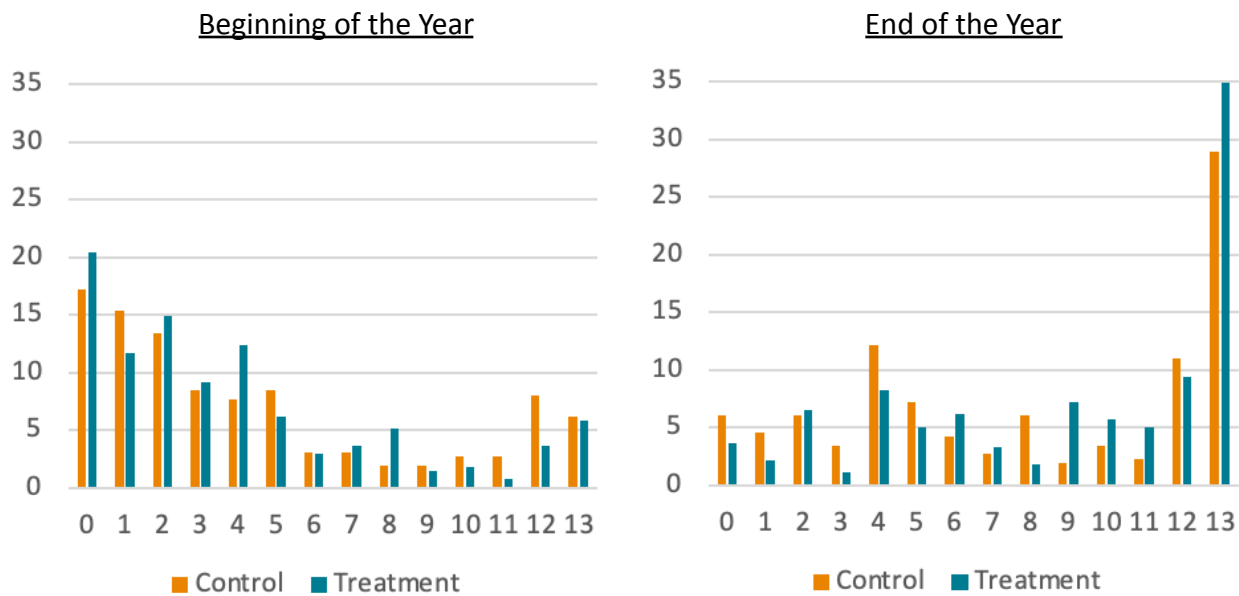
Results for District Assessments of Early Literacy and Reading

Table 4 describes the results for the district's early literacy assessment. In order to reduce the potential for differential attrition from the study schools between the treatment and control groups leading to bias, we include students who left the study schools but remained in the district and took the district

assessments, approximately 15 percent of the sample. This intent-to-treat analysis may underestimate the true effect of the program on those who receive the treatment. Similar to the results for the oral fluency measure, we see no differences at the beginning of the school year in the early literacy score between the treatment and control groups when we adjust for baseline scores. However, due to differential attrition even with the expanded sample, the treatment group scored significantly lower than the control group (14 percent of a standard deviation) without controls, as shown in the third panel.

Table 4. The Effect of Chapter One's Tutoring Program on District Early Literacy Assessments

	Winter Score (1)	Winter Score (2)	Winter At Risk (1)	Winter At Risk (2)	Fall Score (1)	Fall Score (2)	Fall At Risk (1)	Fall At Risk (2)
Treatment	0.0204	0.112	-0.0442	-0.0891**	-0.145*	-0.0165	0.0447	-0.00495
	(0.0841)	(0.0742)	(0.0373)	(0.0335)	(0.0741)	(0.0664)	(0.0391)	(0.0366)
Female		0.138*		0.00571		0.0859		-0.0632
		(0.0751)		(0.0394)		(0.0692)		(0.0382)
Black		0.0642		-0.0525		-0.0833		-0.00897
		(0.188)		(0.119)		(0.188)		(0.109)
Hispanic		0.0398		-0.0428		-0.116		-0.0249
		(0.209)		(0.120)		(0.207)		(0.124)
Other: Race		-0.0922		0.00936		0.0662		0.0639
		(0.261)		(0.167)		(0.242)		(0.165)
English Learner		-0.153*		0.0892*		-0.137*		0.0715
		(0.0843)		(0.0459)		(0.0729)		(0.0454)
Special Education		-0.218		0.123*		-0.206*		0.0452
		(0.166)		(0.0640)		(0.114)		(0.0599)
Fall K Assessment		0.575***		-0.245***		0.622***		-0.249***
		(0.0330)		(0.0190)		(0.0273)		(0.0158)
Imputed Assessment		-0.107		0.0474		-0.212		0.117**
		(0.176)		(0.0832)		(0.143)		(0.0582)
Constant	-0.00885	-0.0676	0.541***	0.551***	0.0804	0.109	0.576***	0.618***
	(0.0698)	(0.190)	(0.0314)	(0.119)	(0.0693)	(0.191)	(0.0369)	(0.107)
Observations	562	562	562	562	601	601	601	601
R-squared	0.000	0.345	0.002	0.251	0.005	0.457	0.002	0.306
Control Group Avg.	-0.009	-0.009	0.541	0.541	0.080	0.080	0.576	0.576

Clustered standard errors at the classroom level **are** shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The early literacy scores from the winter assessments show more rapid gains for students who participated in Chapter One tutoring. While the difference is not statistically significant for the average score, the treatment students scored 11 percent of standard deviation higher than the control students. More concretely, students receiving Chapter One tutoring were nine percentage points (45

percent) less likely to be considered at risk in early literacy than the control group students (54 percent) and this difference is statistically significant at traditional levels, as shown in the second panel of Table 4. If we look only at students who remained in the study schools, we see an 11.5 percentage point reduction in being at risk in early literacy, controlling for baseline characteristics (42 percent versus 54 percent).

Table 5 provides similar results for the district reading exam, given by most schools in the spring of the first grade year. We find no significant differences between the treatment and control on this measure, though five percentage point fewer treatment students (42 percent) are classified at risk than the control group (47 percent) and they scored ten percent of standard deviation higher than treatment, on average.

Table 5. The Effect of Chapter One's Tutoring Program on District Reading Assessments

	Spring Scores		Spring At Risk	
	(1)	(2)	(1)	(2)
Treatment	0.0193	0.0973	-0.0214	-0.0518
	(0.0851)	(0.0754)	(0.0405)	(0.0355)
Female		0.153*		-0.102**
		(0.0774)		(0.0418)
Black		0.234		-0.0722
		(0.197)		(0.0879)
Hispanic		0.174		-0.0836
		(0.214)		(0.0918)
Other: Race		0.324		-0.180
		(0.278)		(0.125)
English Learner		0.0344		0.000636
		(0.0963)		(0.0434)
Special Education		-0.0265		0.00177
		(0.138)		(0.0613)
Fall K Assessment		0.523***		-0.223***
		(0.0291)		(0.0141)
Imputed Assessment		-0.0167		-0.0221
		(0.174)		(0.0756)
Constant	-0.0105	-0.376*	0.471***	0.626***
	(0.0717)	(0.195)	(0.0340)	(0.0935)
Observations	617	617	617	617
R-squared	0.000	0.301	0.000	0.234
Control Group Avg.	-0.011	-0.011	0.471	0.471

Clustered standard errors at the classroom level are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Heterogeneity Analysis

We conduct a heterogeneity analysis to understand the extent to which students' baseline reading abilities impacted the estimates of Chapter One effects. Table 6 shows the effect Chapter One had on kindergarten students who began kindergarten with scores in the bottom half of the study sample compared with those who started in the top half of the sample. The sample overall all scored in the lower part of the early literacy distribution; almost all participants in the study were classified initially as either Early Emergent and Late Emergent Readers.

We see results not only based on starting reading scores but also on the assessment. The reading stages assessment spans the range of readers in the first grade, and we see similar estimating effects, with both the lower half and the upper half of readers in the treatment group gaining approximately one reading stage more than those in the control group (1.17 stages vs. 0.95 stages). The lower scoring students saw a larger boost in the percent of students reaching Stage 5 or above (38.8 vs 26.4 percentage point increases) but that is explained at least in part by the high percentage of students reaching Stage 5 by all students who started in the top half of the distributions. The oral reading fluency instrument measures more advanced skills and here we see greater gains for those starting in the top half of the reading distributions (46 percent of a standard deviation gain, relative to 23 percent for initially lower performing students).

Table 7 shows similar results comparing students classified as English learners with students who are not classified as an English learner. We see similar estimates of the effects between the two groups, with a bit greater gain for English learners in reaching Reading Stage 5, in part because of the lower proportion of the control group that reaches that stage.

Table 6. Heterogeneity of Effects by Initial Reading Performance on Program Outcome Measures

	Reading Stage		Reading Stage 5		Oral Reading Fluency		Oral Reading Fluency Level 9	
	Lower Half	Upper Half	Lower Half	Upper Half	Lower Half	Upper Half	Lower Half	Upper Half
Treatment	1.174***	0.932***	0.388***	0.255***	0.231*	0.440***	0.134**	0.248***
	(0.171)	(0.103)	(0.0590)	(0.0390)	(0.118)	(0.0985)	(0.0557)	(0.0459)
Observations	269	271	269	271	269	271	269	271
R-squared	0.250	0.323	0.194	0.197	0.131	0.232	0.077	0.202
Control Group Avg.	4.167	5.035	0.417	0.754	-0.486	0.201	0.333	0.599

Clustered standard errors at the classroom level are shown in parentheses. All models include controls for gender, race/ethnicity, English learner status, special education status, kinder-entry test score, and an indicator of missing scores. *** p<0.01, ** p<0.05, * p<0.1

Table 7. Heterogeneity of Effects by English Learner Status on Program Outcome Measures

	Reading Stage		Reading Stage 5		Oral Reading Fluency		Oral Reading Fluency Level 9	
	English Learner	Non	English Learner	Non	English Learner	Non	English Learner	Non
Treatment	1.130***	1.036***	0.417***	0.291***	0.355**	0.349***	0.227***	0.188***
	(0.210)	(0.114)	(0.0694)	(0.0388)	(0.133)	(0.0856)	(0.0674)	(0.0365)
Observations	163	377	163	377	163	377	163	377
R-squared	0.266	0.336	0.254	0.235	0.183	0.296	0.145	0.225
Control Group Avg.	4.318	4.745	0.470	0.643	-0.352	-0.033	0.364	0.515

Clustered standard errors at the classroom level are shown in parentheses. All models include controls for gender, race/ethnicity, English learner status, special education status, kinder-entry test score, and an indicator of missing scores. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Conclusion

Early readers continue to see a positive effect of Chapter One tutoring during their first grade year. In the first year of this study, when students were in kindergarten, we saw positive effects on program-collected measures of Reading Stages and Oral Reading Fluency. In this second year of the study, we continue to see positive effects on these measures, as well as on the district measure of early literacy skills - the STAR Early Literacy Assessment. Sample attrition, particularly of higher-scoring students in the treatment group, led to the need to adjust for entry scores in kindergarten.

The Chapter One approach is benefitting students in our partner district. The combination of short bursts of 1:1 instruction by trained staff, with independent practice on digital devices precisely synched to the 1:1 instruction, delivers a program that is affordable and scalable. The program is also likely to be less obtrusive to classroom instruction than tutoring programs that pull out students for greater amounts of time.

We find that implementing this program in kindergarten and first grade can meaningfully improve the literacy ability of students through the end of first grade. We will continue to track students' progress through third grade, but the results from the first two years of the evaluation are encouraging. Given the low-cost of the program and the ease of incorporating the program into the school day, using classroom-based tutors to deliver short bursts of reading instruction, supported by technology that helps tutors address each student's specific needs, may be a promising approach for making early reading tutoring programs sustainable and affordable.

References

Carbonari, M. V., Davison, M., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Hashim, A., Kane, T. J., McEachin, A., & Morton, E. (2022). The Challenges of Implementing Academic COVID Recovery Interventions: Evidence from the Road to Recovery Project.

Cavanaugh, C. L., Kim, A.-H., Wanzek, J., & Vaughn, S. (2004). Kindergarten Reading Interventions for At-Risk Students: Twenty Years of Research. *Learning Disabilities: A Contemporary Journal, 2*(1), 9-21.

Cortes, K., Kortecamp, K., Loeb, S, and Robinson, C. (2023). *A Scalable Approach to High-Impact Tutoring for Young Readers: Results of a Randomized Controlled Trial*. National Student Support Accelerator

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

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393-447.

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

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PloS one, 9*(2), e89900.

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.

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

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., Farkas, G., Fryer Jr, R. G., Mayer, S., & Pollack, H. (2021). *Not too late: Improving academic outcomes among adolescents*.

Hernandez, D.J. (2012). *Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation*. The Annie E. Casey Foundation.

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.

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 Series*(w27476).

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

Sirinides, P., Gray, A., & May, H. (2018). The Impacts of Reading Recovery at scale: Results from the 4-year i3 external evaluation. *Educational evaluation and policy analysis, 40*(3), 316-335.

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.

Sparks, R.L., Patton, J. & Murdoch, A. (2014). Early reading success and its relationship to reading achievement and reading volume: replication of '10 years later'. Read Writ 27, 189–211. https://doi.org/10.1007/s11145-013-9439-2

U.S. Department of Education. (2022). *National Assessment of Educational Progress (NAEP) 2022 Long-Term Trend Assessment Results: Reading and Mathematics*. https://www.nationsreportcard.gov/highlights/ltt/2022/

U.S. Office of the Press Secretary. (2022, June 5, 2022). *FACT SHEET: Biden-Harris Administration Launches National Effort to Support Student Success* https://www.whitehouse.gov/briefing-room/statements-releases/2022/07/05/fact-sheet-biden-harris-administration-launches-national-effort-to-support- student-success/?utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current Evidence on the Effects of Intensive Early Reading Interventions. *Journal of Learning Disabilities, 51*(6), 612-624. https://doi.org/10.1177/0022219418775110

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28*, 551-576.