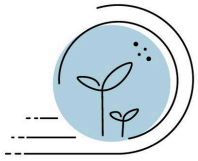


NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Hoot Reading & Kansas City Public Schools: 2024-2025 RCT Report

Hsiaolin Hsieh
Stanford University

Monica G. Lee
Stanford University

Carly D. Robinson
Stanford University

Susanna Loeb
Stanford University

October 2025

studentsupportaccelerator.org

Stanford | SCALE Initiative
Accelerator for Learning

Hoot Reading & Kansas City Public Schools: 2024-2025 RCT Report

Summary

This report evaluates the effectiveness of teacher-led virtual tutoring for early literacy through a large-scale randomized controlled trial conducted in partnership with Kansas City Public Schools (KCPS). The intervention, implemented within the Multi-Tiered System of Support (MTSS) framework, aims to provide a scalable and sustainable approach to increasing access to personalized learning.

The study analyzes data from 1,550 students across 14 elementary schools, focusing on grades 1–4. Findings show that the program had a positive and statistically significant effect, particularly for students who scored well below grade-level benchmarks on beginning-of-year assessments (MTSS Tier III students). Tier III students in the treatment group outperformed their peers in the comparison group by 0.08 standard deviations in end-of-year i-Ready Reading Overall Scale Scores. Among Tier III students who were assigned to and actively participated in tutoring, the effect was slightly larger (0.09 SD). Both effects were statistically significant at the $p < .05$ level.

The intervention also improved other measures of learning. Students randomly assigned to the treatment group demonstrated significantly higher Annual Typical Growth (+10.84 percentage points) and Annual Stretch Growth (+5.24 percentage points) on i-Ready reading growth metrics. For third- and fourth-grade students who took the Missouri Assessment Program (MAP) English Language Arts test, estimated impacts were of similar magnitude to those found for i-Ready, though not statistically significant.

Overall, integrating high-impact tutoring into the MTSS framework significantly enhanced reading outcomes for struggling readers. The findings suggest that schools can reduce disparities in early literacy and improve student learning at scale through structured, technology-supported tutoring without substantial additional cost.

Study Design

In the 2024-2025 school year, Hoot Reading (Hoot), a virtual tutoring provider, partnered with Kansas City Public Schools (KCPS) in Kansas City, MO, to deliver high-impact tutoring focused on early literacy skills for elementary school students in grades one to four. This initiative is part of the city-wide Literacy for All Students (LFAS) Campaign, aimed at improving reading growth for students in grades K-3, and is funded by SchoolSmartKC.

A total of 16 elementary schools within KCPS participated in the program. The number of seats available for Hoot tutoring varied by school, ranging from 40 to 220, for a total of 1,188 seats. Enrollment prioritized MTSS-Tier III students who scored well below (two or more levels below) grade-level benchmarks on the beginning-of-year (BOY) i-Ready reading assessment. Any unfilled seats were then allocated to Tier II students, who were one level below grade-level benchmarks.

We conducted a student-level randomization process, taking into account school, grade level, and MTSS Tier. For each school, based on the number of allocated seats, we assigned half of the eligible students to the Treatment group and the other half to the Comparison (control) group. Students in the Treatment group received one-on-one virtual tutoring from Hoot for 30 minutes, at least three times a week, over a period of 20 weeks, from late October/early November 2024 to early May 2025. Meanwhile, students in the Comparison group continued to receive the standard instructional support they would have had without Hoot tutoring.[1]

To maintain the integrity of the control group despite attrition from the treatment group, we randomly assigned 10 students per school to a waitlist and provided schools with a protocol for backfilling tutoring seats if attrition occurred. To maintain the value of the randomization, we do not include the waitlist students in the analyses, while we do include the Treatment students regardless of whether they attrited from the program.

During randomization, two schools did not have enough eligible students to evenly assign to the Treatment and Comparison groups. Consequently, all eligible students from those schools were assigned to the Treatment group (with no Comparison group), and those schools were excluded from the study. Overall, this study included 1,550 students across 14 elementary schools, with 707 students in the Treatment group, 843 in the Comparison group, plus 140 on the waitlist. More students are in

[1] Various district-wide reading interventions sponsored by the Literacy for All Students Campaign occurred simultaneously during the study period. See Table 1.

the Comparison group because the number of available seats for tutoring was limited. Table 1 provides the breakdown of random assignments by MTSS tier and school.

Table 1. Breakdown of Random Assignment by MTSS Tier and School

	Tier-II		Tier-III		
School	**Comparison**	**Treatment**	**Comparison**	**Treatment**	**All**
ACPrep	30	19	21	21	91
Banneker[1]	22	71	0	0	93
Carver	19	6	33	34	92
Garcia[2]	64	13	66	67	210
Garfield[2]	0	0	78	80	158
James[1]	6	20	0	0	26
JARogers	0	0	96	40	136
King	0	0	118	80	198
Melcher	0	0	43	40	83
Phillips	32	17	23	23	95
Pitcher	10	27	13	13	63
Trailwoods	0	0	43	40	83
Wheatley[2,3]	30	25	54	55	164
Whittier[1]	42	16	0	0	58
Subtotal	**255**	**214**	**588**	**493**	**1550**

Note. Superscripts next to school names are used to annotate different known interventions that are part of the district-wide Literacy for All Students Campaign and occurred simultaneously during the study period.[1]Literacy for All Students (LFAS); top-down; principal training, literacy coach provided; [2]Hall Family Schools: intervention = Hoot, control group received BAU - likely that they received some kind of literacy intervention; [3]PlaBook- AI literacy.

Data

Student-Level Administrative and Tutoring Data

We compile a student-level dataset using administrative data and tutoring session data from KCPS and Hoot. This dataset includes demographic information such as grade, gender, race, ethnicity, designation as an English Learner (EL) or as eligible for Special Education (SPED) services, and enrollment in the Supplemental Nutrition Assistance Program (SNAP). It also contains performance data from assessments, including test scores from the i-Ready Reading Diagnostics and the Missouri Assessment Program (MAP), as well as information on tutoring attendance. Hoot provided session-level metadata from their platform, including the number of sessions scheduled and attended for each student, allowing the research team to analyze tutoring participation.

We examine the impact of Hoot tutoring on students' end-of-year (EOY) reading performance, primarily focusing on the Overall Scale Scores. These scores are criterion-referenced and compare a student's knowledge and skills to grade-level standards. In this report, our primary outcome is the i-Ready assessment. We also explore the impact on the state-mandated MAP assessment, which was administered only to students in grades 3 and 4 of the study sample.

We also assess the impact on students' growth measures as reported by i-Ready, including Typical Growth and Stretch Growth. According to i-Ready's manual, the Annual Typical Growth metric indicates the progress a student is making in their academic skills compared to the national average expected growth (at the 50th percentile) for their grade level and starting point over the course of a school year. A "Progress to Annual Typical Growth" value of 100 or more signifies that a student has made 100 percent (or more) of the expected academic progress for their grade level and starting point within a given school year.

Conversely, the Stretch Growth metric establishes an ambitious yet attainable growth target for students, defining the progress a student needs to reach grade-level proficiency. This metric is informed by observations of students who began below grade level and subsequently improved their scores to achieve grade-level proficiency over time. The Stretch Growth values represent above-average growth but do not exceed the 80th percentile of growth for students in a given placement, ensuring these goals are attainable with additional instructional support and effort.

Analytical Sample

The analytical sample for the study includes 1,550 students across 14 elementary schools. Of these students, 707 were assigned to the Treatment group and 843 to the Comparison group. Overall, 70 percent of the students in the analytical sample qualified for MTSS-Tier III instruction. In the Treatment group, 493 students qualified for Tier III instruction, while 214 students qualified for Tier II instruction. In the Comparison group, 588 students qualified for Tier III, and 255 qualified for Tier II instruction. There are more students in Grades 2 and 3 (222 and 227 in the Treatment group; 265 and 270 in the Comparison group) because some schools did not include Grade 1 and/or Grade 4 students in the program. Grade 4 has the smallest sample sizes, with 103 in the Treatment group and 121 in the Comparison group, while there are 155 students in the Treatment and 187 in the Comparison in Grade 1.

Approximately 47 percent of the students in the analytical sample are Black, four percent are White, and three percent are Asian. More than 41 percent identify as Hispanic. About 38 percent of the students are English learners, 56 percent are economically disadvantaged (enrolled in SNAP), and 10 percent receive special education services.

We compare student characteristics—including gender, race, ethnicity, eligibility for English learner or special education services, and eligibility for free or reduced-priced lunch—along with their beginning-of-year i-Ready overall scores, between the Treatment and Comparison groups (see the balance Table A1 in Appendix A). Our findings indicate that the groups are similar, suggesting that randomization was successful and supporting a valid causal analysis.

Approximately 10 percent of the students in the analytical sample (balanced between the Treatment and Comparison groups) did not have any EOY assessment scores, either because they left the district or missed assessments. Attrition rates were similar between the Treatment and Comparison groups. We excluded these students from the outcome analyses.

Method

We present both aggregated (the analytical sample as a whole) and disaggregated results based on MTSS tiers, grade levels, and schools to better understand any differentiated treatment effects influenced by contextual factors.

We begin by conducting a series of descriptive analyses to summarize the implementation of Hoot tutoring, focusing on students' participation in tutoring sessions.

To examine the effects of HOOT on reading performance, we utilize both the original (raw) and standardized Overall Scale Scores. The i-Ready Overall Scale Score is a composite derived from individual tested sub-skill scores, calculated by the testing organization; the formula used can be found in the i-Ready technical report. We standardized the Overall Scale Scores by the mean of the Comparison group within each grade level to facilitate comparisons across grade levels.

To obtain causal estimates, we performed an Intent-to-Treat (ITT) analysis, which compares students' outcomes based on their initial random assignment to either the tutoring or comparison group, regardless of whether they actually received tutoring during the study period. Our statistical model controls for baseline performance (specifically, BOY i-Ready reading scores) and student covariates (such as demographic information and eligibility for individualized services). It also accounts for the school-grade-MTSS tier contexts by including strata as fixed effects (n=66) and employs robust standard errors at the student level.

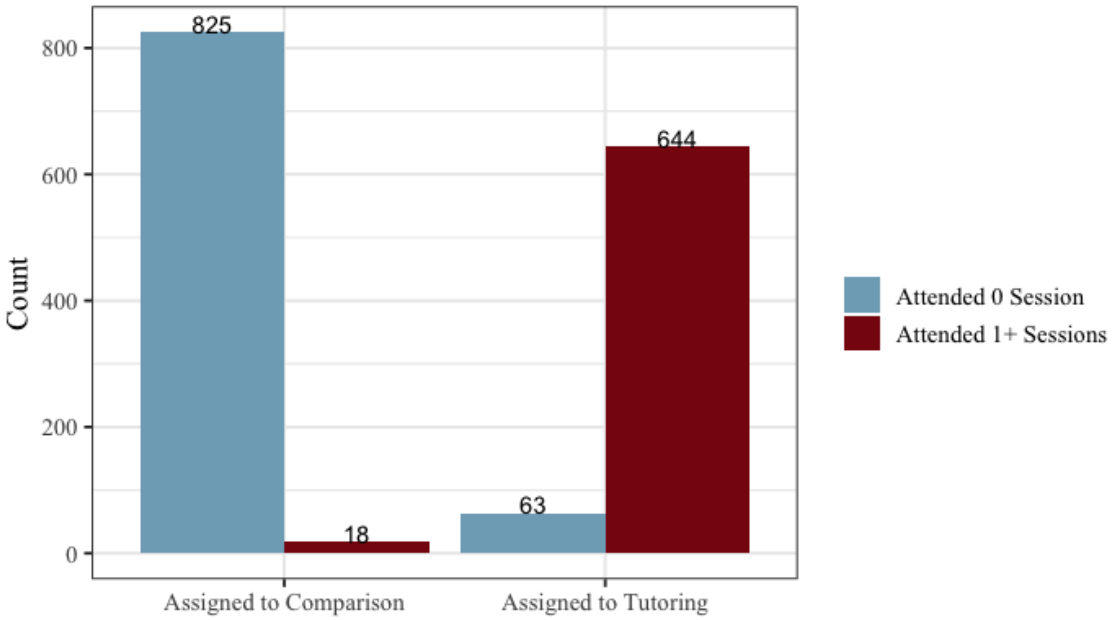
Due to logistical complexities, some students did not receive treatment even though they were initially assigned to tutoring. To adjust for this non-compliance, we conduct a Treatment-on-the-Treated (TOT) analysis to estimate the average effect of tutoring among those who participated in the tutoring program. The TOT model considers the same baseline performance, student covariates, fixed effects, and robust standard errors as the ITT model.

Findings

How Much Tutoring Did Students Receive?

Figure 1 illustrates the participation in tutoring among students in the Treatment and Comparison groups. Out of the 707 students assigned to the Treatment group for Hoot tutoring, 644 attended at least one tutoring session during the study period. A total of 63 students in the Treatment group did not attend any sessions. In contrast, 18 out of 843 students in the Comparison group, who were not assigned to receive Hoot tutoring, still attended at least one tutoring session.

Figure 1. Implementation Compliance with Random Assignment



On average, Hoot scheduled 56.30 tutoring sessions (SD=17.02) for each student in the Treatment group, as Figure 2 illustrates. The mean number of sessions attended per student was 32.01 (SD=15.19), which is slightly more than half of the total intended dosage (60 tutoring sessions). *The average attendance rate per student, considering the maximum intended dosage, was 53.35 percent (SD=25.32)*.

Figure 2. Average Number of Attended Tutoring Sessions by Treatment Condition

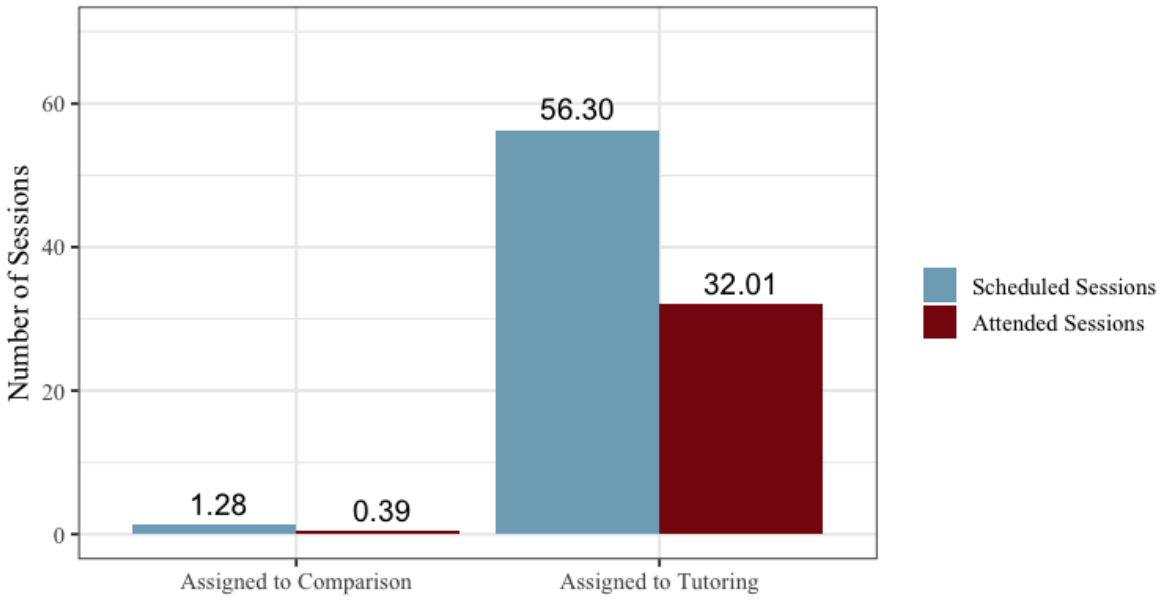
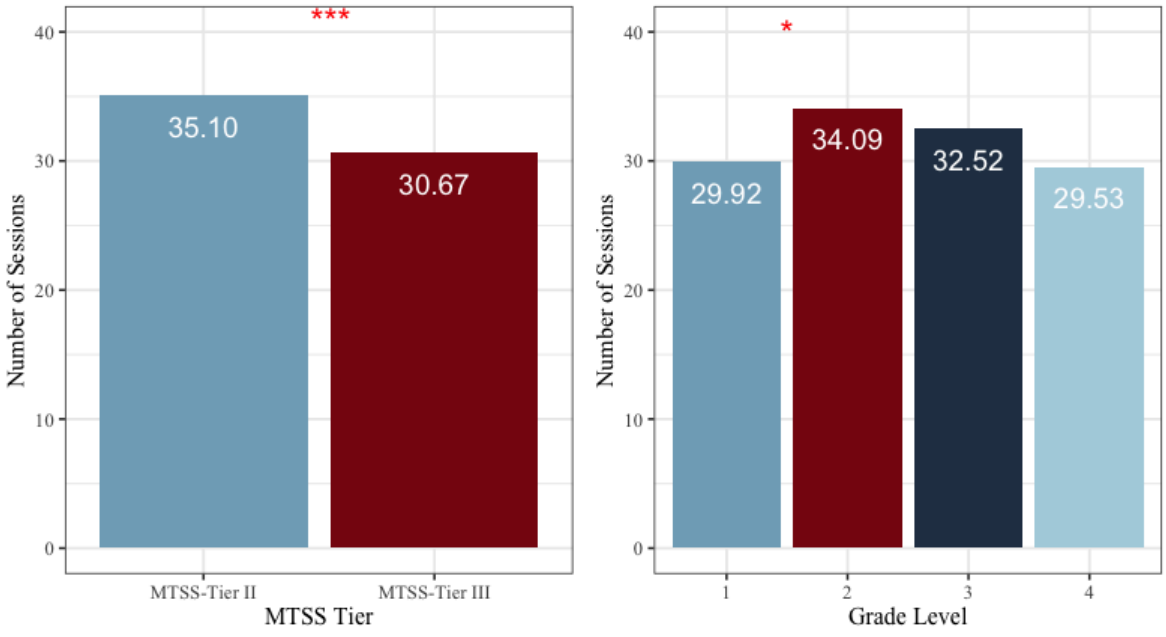


Figure 3. Average Attended Sessions of Treatment Students by MTSS Tier and Grade Level

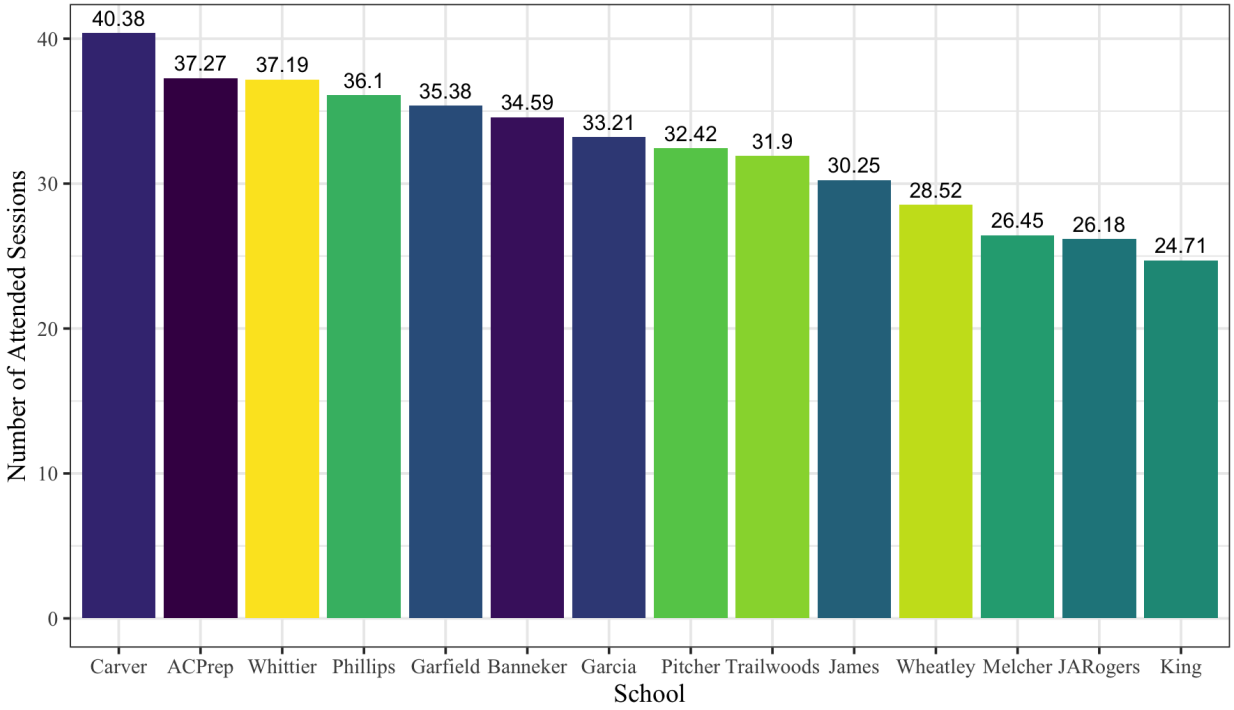


When breaking down attendance by MTSS tier and grade level (see Figure 3), we find that Tier II students attended an average of 4 more sessions (M=35.10, SD=13.46) compared to Tier III students (M=30.67, SD=15.71). This difference is statistically significant (p< .001). Among grade levels, Grade 2

students attended the most sessions (M=34.90, SD=14.54), followed by Grade 3 (M=32.52, SD=14.62), Grade 1 (M=29.92, SD=16.40), and Grade 4 (M=29.53, SD=15.39). The average number of sessions attended by Grade 2 students was significantly higher than that of Grade 1 students ($p < .05$).

Tutoring participation of Treatment students differed by schools. Figure 4 displays the average attended sessions by school from the highest to the lowest on the x-axis. On average, in 10 out of 14 schools, students attended at least 50 percent of the intended dosage (60 sessions in total), with 40.38 as the highest average number of sessions. Four schools had average attended sessions of fewer than 30 sessions, with 24.71 as the lowest average number of sessions.

Figure 4. Average Attended Sessions of Treatment Students by School



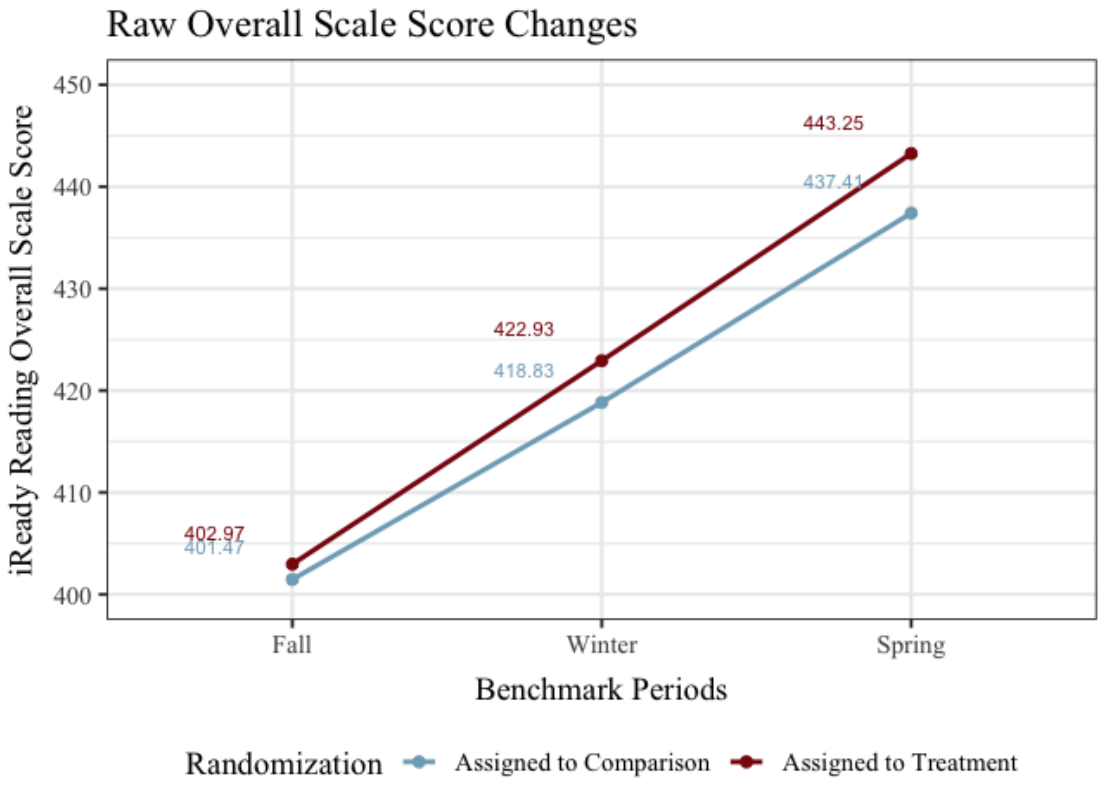
Did Hoot Tutoring Affect Students' End-of-Year Reading Proficiency?

Overall Scale Scores

On average, students demonstrated growth in their i-Ready reading assessments from the beginning of the year (BOY) to the end of the year (EOY), regardless of whether they were in the Treatment or Comparison group, as Figure 5 illustrates. The average Overall Scale Score during the BOY period was 402.97 (SD = 55.79) for the Treatment group and 401.47 (SD = 56.64) for the Comparison group. By the

EOY assessment, the Treatment group achieved an average Overall Scale Score of 443.25 (SD = 61.41), while the Comparison group scored an average of 437.41 (SD = 62.84). Similar upward trends are observed when the sample is disaggregated by MTSS Tier and grade level (see Figure B1 in Appendix B).

Figure 5. Average Overall Scale Scores by Treatment Condition and Benchmark Periods



Intent-to-Treat (ITT) Analysis

We performed an Intent-to-Treat (ITT) analysis to obtain causal estimates. Table 2 summarizes the results from four different ITT model configurations. Models 1 and 2 include all students in the analytical sample (N=1,384), while Models 3 and 4 focus on MTSS-Tier II (N=428) and MTSS-Tier III students (N=956), respectively. Model 1 represents the parsimonious configuration that accounts for strata. It shows that students assigned to Hoot tutoring surpassed their comparison peers at the EOY reading assessment by 0.08 standard deviation units (SE = 0.04, p = .068) after accounting for the school-grade-MTSS factors. Model 2 additionally accounts for student-level factors, resulting in a positive but imprecise effect size of 0.05 standard deviations (SD), with a standard error (SE) of 0.03 (p=.12). The effect size is equivalent to 3.01 (SD=1.89) raw Overall Scale Score points.

Table 2. ITT Effect on Standardized EOY iReady Reading Overall Scale Scores

	iReady EOY Overall Scale Scores (Std.)							
	All STUDENTS				MTSS TIER-II STUDENTS		MTSS TIER-III STUDENTS	
	(1)		(2)		(3)		(4)	
	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*
Assigned to Tutoring	0.08+ (0.04)	.068	0.05 (0.03)	.116	-0.02 (0.06)	.712	0.08* (0.04)	.032
FE: Strata	V		V		V		V	
BOY std scores			V		V		V	
Student-covariates			V		V		V	
Control Mean	0.00		0.00		0.816		-0.367	
Student Observations	1384		1384		428		956	
R^2 / R^2 adjusted	0.666 / 0.647		0.390 / 0.359		0.490 / 0.439		0.558 / 0.536	

Notes. Student covariates include demographic variables (gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP); BOY iReady overall scale scores are standardized by grade level and benchmark periods; Fixed effect (FE) is on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

The positive effects overall are driven by positive effects for students in MTSS Tier III. *Tier III students in the Treatment group significantly outperformed their Comparison peers by 0.08 standard deviations (SE = 0.04) or 4.74 (SD=2.27) raw score points* (p < .05). No similar trend is observed among Tier II students in the Treatment compared to the Comparison group (Model 3).

A breakdown of treatment effects by grade level (Table A2 in Appendix A) reveals that, except for Grade 1, students in the Treatment group in Grades 2, 3, and 4 demonstrate positive outcomes compared to their peers in the Comparison group. The effect sizes range from 0.05 to 0.10 SDs (or 3.08 to 7.02 raw score points), which are medium in scale but statistically insignificant.

Treatment-on-Treated (TOT) Analysis

We conduct a Treatment-on-Treated (TOT) analysis to assess the average effect of tutoring on students who actually attended Hoot tutoring sessions, as Table 3 shows. Our findings indicate that students assigned to the Treatment group who participated in tutoring attended an average of 33.12 sessions (p<.001, see Model 1 in Panel B: First Stage). The average number of sessions attended was higher for students in MTSS-Tier II (M = 36.31, p < .001) compared to those in Tier III (M = 31.91, p < .001).

Table 3. TOT Effect on Standardized EOY iReady Reading Overall Scale Scores

Panel A. 2SLS	iReady EOY Overall Scale Scores (Std.)					
	All Students (1)		MTSS-TIER II (2)		MTSS-TIER III (3)	
	Est.	p	Est.	p	Est.	p
Amount of Tutoring	0.0016 (0.001)	.11	-0.001 (0.002)	.71	0.003 (0.001)	.031*
R2/R2 Adj.	0.668/0.649		0.489/0.438		0.560/0.538	
Panel B. First Stage						
Assigned to Treatment	33.118*** (0.529)		36.306*** (0.915)		31.905*** (0.642)	
F-Statistic	3,915.9		1,573.7		2,473.3	
Observations	1384		428		956	

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

On average across all students assigned to the Treatment group who engaged in Hoot tutoring, each additional tutoring session is associated with a slight improvement in their EOY i-Ready reading scores—approximately 0.0016 SD (or 0.091 raw score points) - which is not statistically different from zero as Model 1 in Table 3 shows. However, for MTSS-Tier III students in the Treatment group who participated in Hoot tutoring (Model 3), each additional session resulted in a statistically significant improvement of 0.003 SD (or 0.149 raw score points, p<.05). Based on this estimate, completing 30 sessions could lead to an approximate improvement of 0.09 SD on the EOY assessment.

When we break down the effect of the tutoring intervention on students' EOY Overall Scale Scores by grade level (see Table A3 in Appendix A), we observe a trend similar to the results from the ITT analysis. While the estimated effect sizes range from 0.002 to 0.003 SD for students in Grades 2, 3, and 4, these estimates are not statistically significant. The estimates for students in Grade 1 are smaller in magnitude.

Growth Measures

Descriptive Analysis

Overall, students in the Treatment group who received Hoot tutoring outperformed their peers in the Comparison group on their EOY reading performance using growth measures. Figure 6 illustrates the mean differences based on treatment conditions and growth outcomes. On average, students in the Treatment group exceed the 100% Annual Typical Growth by 10.35 percentage points. In contrast, students in the Comparison group miss the 100% Annual Typical Growth goals by 0.17 percentage points. Furthermore, students in the Treatment group achieve 61.73% progress toward Annual Stretch Growth, which is 5.36 percentage points higher than that of students in the Comparison group.

Figure 6. Mean Percent Progress toward Annual Typical and Stretch Growths

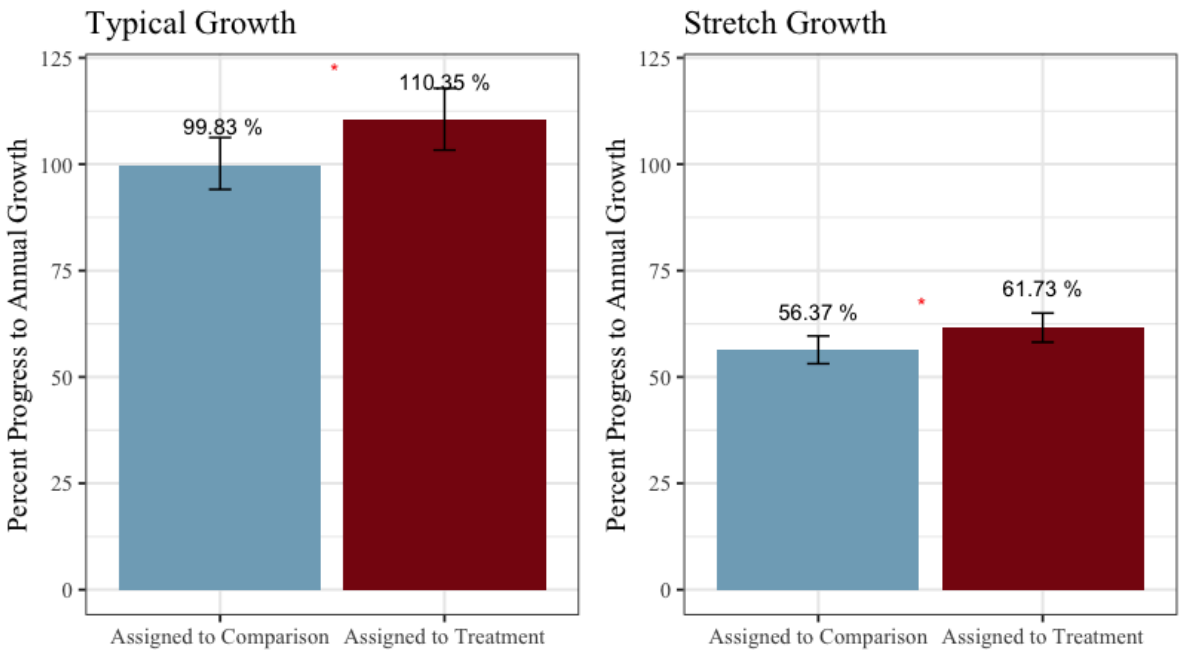
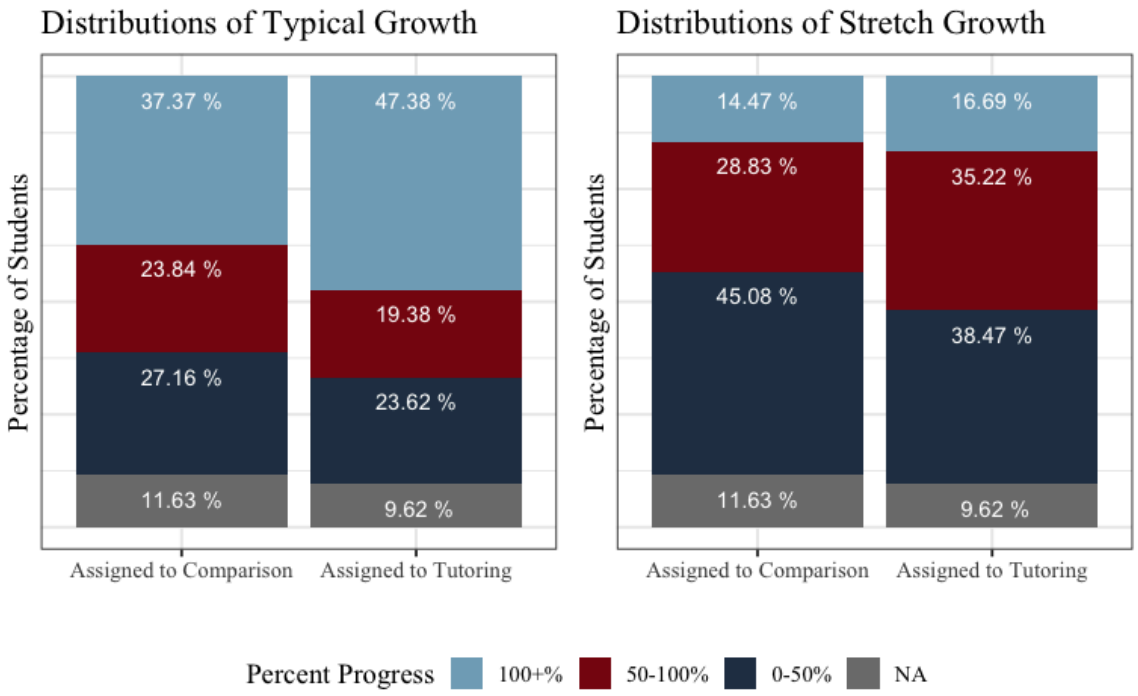


Figure 7 shows the distribution of students in the analytical sample who made 0-50%, 50-100%, or 100%+ progress toward their annual growth goals. By the end of the school year, 47.38 percent of students in the Treatment group and 37.37 percent of students in the Comparison group met or exceeded their Annual Typical Growth goals. Additionally, 16.69 percent of Treatment group students and 14.47 percent of Comparison group students met or exceeded their Annual Stretch Growth goals.

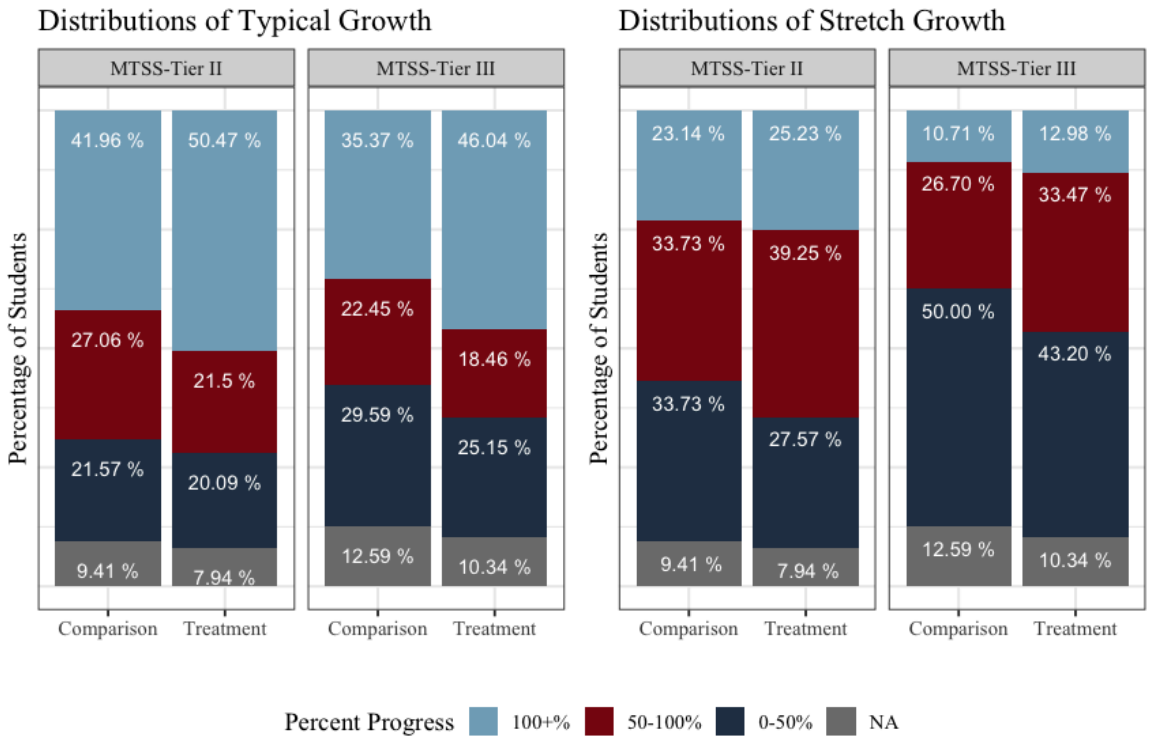
Figure 7. Distributions of Student Percent Progress to Annual Growth Goals



Both Tier II and Tier III students made significant progress toward their Annual Typical Growth goals. Tier II students showed more considerable progress toward grade-level proficiency (Annual Stretch Growth, see Figure B2 in Appendix B). A higher percentage of Tier II students met and exceeded both Annual Typical and Stretch Growth goals compared to their Tier III peers as shown in Figure 8.

The descriptive analyses show some differences between the Treatment and Comparison groups. The treatment-control differences are more pronounced among Tier III students than among Tier II students. Across grade level (Figure B3 in Appendix B), students in the Treatment group perform better toward their annual growth goals compared to their peers in the Comparison group, though these differences are not statistically significant. Additionally, a larger proportion of students in Grade 4 meet or exceed their Annual Typical and Stretch Growth goals (Figure B4 in Appendix B) compared to students in other grades.

Figure 8. Distributions of Student Percent Progress to Annual Growth Goals by MTSS Tier



Intent-to-Treat (ITT) Analysis

An ITT analysis of students' learning outcomes using iReady's growth measures, shown in Table 4, estimates that *Hoot Reading leads to a statistically significant improvement in Annual Typical Growth (Model 1) of 10.84 percentage points (SD=4.58) and in Annual Stretch Growth goals (Model 4) by 5.24 percentage points (SD=2.50)*. These findings account for baseline performance, student-level covariates, and school-level fixed effects, with the differences statistically significant (p<.05).

For MTSS-Tier III students, the effect was larger with Hoot Reading leading to a 13.77 percentage points (SD=5.53) increase in progress toward the Annual Typical Growth goals (Model 3) and a 6.64 percentage points (SD=2.76) increase in Annual Stretch Growth goals (Model 6). These differences are statistically significant (p<.05) after controlling for baseline performance, student-level covariates, and school-level fixed effects. Differences between the Treatment and Comparison groups are smaller and not statistically significant for Tier II students (Models 2 and 5).

Table 4. ITT Effects on Students' Percent Progress to Annual Growth Goals

	EOY Percent Progress to Annual TYPICAL Growth						EOY Percent Progress to Annual STRETCH Growth					
	ALL		MTSS-T2		MTSS-T3		ALL		MTSS-T2		MTSS-T3	
	(1)		(2)		(3)		(4)		(5)		(6)	
	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*
Assigned to Tutoring	10.84 (4.58)	.018*	3.86 (8.22)	.638	13.77 (5.53)	.013*	5.24 (2.50)	.036*	1.98 (5.47)	.717	6.64 (2.76)	.016*
Observations	1384		428		956		1384		428		956	
R^2	0.129		0.136		0.133		0.140		0.098		0.115	
R^2 adjusted	0.079		0.049		0.090		0.091		0.007		0.072	

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

We also observe variations in the treatment effects on students' progress toward their annual growth goals by grade level (see Table A4 in Appendix A). Treatment group students in Grade 4 showed a particularly positive trend, outperforming their peers in the Comparison group by 28.74 percentage points (SD=16.52, p=.084) for Annual Typical Growth and by 13.88 percentage points (SD=8.06, p=.087) for Annual Stretch Growth. Similarly, Grade 2 students in the Treatment group performed better than their counterparts, achieving an increase of 12.24 percentage points (SD=7.35, p=.097) toward their Annual Typical Growth goals.

Did Hoot Tutoring Improve Students' Performance on the State Test?

Students in grades 3-8 are required to take the state-mandated Missouri Assessment Program (MAP) in English Language Arts (ELA) each year. In our analytical sample of 1,550 students, 46.5 percent were in Grades 3 and 4, where students are expected to take the MAP test. However, 11.7 percent of Grade 3 students (N=497) and 14.3 percent of Grade 4 students (N=224) did not have a MAP score at the end of the school year. Missing data on these assessments does not differ between the Treatment and Comparison groups, leaving a sample of 294 Treatment-group students and 337 Control-group students with MAP data.

Table 5 shows the effect of Hoot tutoring on students' EOY MAP assessments in ELA. Overall, the results reflect patterns similar to those observed in the i-Ready assessments, with an estimated treatment effect of .07 standard deviations. For fourth grade, the estimate is larger at 0.16 standard deviations, compared to 0.04 for third grade. None of the estimates is statistically significant at traditional levels.

Table 5. ITT Effect on EOY MAP Overall Scale Scores by Grade Level

	G3+G4 (N=721)		G3 (N=497)		G4 (N=224)	
	Est. (Raw)	*Est. (Std.)*	*Est. (Raw)*	*Est. (Std.)*	*Est. (Raw)*	*Est. (Std.)*
Tutoring	2.55 (2.11)	0.07 (0.06)	1.55 (2.56)	0.04 (0.07)	5.54 (3.77)	0.16 (0.11)
Observations	631		439		192	
R2 / R2 Adj.	0.556 / 0.528		0.510 / 0.475		0.562 / 0.520	

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Did Treatment Effects Differ by School?

Randomizing students within schools to receive Hoot Reading allows us to assess differences in the effects of the program across schools. Schools may differ both in the quality of their implementation program for the treatment group and in the experiences of the comparison group.

During the implementation of Hoot Reading tutoring, the district conducted three additional known reading interventions. The Literacy for All Students (LFAS) intervention, based on the Mississippi model, employed a top-down approach that included principal training and the use of literacy coaches. Three schools in the study—Banneker, James, and Whittier—were part of the LFAS intervention. Simultaneously, teachers in the Wheatley school were provided with an AI literacy tool called PlaBook; however, anecdotal evidence suggests that teachers did not have a positive experience with it, and it is unclear how effectively they used it to enhance teaching and learning. The Halls Family Foundation provided a bottom-up intervention in which it funded literacy supports, which could include Hoot

Reading. If schools choose Hoot Reading as part of their use of the funds or for all their funds, we were able to randomize students into Treatment and Comparison groups for Hoot. Students in Garcia, Garfield, and Wheatley participated in this intervention: those assigned to the treatment group received Hoot tutoring, while control students continued with standard instructional practices without Hoot tutoring, although they might have received other literacy interventions.

Because these literacy interventions may have clouded the effects of Hoot Reading, we estimated the effects of the program without schools in the other interventions (see Table A5 in Appendix A for detailed treatment effects). *When we focused on schools that did not receive the top-down Mississippi literacy intervention (non-LFAS schools), students in the Treatment group who received Hoot tutoring outperformed their peers in the Comparison group by 0.07 SD (SE=0.03) in their EOY reading Overall Scale Scores* (Model 1). This effect size is statistically significant ($p<.05$). The estimated effects on growth for the non-LFAS schools are 13.19 percentage points for Annual Typical Growth and 6.71 percentage points for Annual Stretch Growth, both significant at $p<.01$. For Tier III students in non-LFAS schools, the estimated effects are 13.77 percentage points and 6.64 percentage points, significant at the $p<.05$ level. When we limit the sample to exclude only the Hall schools (Model 4) or both the LFAS and Hall schools (Model 7), we find similar positive trends, where treatment students perform better than their peers in the comparison group; however, the effect sizes are smaller and less precise.

We also examine the impact of Hoot tutoring on a school-by-school basis, summarizing the treatment effects on students' EOY reading Overall Scale Scores (both original/raw and standardized) by school. Results show that students in the Treatment group at King and Wheatley significantly outperformed their peers in the Comparison group by 0.20 SD (10.51 raw score points) and 0.15 SD (8.94 raw score points), respectively, in their EOY reading performance, which are statistically significant. However, students in the Treatment group at Carver performed worse than their Comparison peers by 0.22 SD (12.39 raw score points), which is marginally statistically significant. Implementation quality ratings from Hoot do not explain this variation across schools.

Takeaways

This study examines the impact of Hoot Reading tutoring on students' early literacy skills as measured by the i-Ready Reading assessment. We focused specifically on the effectiveness of integrating one-on-one, high-impact tutoring in MTSS-Tier III instruction for students in grades one to four, though we included some students in MTSS-Tier II.

The intervention had a positive and statistically significant impact, especially for students who scored well below grade-level benchmarks on the beginning-of-year (BOY) assessment (i.e., MTSS-Tier III students). Seventy percent of students in the analytical sample were in this group. MTSS-Tier III students in the Treatment group outperformed their peers in the Comparison group by 0.08 SD (SE = 0.04, $p<.05$) in their EOY i-Ready Reading Overall Scale Scores. The effect of the program is larger (0.09 SD, SE=0.04, $p<.05$) when accounting for Tier-III students who were assigned to Treatment and took up tutoring. To translate the standard deviation into months of learning, a 0.08 SD increase in reading growth corresponds to an additional 0.82, 1.33, 2.11, and 2.00 months of learning for grades 1, 2, 3, and 4, respectively.[2]

Hoot Reading also demonstrates positive effects on other measures of learning. Being randomly assigned to the treatment group led to a statistically significant improvement in Annual Typical Growth of 10.84 percentage points and in Annual Stretch Growth goals by 5.24 percentage points. Moreover, for the subset of third and fourth graders who took the state assessment (MAP) at the end of the year, the estimated effect sizes for the impact of Hoot Reading are similar to the i-Ready results, though the estimates are not statistically significant.

For many schools involved in this study, this was their first randomized controlled trial, which posed challenges for administrators, teachers, students, and coordinators in launching the program. Students assigned to Hoot Reading received only 53% of the total tutoring that was the goal for the program. Some schools faced technical issues, such as limited internet bandwidth at the program's start, and many had to cancel classes and tutoring sessions due to unexpected snow days. Despite these implementation challenges, the effect size observed for MTSS-Tier III students (0.08 SD) in our study aligns with the effect sizes reported by other virtual tutoring programs in the field.

By utilizing high-impact tutoring within the MTSS framework, schools can significantly boost student learning and help reduce disparities in educational experiences and outcomes without incurring

[2] This calculation is based on Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). "Empirical benchmarks for interpreting effect sizes in research." Child development perspectives, 2, no. 3, 172-177.

substantial additional costs. Successful implementation requires careful planning, effective resource allocation, and sustained monitoring.

Appendix A

Table A1.

Balance Table of the Analytical Sample

	Comparison			Treatment			All Students			
	Mean	SD	n	Mean	SD	n	Mean	SD	n	*p*-value
Female	0.49		843	0.48		707	0.49		1550	.66
Black	0.47		843	0.48		707	0.47		1550	.72
White	0.04		843	0.05		707	0.04		1550	.90
Asian	0.02		843	0.04		707	0.03		1550	**.06+**
Hispanic	0.41		843	0.40		707	0.41		1550	.69
Multirace	0.05		843	0.03		707	0.04		1550	.12
EL	0.38		843	0.38		707	0.38		1550	.99
SNAP	0.55		843	0.56		707	0.56		1550	.81
IEP	0.11		843	0.08		707	0.10		1550	**.07+**
BOY iReady Scale Score										
All	401.47	56.64	843	402.97	55.79	707	402.15	56.24	1550	.60
G1	358.88	37.59	187	360.83	37.69	155	359.77	37.59	342	.63
G2	387.84	47.60	265	390.87	43.76	222	389.22	45.87	487	.47
G3	423.67	46.42	270	420.96	50.49	227	422.43	48.29	497	.53
G4	447.58	63.34	121	452.79	57.91	103	449.97	60.82	224	.52
MTSS Tier-III										
All	385.26	48.60	588	388.68	50.03	493	386.82	49.26	1081	.26
G1	321.12	18.54	75	320.91	22.13	58	321.03	20.10	133	.95
G2	369.17	34.80	206	373.43	31.80	171	371.10	33.50	377	.22
G3	409.34	38.39	221	405.94	42.90	186	407.78	40.50	407	.40
G4	417.86	49.33	86	431.33	49.36	78	424.27	49.66	164	**.08+**
MTSS TierII										
All	438.84	56.26	255	435.88	54.52	214	437.49	55.44	469	.57
G1	384.17	23.02	112	384.70	21.21	97	384.42	22.15	209	.86
G2	453.03	21.70	59	449.33	22.38	51	451.32	22.00	110	.38
G3	488.33	10.51	49	489.12	8.95	41	488.69	9.78	90	.70
G4	520.60	18.77	35	519.72	16.50	25	520.23	17.72	60	.85

Table A2.

ITT Effect on Standardized EOY iReady Reading Overall Scale Scores by Grade Level

	iReady EOY Overall Scale Scores (Std.)							
	Grade 1		**Grade 2**		**Grade 3**		**Grade 4**	
	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*
Assigned to Tutoring	-0.01	.94	0.08	.16	0.05	.35	0.10	.158
	(0.08)		(0.06)		(0.06)		(0.07)	
Observations	307		441		441		195	
R^2 / R^2 adjusted	0.642 / 0.607		0.632 / 0.606		0.704 / 0.683		0.765/ 0.743	

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A3.

TOT Effect on Standardized EOY iReady Reading Overall Scale Scores by Grade Level

Panel A. 2SLS	iReady EOY Overall Scale Scores (Std.)							
	Grade 1		**Grade 2**		**Grade 3**		**Grade 4**	
	Est.	*p*	*Est.*	*p*	*Est.*	*p*	*Est.*	*p*
Amount of Tutoring	-0.0002	.94	0.002	.16	0.002	.35	0.003	.16
	(0.003)		(0.002)		(0.002)		(0.002)	
Panel B. First Stage								
Assigned to Treatment	30.795***		34.781***		34.451***		30.396***	
	(1.388)		(0.884)		(0.830)		(1.442)	
F-Statistic	492.2		1,549.1		1,724.7		444.0	
Observations	307		441		441		195	

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A4.

ITT Effects on Percent Progress to Annual Growth Goals by Grade Level

	EOY Percent Progress to Annual TYPICAL Growth				EOY Percent Progress to Annual Stretch Growth			
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 1	Grade 2	Grade 3	Grade 4
Assigned to Tutoring	1.19 (7.38)	**12.24+ (7.35)**	10.85 (8.49)	**28.74+ (16.52)**	0.76 (5.03)	6.37 (4.30)	5.26 (4.32)	**13.88 + (8.06)**
P-value	.872	**.097**	.202	**.084**	.880	.139	.224	**.087**
Observations	307	441	441	195	307	441	441	195
R^2	0.091	0.099	0.187	0.091	0.150	0.129	0.224	0.110
R^2 adjusted	0.003	0.036	0.129	0.004	0.068	0.068	0.170	0.025

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A5.

ITT Effect on EOY iReady Reading Overall Scale Scores by Subsets of School

	iReady EOY Overall Scale Scores (Std.)								
	Non-LFAS schools			Non-Hall schools			NonLFAS/Hall schools		
	(1) All	(2) T2	(3) T3	(4) All	(5) T2.	(6) T3	(7) All	(8) T2.	(9) T3
Assigned to Tutoring	**0.07* (0.03)**	**0.03 (0.07)**	**0.08* (0.04)**	0.03 (0.04)	-0.04 (0.08)	0.07 (0.05)	0.06 (0.05)	0.03 (0.10)	0.03 (0.10)
p-value	**.035**	**.635**	**.032**	.438	.574	.188	.178	.733	.188
Control Mean	-0.059	0.886	-0.367	0.0338	0.868	-0.324	-0.0568	1.05	-0.324
Observations	1224	268	956	907	307	600	747	147	600
R^2	0.671	0.559	0.558	0.647	0.443	0.535	0.656	0.481	0.535
R^2 adjusted	0.653	0.507	0.536	0.624	0.380	0.508	0.635	0.399	0.508

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. +p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A6.

ITT Effect on EOY iReady Reading Overall Scale Scores by School

Schools	Assigned to Tutoring		Observations	Implementation Rating
	Estimate (Raw)	*Estimate (Std.)*		
ACPrep	0.02 (7.62)	0.01 (0.14)	88	11
Banneker	-9.99 (10.62)	-0.19 (0.20)	83	7
Carver	**-12.39 (6.93)+**	**-0.22 (0.12)+**	89	11
Garcia	4.54 (6.06)	0.08 (0.11)	186	5
Garfield	2.86 (4.66)	0.04 (0.08)	144	6
James	**-23.41 (9.86)***	**-0.41 (0.18)***	24	8
JARogers	8.09 (7.43)	0.13 (0.13)	124	8
King	**10.51 (5.75)+**	**0.20 (0.10)***	171	8
Melcher	-1.29 (8.82)	-0.03 (0.15)	66	8
Phillips	3.32 (6.84)	0.07 (0.12)	85	11
Pitcher	-1.10 (11.27)	-0.01 (0.21)	50	10
Trailwoods	2.04 (7.18)	0.04 (0.13)	74	5
Wheatley	**8.94 (4.51)***	**0.15 (0.07)***	147	5
Whittier	-2.54 (11.65)	-0.03 (0.21)	53	7

Notes. All models account for students' baseline performance (i.e., BOY iReady overall scale scores standardized by the control mean), student covariates (including gender, Black, Asian, White, Hispanic, EL, IEP, and SNAP), and fixed effects on the strata level (i.e., school-grade-MTSS, n=66). Robust standard errors in parentheses. $+p < 0.1$, $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$

Appendix B

Disaggregated Student Learning Outcomes by MTSS TIER and Grade Level

Figure B1.

Raw Overall Scale Score Changes by Treatment Condition, MTSS-Tier, and Benchmark Periods.

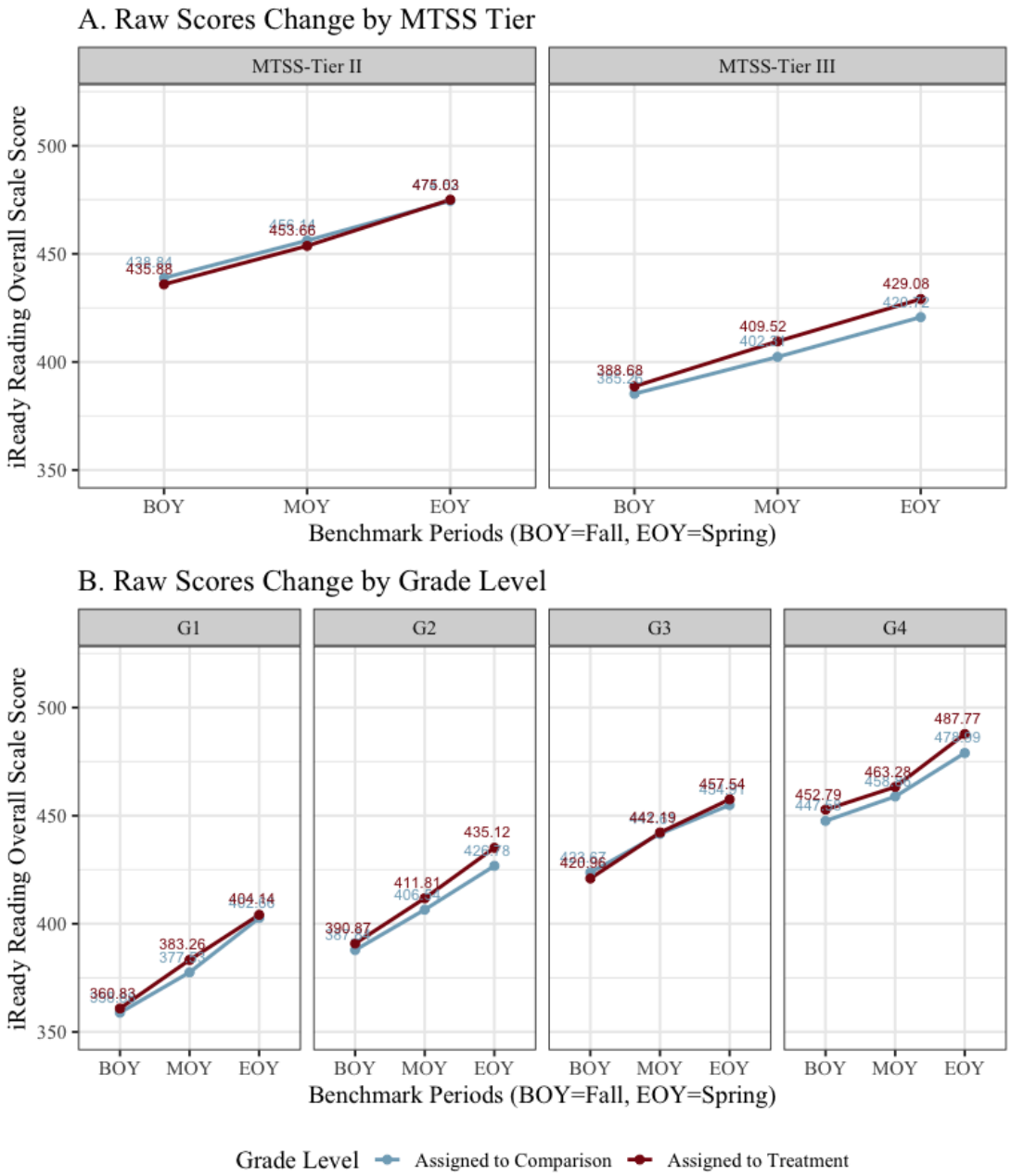


Figure B2.

Mean Percent Progress to Annual Typical and Stretch Growth by Treatment Condition and MTSS-Tier

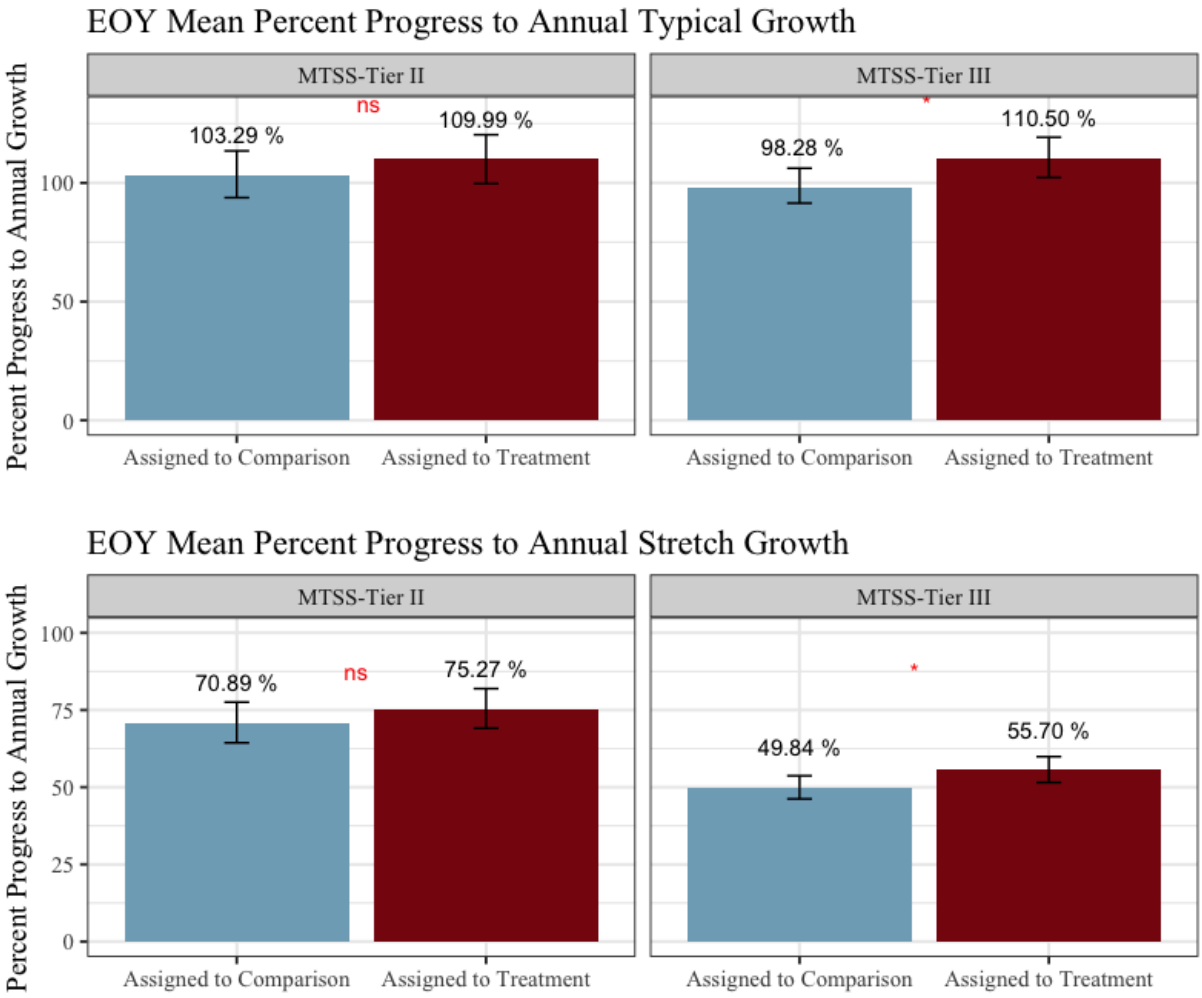


Figure B3.

Mean Percent Progress to Annual Typical and Stretch Growth by Treatment Condition and Grade Level

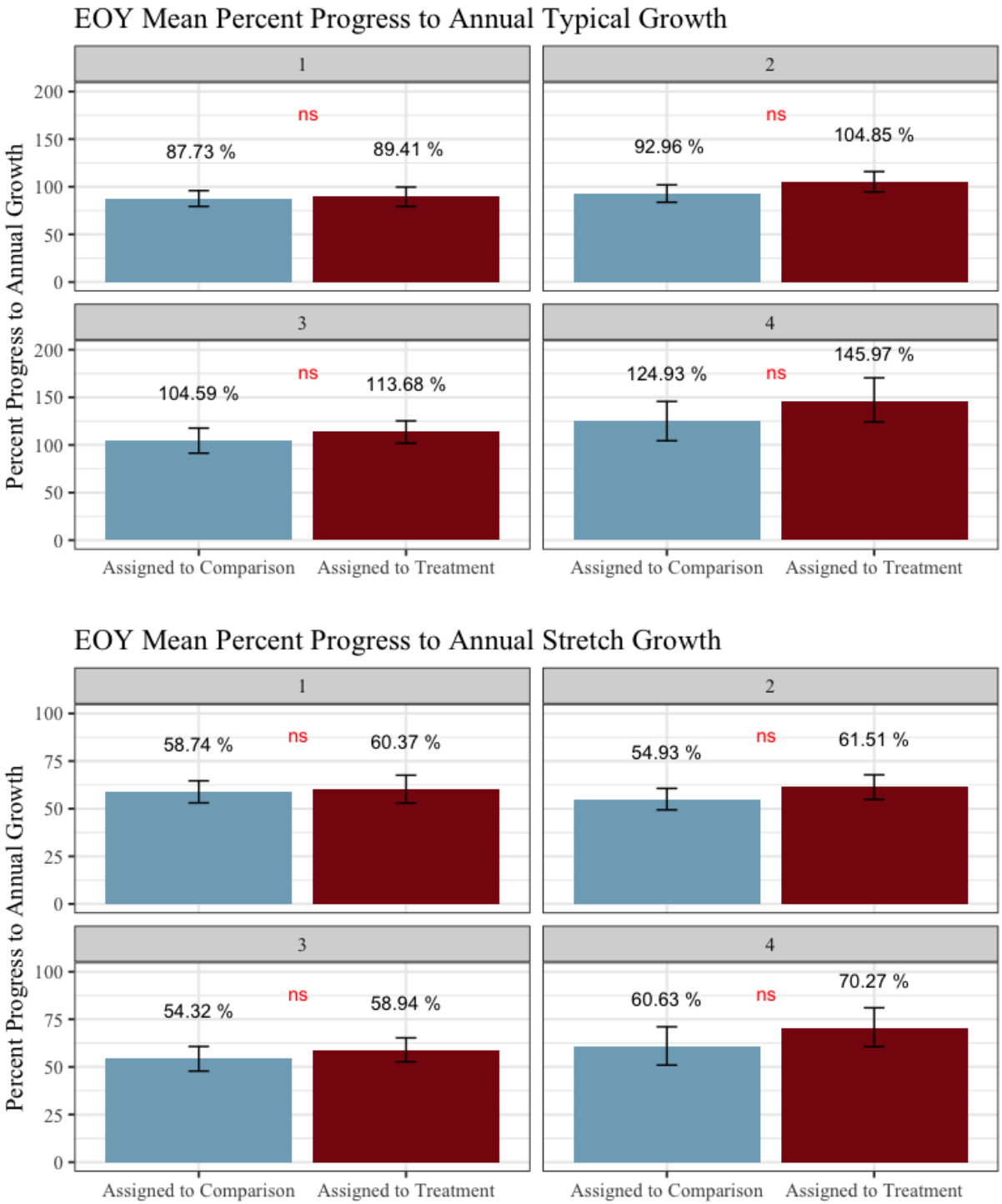


Figure B4.

Distributions of Students and their Percent Progress to Annual Growth Goals by Treatment Condition and Grade Level