NATIONAL STUDENT SUPPORT ACCELERATOR

equalizing access to quality tutoring

Effects of High-Impact Tutoring on Early Literacy Outcomes: A Pilot Study of a 1:1 Program With Existing Staff

Evan Bennett
Monica Lee
Susanna Loeb
Carly Robinson
Stanford University

May 2024

studentsupportaccelerator.org

Executive Summary

During the 2022-23 school year, Try Once, Inc. ("Once") partnered with a large, urban school district on the East Coast to provide high-impact early literacy tutoring to 105 kindergarten and first grade students in 13 schools. The district identified students as eligible for tutoring services if they scored below grade-level benchmarks in their early literacy skills. The Stanford research team randomly assigned eligible students into a tutoring program group (n=105) and a comparison group (n=199). Students in the program group were supposed to receive tutoring for 15 minutes every day during the school day between November 2022 and June 2023, one-on-one from a non-teaching staff member at their school. This report describes the research study design, the characteristics of students who participated in the study, tutoring participation rates, and the effect of receiving tutoring on end-of-year early literacy skills, both overall and within various subgroups.

Findings

1. Students randomly assigned to receive Once tutoring performed two points better on the end of year DIBELS (early literacy) assessment compared to students in the comparison group. The sample size of the study was too small to determine whether this difference is the result of the tutoring, but points to the promise of large-scale evaluations.

2. Students in the program group received an average of 42 sessions of tutoring, which was far less than the 140 sessions that the program aimed to deliver. The number of sessions students received varied meaningfully across schools and students designated as English Learners or eligible for Special Education received more tutoring, on average, than did other students.

Introduction

High-impact tutoring – intensive, personalized, instruction provided at least several times a week during the school day – is one of few school-based interventions shown to have consistently large, positive effects on student academic achievement (Harris, 2009). Despite the strong evidence for high-impact tutoring, it is often challenging to implement these types of interventions at scale, and one barrier to scaling an effective intervention is the human capital resources needed to do so. Tutoring is generally more effective when tutors are classroom teachers (Gerston et al., 2020), but allocating classroom teachers' hours to implementing small-group instruction is not feasible without taking away the hours they would use to do whole-classroom instruction, and adding additional teachers to the school staff is costly (Slavin et al., 2011).

This report describes a randomized controlled trial (RCT) designed to investigate whether tutoring, when provided by paraprofessionals as tutors, improved student outcomes in early literacy skills. Paraprofessionals are non-teaching staff who take on a diverse array of roles in schools, including classroom support, support for special education students, clerical work, and behavior management (Hemelt et al., 2021). Paraprofessionals are already present and staffed by the school, which means that they can potentially be brought on for in-school tutoring settings much more easily than volunteers. At the same time, they do not have whole-classroom responsibilities to the same extent as traditional classroom teachers, although they have the potential to provide instruction and build relationships with students in small group settings.

The Once Tutoring Model

Once was developed in early 2022 with the objective to promote kindergarten and first grade reading fluency. The organization's high-impact tutoring model contains evidence-based features such as:

- Daily delivery of early literacy instruction, for 15 minutes a day
- Tutoring during the school day in a one-on-one setting
- Highly scripted instructional sessions, along with weekly, synchronous coaching for tutors from Once's instructional coaches to review and preview materials
- Dashboards to monitor student progress through the tutoring curriculum and ongoing needs

As noted, Once typically leverages educators who are already employed by their school district partners to act as tutors; this was also the case for the district in which the study took place. For the most part this included paraprofessionals, but in some cases it also included lunch monitors, assistant principals, or other staff in the building with time in their work schedule to tutor students. The use of pre-existing school staff as tutors may allow for deeper relationships or a greater understanding of the school

context than would occur with external staff. This operational model was set in place with the idea that districts can proceed with program implementation with minimal (if any) additional human resources investments, often a key barrier for scaling a program successfully (White et al., 2021).

From an operational perspective, Once employed several strategies to fully embed tutoring into the school day. First, Once worked to identify a point of contact within the school building to help set up tutoring schedules and lead the effort. Second, each tutor received a "caseload" of a specific number of students to work with, which had two implications. First, this strategy implied that students were consistently interacting with the same adult tutor. It also established tutoring as a significant portion of each tutor's work week, legitimizing the practice as an important part of the school day. Third, Once trained teachers about their tutoring curriculum and implementation process in addition to the tutors themselves, in an effort to increase educator buy-in.

Study Design

At the start of the 2022-23 school year, the participating school district identified 13 elementary schools to participate in the study. The number of tutoring seats available varied across schools due to a limited number of school staff who could serve as tutors at each school. Generally, there was a cap of approximately five students who could receive tutoring in each classroom.

To understand the effectiveness of tutoring, the research team conducted an RCT where 304 students were randomly assigned to receive Once tutoring. In order to be eligible for the study, students had to score one or more levels below what was expected for their grade in early literacy skills at the beginning of the school year. Specifically, within each classroom and baseline reading level band, we randomly assigned 105 eligible students to the program 199 to the comparison group (or business as usual instruction).[1]

Beginning in November 2022, kindergarten and first grade students in the program group received tutoring using the Once curriculum and model (daily one-on-one lessons for 15 minutes a day). In line with the Once curriculum, all students started at the beginning of the curriculum ("Cycle 0") regardless of their grade level or baseline knowledge of early literacy skills. Students in the program group proceeded at their own pace through Once cycles (i.e., lessons) throughout the school year, while

---

[1] In addition to random assignment to the program, the research team randomly assigned and ordered comparison students on a waitlist for the program, assuming that program students could leave the study and open up additional tutoring slots for other students to take up. If comparison group students were taken off the waitlist to receive tutoring before November 15th, the first day of Once tutoring at the district, we consider these students as those who received the program in our analysis. Nine students in the comparison group backfilled the program group later in the year after the November 15th start date; these cases are considered to be "waitlisted-then-tutored", and they are considered as part of the comparison group in our main analysis.

students in the comparison group continued to receive mainstream classroom instruction, or another activity ongoing for the rest of the classroom during the hours Once tutoring took place for the students in the program group. Once tutoring for students in the program group concluded in late May to early June 2023.

Data

We use several sources of data to conduct our analyses. First, we collected administrative data from the partnering school district consisting of student gender, race/ethnicity, an indicator for being designated as an English Learner, and an indicator for being designated as receiving special education services.[2] The district also provided information on student performance on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment and the i-Ready Math assessment, both administered three times each year. We also collected roster and program data from Once, which included the date when students started and finished Once tutoring, the number of cycles (distance into the Once early literacy curriculum) completed, how many sessions students attended, the name of their tutor, and student/tutor absences.

Our primary outcome of interest is the end-of-year DIBELS composite score. The DIBELS 8th edition is used nationwide to assess and track progress in literacy skills, in addition to screening for learning disabilities such as dyslexia. The DIBELS composite score is made of subtests that measure grade level-appropriate skills, which are aligned with the early literacy skills targeted by the Once intervention and aligned to the sequence of elements incorporated into the science of reading (National Reading Panel, 2000). Classroom teachers administer each subtest, each of which last about 60 seconds in duration and test for specific skills like phonemic awareness, phonics, reading fluency, vocabulary, etc. At the beginning of the school year prior to the study, 56 percent of kindergarten through first grade students at the district overall performed at or above grade level on the DIBELS. Because raw DIBELS scores are always higher for older students and do not account for grade levels, we use standardized versions of these assessments, which accounts for grade level differences, in our analyses.

While 304 students were initially included in the study, one school pulled most of its students from the study (n=14 students) prior to program implementation. During program implementation, five students in the study left the school or district, also dropping from the study. At the end of the school year, 15 students had missing demographic and/or DIBELS data necessary for analysis. This left us with a final sample of 270 students for the main analysis, with 11.2% (n=34) of the original sample experiencing either scenario for attrition or having missing data.

---

[2] All schools participating in the pilot study were eligible for free and reduced price lunch, so we do not collect income data from students in the study.

Findings

Who Participated in the Study?

Table 1 shows the average demographic characteristics and achievement levels of the students in our study, and separately by program and comparison groups. Half the students in our study are female. The majority of students in our sample are Black (69 percent) and a quarter are Hispanic, approximately mirroring the characteristics of the district. About 13 percent and 23 percent of students are designated as receiving Special Education services or as English Learners. Kindergarten students in the study had a mean score of 267.8 on the fall DIBELS assessment while 1st grade students had a mean score of 313.7. As expected, given the eligibility criteria for the study, these scores are below the range within which we would expect students to be scoring at each respective grade level. Average student demographics and early literacy scores are similar between the program and comparison groups.

Table 1: Descriptive Statistics and Balance on Baseline Demographics and Achievement

	Comparison			Program			All Students			
	Mean	SD	N	Mean	SD	N	Mean	SD	N	p-value
Sex: Female	0.47		171	0.54		99	0.50		270	0.331
Race: Black	0.65		171	0.75		99	0.69		270	0.114
Race: Hispanic	0.26		171	0.23		99	0.25		270	0.575
Race: Other	0.02		171	0.00		99	0.01		270	0.126
Race: White	0.06		171	0.02		99	0.04		270	0.142
SPED	0.14		171	0.10		99	0.13		270	0.35
English Learner	0.24		171	0.22		99	0.23		270	0.745
Beginning of year DIBELS Composite Score										
K (Standardized)	0.00	(0.97)	120	0.06	(1.06)	87	0.02	(1.01)	207	0.671
1st (Standardized)	-0.02	(0.99)	51	-0.14	(1.05)	12	-0.04	(0.99)	63	0.713
K (Raw)	267.29	(19.32)	120	268.49	(21.12)	87	267.80	(20.05)	207	0.671
1st (Raw)	313.92	(9.07)	51	312.83	(9.66)	12	313.71	(9.11)	63	0.713
Beginning of year iReady Math Score										
K (Standardized)	0.00	(0.97)	118	-0.14	(0.97)	85	-0.06	(0.97)	203	0.32
1st (Standardized)	-0.02	(1.01)	51	-0.56	(0.96)	12	-0.12	(1.01)	63	0.093
K (Raw)	330.53	(17.88)	118	327.99	(17.98)	85	329.46	(17.92)	203	0.321
1st (Raw)	354.67	(20.28)	51	343.67	(19.34)	12	352.57	(20.42)	63	0.093
Observations			171			99			270	

Notably, the district provided additional tutoring services *beyond* the Once program, which included programs from 7 district-approved tutoring providers as well as other tutoring provided by the district or school staff. Many students in the study received other types of tutoring in addition to Once. Figure 1 shows the number of students in each group who received some type of tutoring, including Once. All but one student in the program group received tutoring, while 44 students (more than a quarter) of the comparison group received some type of tutoring service from a district-approved provider.

Figure 1. Count of Students Who Received Tutoring in Program and Comparison Groups
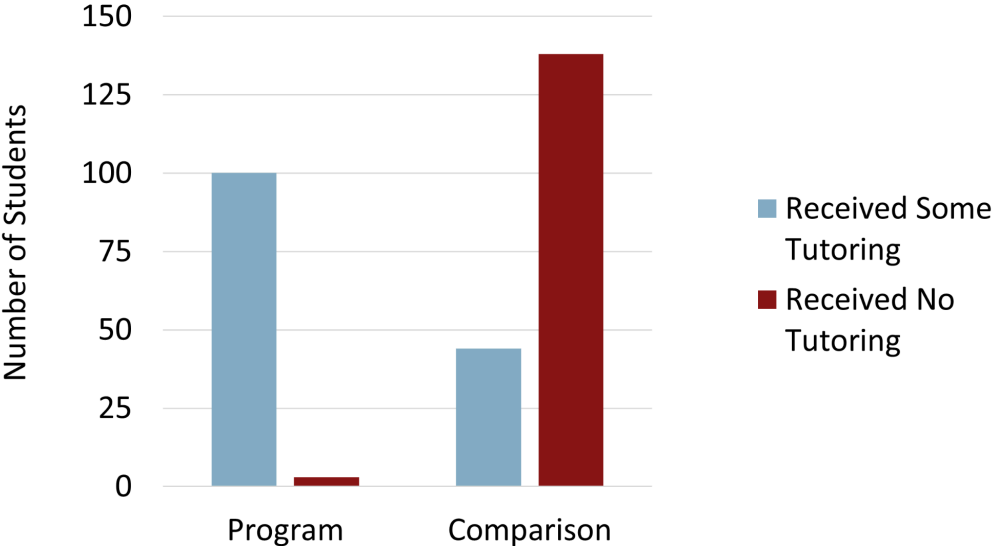

How Much Tutoring Did Students Receive?

Because implementation occurred later than planned, students received fewer sessions than planned. The expectation was that students should receive 140 daily sessions out of 180 possible school days in a year; however, the actual number of sessions received varied greatly among the students in the program, ranging from 0 to 113 sessions. On average, students received 42 sessions, which translates to 10.5 hours of 1:1 literacy instruction throughout the school year.

Implementation varied greatly across schools. Figure 2 visualizes the average number of sessions that students in the program group received at each school (vertical axis), and the number of students tutored at each school (horizontal axis). Most schools had between 6 to 12 students assigned to receive tutoring. The number of sessions offered on average differed across schools, but all schools fell far below the ideal of 140 sessions for the study year.

Figure 2: Scatterplot of Mean Sessions per School by Number of Students in the Program
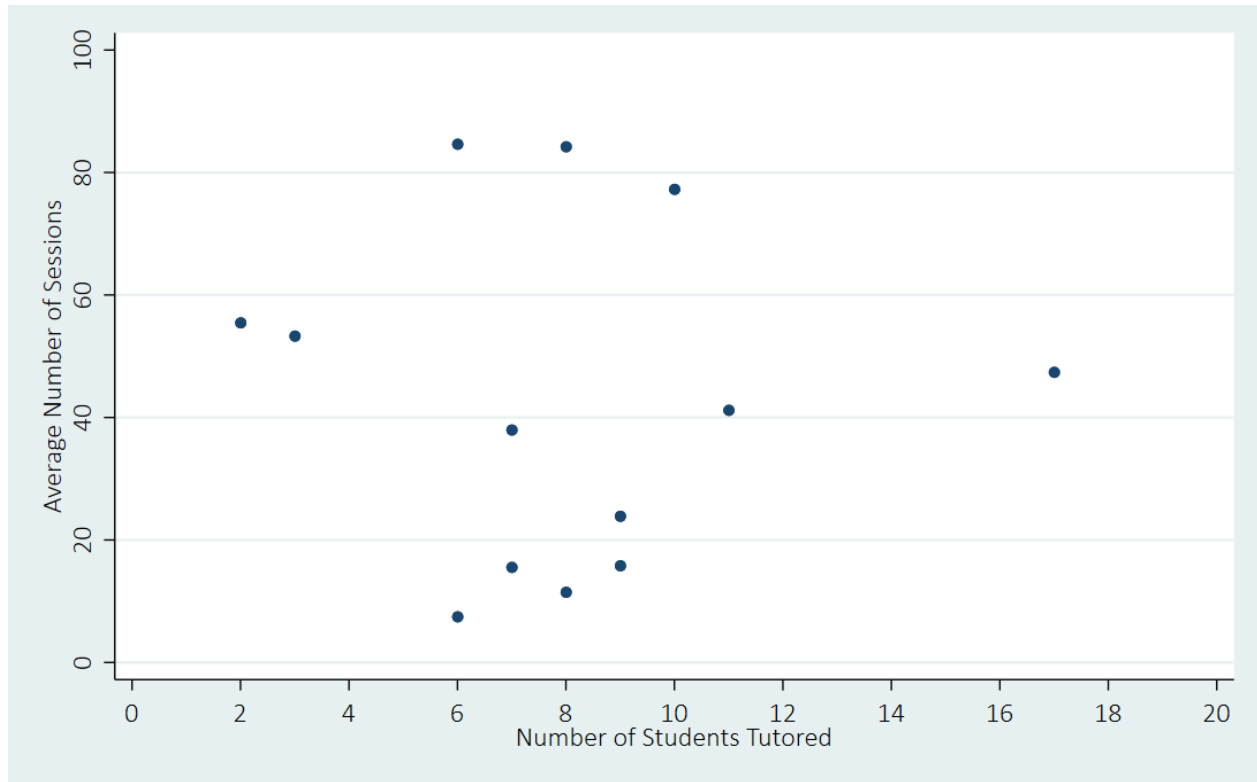

Note: Average Number of Sessions is derived by taking the average of the number of sessions received by all students at each school.

Underserved students and students who typically may need additional services received more Once tutoring on average than their counterparts. As shown in Figure 3, students designated as receiving Special Education services attended six more tutoring sessions relative to non-Special Education students, and Students classified as English Learners attended 19 more tutoring sessions than non-English Learners. Hispanic/Latino students and male students also attended more tutoring sessions compared to White and female students, respectively.

Figure 3. Average Number of Tutoring Sessions Completed by Student Characteristic
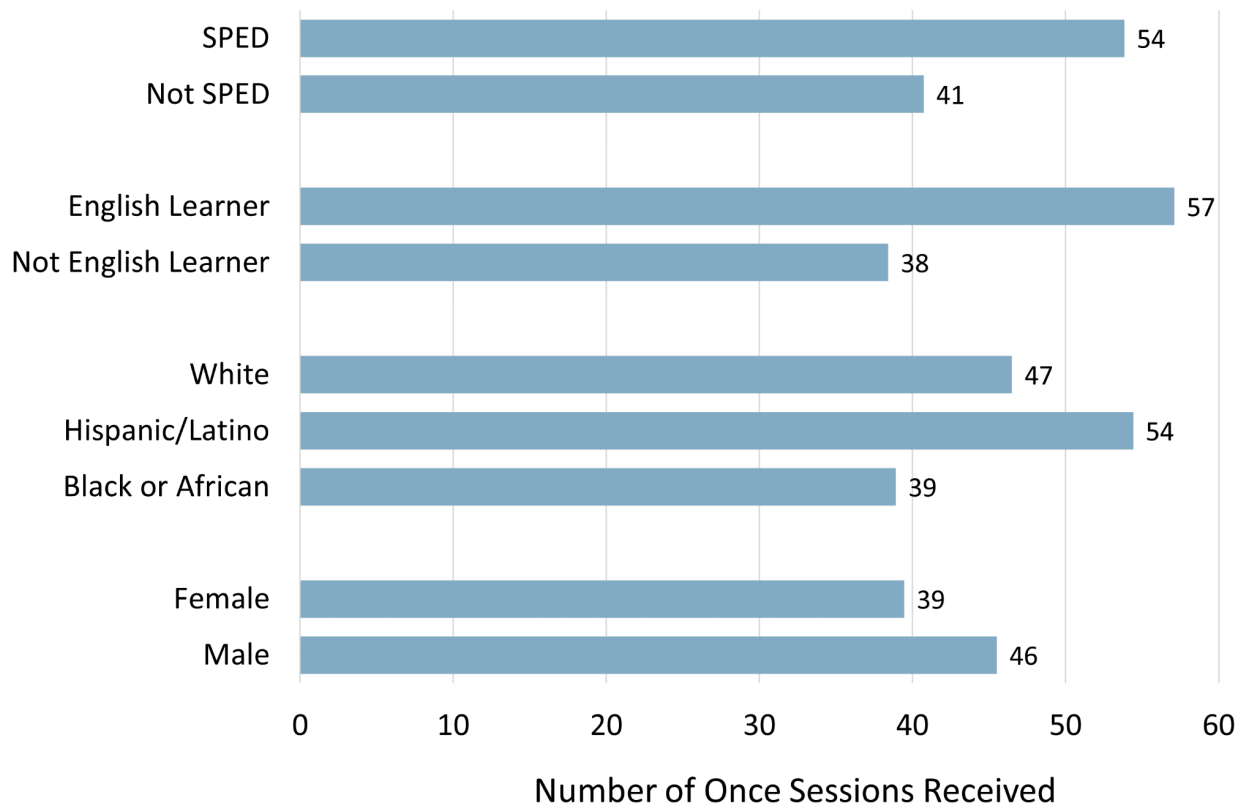

Did Once Tutoring Improve Students' End-of-Year Early Literacy Skills?

Participation in Once tutoring is associated with a 1-point increase in students' DIBELS composite scores, accounting for baseline achievement, student demographics, and classroom characteristics.[3] However, we do not have a sufficient sample size to detect whether the program group performed better than the comparison group as a result of tutoring.

---

[3] For a full set of iterative multivariate regression models demonstrating our analyses, see Appendix Table 1.

**Figure 4. Mean DIBELS Scores by Intervention Group**
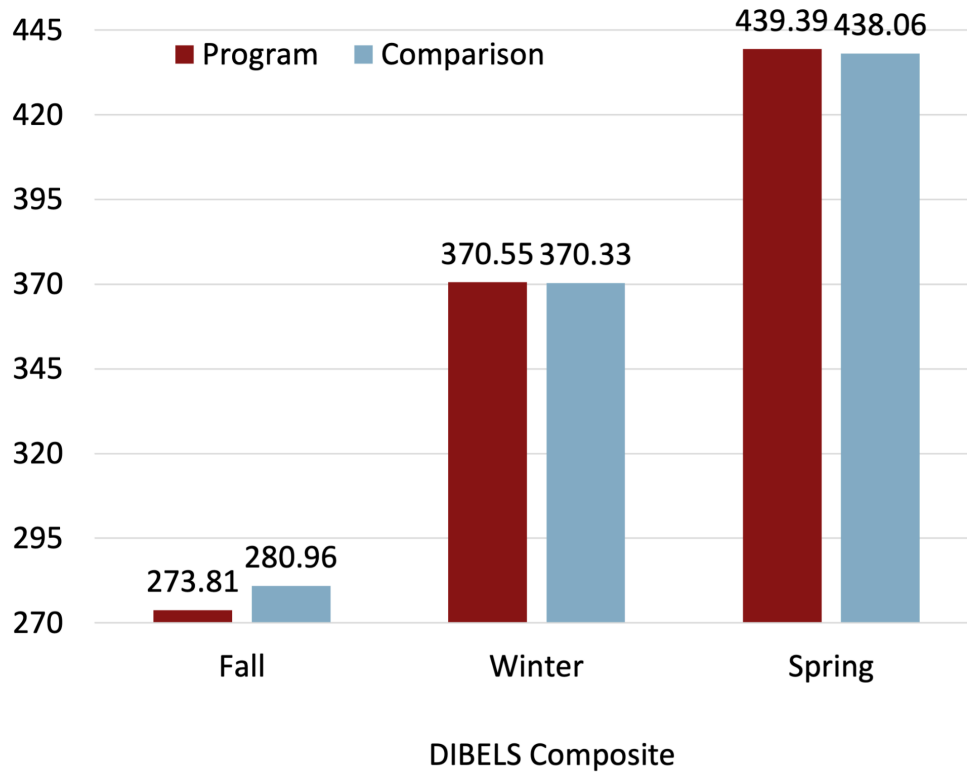
DIBELS Composite

Note: Each bar in the figure represents the average DIBELS score for each time point measured in the year (Fall, Winter, or Spring) by intervention group. Standardized values are within grade level. Total N=290, 278, and 273 respectively for Fall, Winter, and Spring assessments.

In addition to comparing average differences between the program and comparison groups, we assessed potential differences in the effect of the program for different students. Below, we report estimates of the effect of Once tutoring for students by grade level; baseline reading scores; and gender.[4]

Grade level: Both kindergarten and first grade students who received tutoring ended the year with average DIBELS scores that were greater than those in the comparison groups. These estimates are positive, but, again, are not statistically distinguishable from zero. Appendix Table 2 contains full regression analysis results separately by grade level. Figures 5 and 6 show average DIBELS gains for the program and comparison group students in kindergarten and first grade.

---

[4] We also conducted multivariate regression analyses to examine potential differences in outcomes by student race/ethnicity, differences when we examine DIBELS subtests as outcomes, and potential spillover effects of the tutoring affecting math scores. Additionally, we examined the potential effect of tutoring on students' attendance and chronic absenteeism rates. None of these models yielded meaningful results, and we omit discussing these for conciseness in this report.

Figure 5. DIBELS Gains by Intervention Group: Kindergarten Only
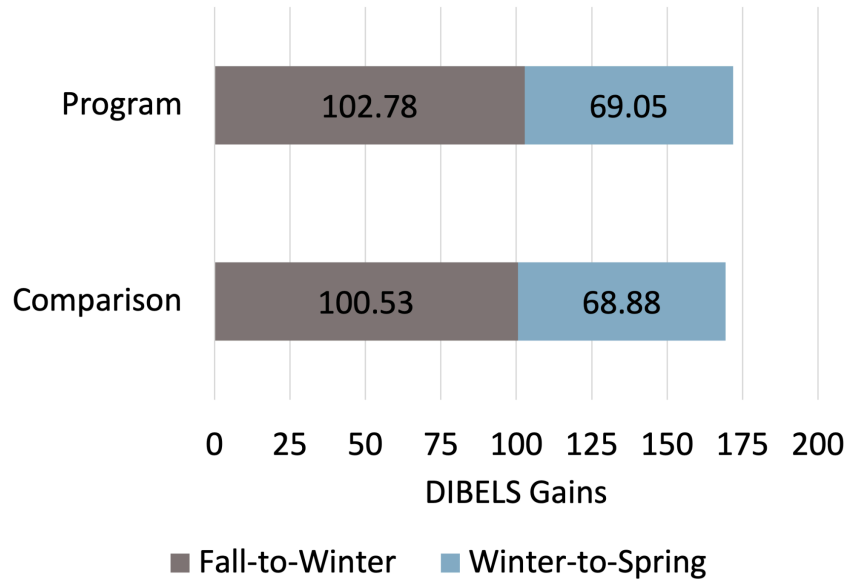
Note: Each bar in the figure represents the average DIBELS gains between two time points measured in the year (Fall to Winter, or Winter to Spring) within each of the experimental conditions. Total N=222, 212, and 209 respectively for kindergarteners in each testing period.

Figure 6. DIBELS Gains by Intervention Group: First Grade Only
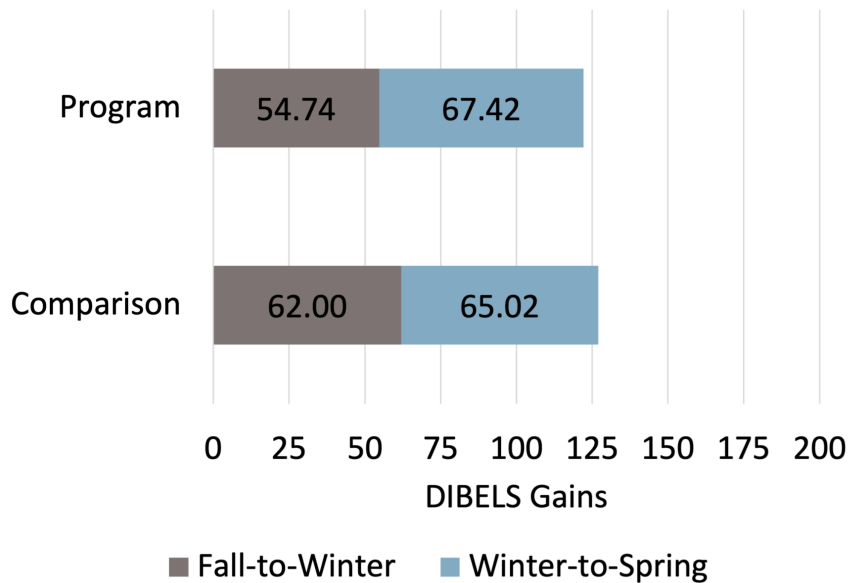
Note: Each bar in the figure represents the average DIBELS gains between two time points measured in the year (Fall to Winter, or Winter to Spring) within each of the experimental conditions. Total N=68, 66, and 64 respectively for first graders in each testing period.

Baseline Skills: We also examined the effect of Once tutoring separately by baseline DIBELS assessment score, splitting the sample into two subgroups: one consisting of students who scored in the Well Below Benchmark level – the lowest possible score band – in the fall DIBELS, and the other consisting of students who scored in the Below Benchmark level – second lowest of four possible score bands – at that time. Figures 7 and 8 show average DIBELS gains for the program and comparison group students by baseline skill level.

Generally, students who received tutoring made larger gains by the end of the year relative to students who did not receive tutoring, regardless of where they scored at baseline. Students who scored at Well Below Benchmark at baseline and received tutoring made particularly large gains. In all cases, the estimated program effect is not statistically distinguishable from zero due to small sample sizes (see Appendix Table 3). However, this provides early evidence that the positive learning results might be driven by students with the least developed early literacy skills at baseline.

Gender: Figures 9 and 10 show average DIBELS scores by time period administered and intervention group for male and female students, respectively. Male students started the school year at a slightly lower average baseline score, but scored at or above where female students scored by the end of the year. Among male students (n=136), those in the group who received tutoring experienced a 172-point gain, about 14 points higher than the gains made by male students in the comparison group. Female students (n=134) experienced a 160-point gain as a result of tutoring, which is about five points higher than the gains made by female students in the comparison group. Students in the comparison group make similar gains to one another (155-158 points) regardless of gender. See Appendix Table 4 for full results using standardized scores. While, again, subsample analyses are not statistically significant, we do observe robust and consistent evidence in other analyses (as seen in Appendix Tables 1, 2, and 3) that male students made noticeably larger gains, on average, than female students during the year.

Figure 7. DIBELS Gains by Intervention Group: Well Below Benchmark at Baseline
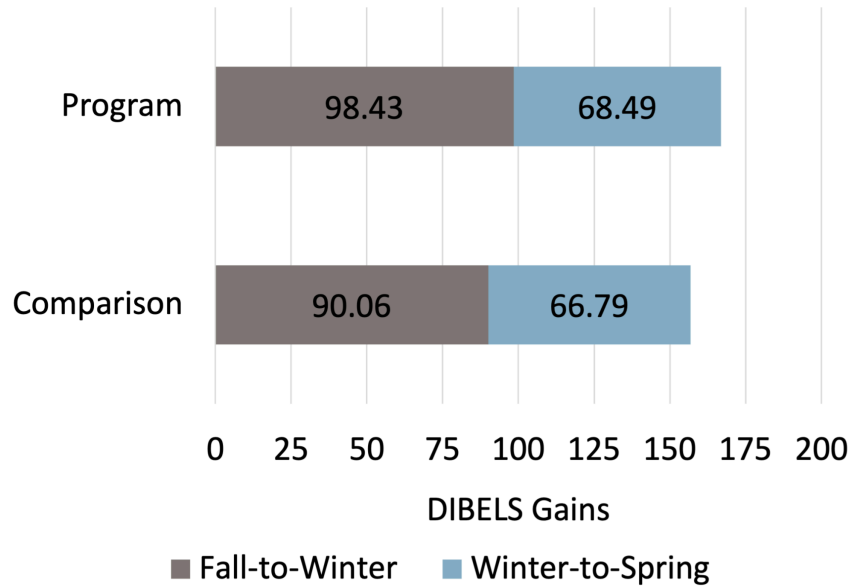
Note: Each bar in the figure represents the average DIBELS gains between two time points measured in the year (Fall to Winter, or Winter to Spring) within each of the experimental conditions. Total N=207, 196, and 195 respectively for those scoring at Well Below Benchmark at baseline in each testing period.

Figure 8. DIBELS Gains by Intervention Group: Below Benchmark at Baseline
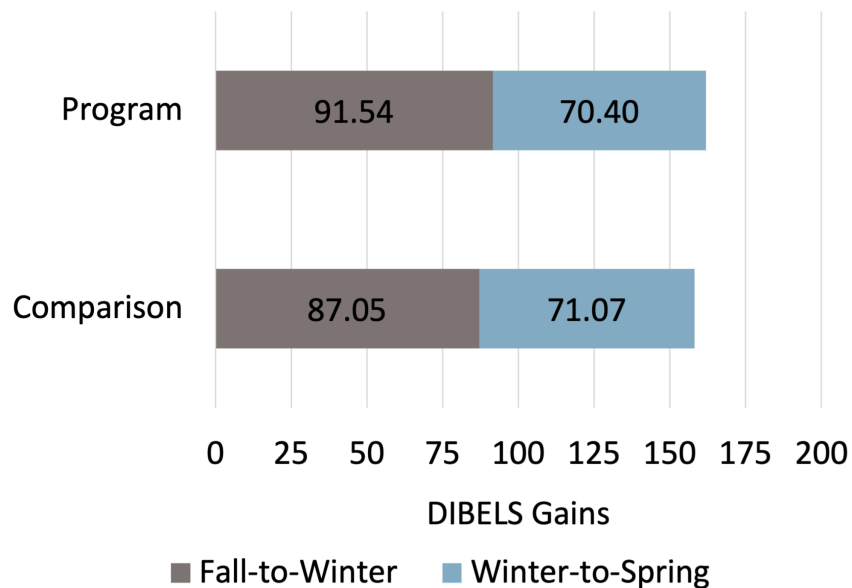
Note: Each bar in the figure represents the average DIBELS gains between two time points measured in the year (Fall to Winter, or Winter to Spring) within each of the experimental conditions. Total N=83, 82, and 78 respectively for those scoring at Below Benchmark at baseline in each testing period.

Figure 9. Mean DIBELS Scores by Intervention Group: Male Students
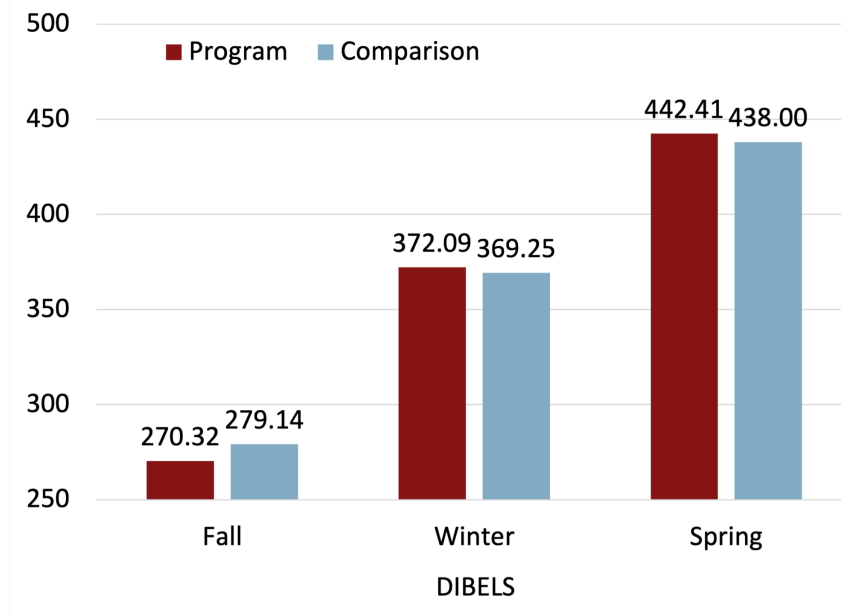
Note: Each bar in the figure represents the average DIBELS score for each time point measured in the year (Fall, Winter, or Spring) for by intervention group. Standardized values are within grade level. Total N=146, 140, and 138 respectively for Fall, Winter, and Spring assessments for boys.
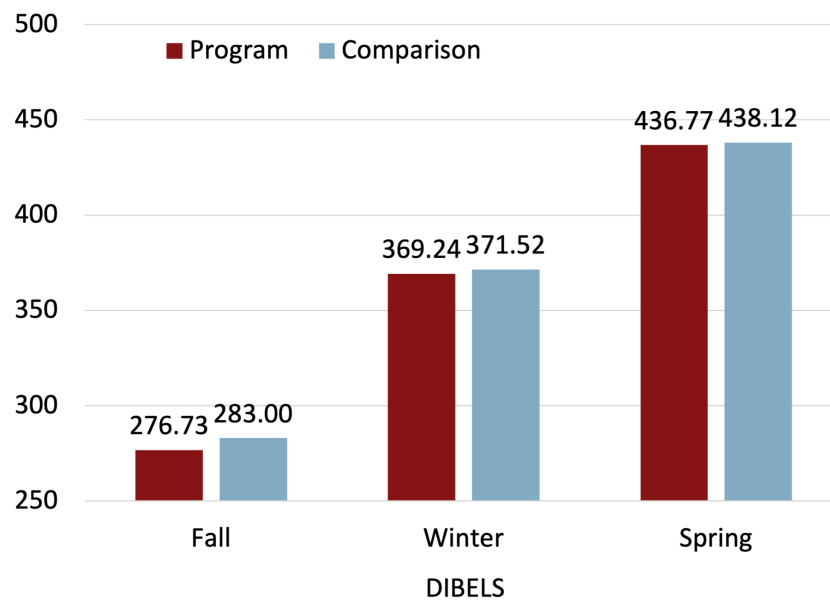
Figure 10. Mean DIBELS Scores by Intervention Group: Female Students

Note: Each bar in the figure represents the average DIBELS score for each time point measured in the year (Fall, Winter, or Spring) for by intervention group. Standardized values are within grade level. Total N=143, 138, 135 respectively for Fall, Winter, and Spring assessments for girls.

School: We observe some variation in the effectiveness of Once tutoring on end-of-year DIBELS scores across the schools that participated in the study (Figure 11; Appendix Table 5). However, none of the estimates are statistically significant to the extent that we can rule out the possibility that the differences may have occurred by chance.

Figure 11: Effect of Once Tutoring on End-of-Year DIBELS Scores by Participating School
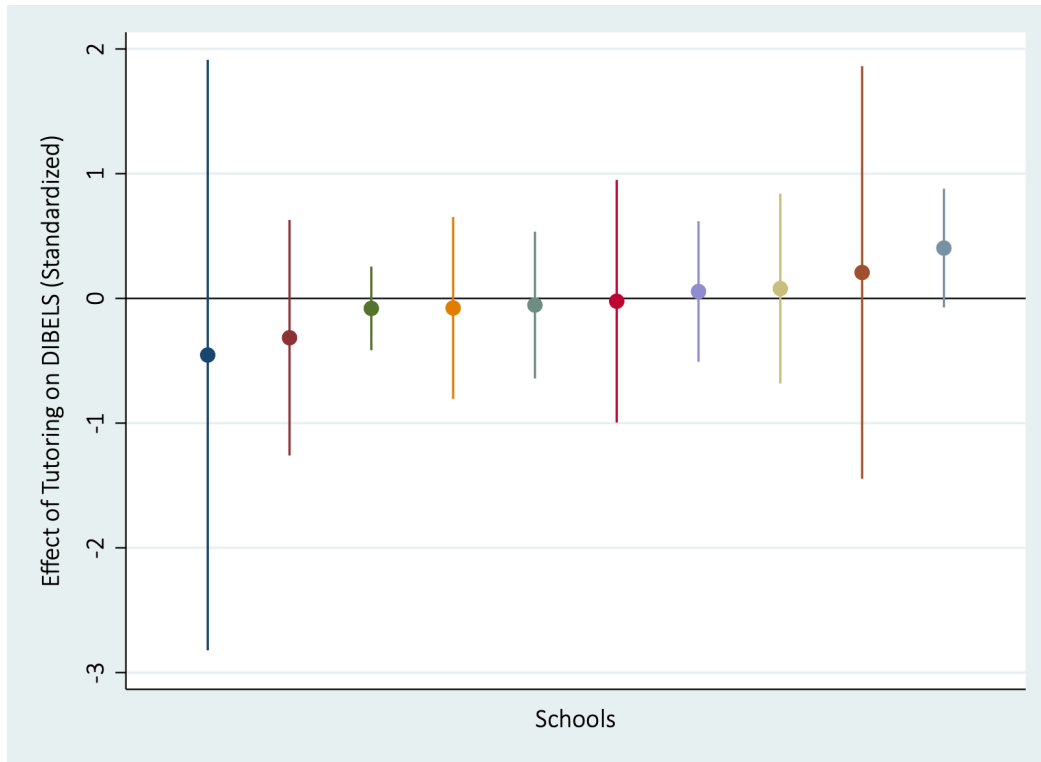
Note: Three schools are omitted from display due to calculation errors resulting from small sample size issues.

Takeaways

While this pilot study report did not yield statistically significant results, it does provide several notable takeaways.

First, many of the students who received Once tutoring attended traditionally low-performing and under-resourced schools. All but one of the schools providing tutoring were certified for the Community Eligibility Provision (CEP), a program that allows high-poverty schools to provide meals for free to all of its students without requiring individual families to enroll. Participating students primarily came from low income families and identified as Black or Hispanic/Latino. Over a quarter of these students qualified to receive Special Education services, a frequency which is uncommonly high, especially at these early grade levels. Additionally, more than one in ten students were classified as

English Learners. All students had assessment scores indicating early literacy skills at levels below the expectations for their grade level.

Second, while students in the study received far less tutoring – 42 15-minute sessions, on average – than the goal of 140 sessions, students with additional needs for learning support (such as those designated for Special Education or classified as English Learners) received more sessions than other students. This pattern is notable because the study began later than initially planned, and many paraprofessionals who were assigned to be Once tutors were pulled to conduct other high-priority duties.

Most importantly, while the study was a pilot and not large enough to detect significant differences between the program and comparison groups, it does provide indications that Once tutoring could lead to positive impacts on students' early literacy skills. Students in the program group, particularly male students and students scoring far below average on initial assessments, scored higher on the end-of-year assessments than their peers in the comparison group. Continued research is needed to understand whether these differences were caused by the program.

Overall, the average fall-to-spring gains in DIBELS scores observed in kindergarten and first grade students in the study are on par with or greater than what we observe in other research sites, indicating that all students (regardless of their participation in Once tutoring) made progress in their early literacy skills. District data suggests that schools used several tutoring providers to provide early literacy tutoring to support students, which may have led to these gains though obfuscated our ability to investigate the causal effect of Once.

References

Brown, K. J., Morris, D., & Fields, M. (2005). Intervention after grade 1: Serving increased numbers of struggling readers effectively. *Journal of Literacy Research*, *37*(1), 61–94. https://doi.org/10.1207/s15548430jlr3701_3

Cortes, K., Kortecamp, K., Loeb, S., & Robinson, C. D. (2023). Year two results assessing the effects of a scalable approach to high-impact tutoring for young readers. In *studentsupportaccelerator.org*. National Student Support Accelerator. https://studentsupportaccelerator.org/briefs/year-two-assessing-effects-high-impact-tutoring-young-readers

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

Harris, D. N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, *31*(1), 3–29. https://doi.org/10.3102/0162373708327524

Hemelt, S. W., Ladd, H. F., & Clifton, C. R. (2021). Do teacher assistants improve student outcomes? Evidence from school funding cutbacks in North Carolina. *Educational Evaluation and Policy Analysis*, *43*(2), 016237372199036. https://doi.org/10.3102/0162373721990361

Lee, M., Robinson, C. D., & Loeb, S. (2023). Effects of High-Impact Tutoring on Early Literacy Outcomes. The Open Science Framework. Retrieved from https://osf.io/zea9k

Loeb, S., Novicoff, S., Pollard, C., Robinson, C., & White, S. (2023). The effects of virtual tutoring on young readers: Results from a randomized controlled trial. In *studentsupportaccelerator.org*. National Student Support Accelerator. https://studentsupportaccelerator.org/briefs/effects-virtual-tutoring-young-readers

Markovitz, C. E., Hernandez, M. W., Hedberg, E. C., & Whitmore, H. W. (2021). Evaluating the effectiveness of a volunteer one-on-one tutoring model for early elementary reading intervention: A randomized controlled trial replication study. *American Educational Research Journal*, *59*(4), 000283122110668. https://doi.org/10.3102/00028312211066848

National Reading Panel. (2013). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. In *National Reading Panel* (Vol. 88, Issue 1, pp. 3–6). Eunice Kennedy Shriver National Institute of Child Health and Human Development. https://doi.org/10.1215/00031283-2322610

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2021). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, *57*(1). https://doi.org/10.1002/rrq.379

Pace Miles, K., Lauterbach, M. D., Murano, D. M., & Dembek, G. A. (2018). Reading Rescue: A follow-up on effectiveness of an intervention for struggling readers. *The Journal of Educational Research*, *112*(2), 255–269. https://doi.org/10.1080/00220671.2018.1514358

Robinson, C. D., & Loeb, S. (2021, May 7). *High-Impact tutoring: State of the research and priorities for future learning*. Edworkingpapers.com; Annenberg Institute for School Reform at Brown University. https://edworkingpapers.com/ai21-384

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

White, S., Carey, M., O'Donnell, A., & Loeb, S. (2021). *Early Lessons from Implementing High-Impact Tutoring at Scale*. National Student Support Accelerator; Stanford Graduate School of Education. https://studentsupportaccelerator.org/sites/default/files/Early%20Lessons%20from%20Implementing%20High-Impact%20Tutoring%20at%20Scale.pdf

Appendix

Appendix Table 1: Program Effect on End of Year DIBELS Score

	Standardized DIBELS				Raw DIBELS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program Effect	0.055	0.046	0.071	0.049	1.335	1.621	2.505	1.524
	(0.117)	(0.114)	(0.115)	(0.092)	(4.171)	(4.039)	(4.068)	(3.314)
Female			-0.101	-0.270**			-3.532	-8.084*
			(0.111)	(0.091)			(3.924)	(3.244)
Black			-0.407+	-0.004			-14.326+	-1.156
			(0.238)	(0.192)			(8.438)	(6.950)
Hispanic			-0.297	0.096			-10.827	3.379
			(0.252)	(0.234)			(8.921)	(8.466)
ELL				0.185				5.322
				(0.167)				(6.046)
Special Ed.				-0.285*				-10.299*
				(0.135)				(4.894)
Beginning of Year DIBELS				0.539***				-4.390*
				(0.065)				(2.058)
Beginning of Year Quadratic				-0.113				0.010*
				(0.069)				(0.004)
Minimum DIBELS Scorer				-0.094				-14.683*
				(0.177)				(5.696)
R-squared	0.001	0.319	0.331	0.589	0.000	0.323	0.334	0.573
Comparison Mean	0.000	0.003	-0.006	0.002	438.058	437.954	437.630	437.989
Observations	270	270	270	270	270	270	270	270
Classroom FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Standard errors in parentheses. Constant omitted from display. Reference category for race/ethnicity is White/Other Race. DIBELS is standardized within grade and time of year administered within sample. Minimum BOY DIBELS Scorer is an indicator that equals one if a student scored the minimum possible score on the BOY DIBELS.

+ $p<0.10$ * $p<0.05$ ** $p<0.010$ *** $p<0.001$.

Appendix Table 2: Program Effect on End-of-Year DIBELS Score (Standardized) by Grade

	Grade K				Grade 1			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program Effect	0.092	0.039	0.068	0.033	-0.218	0.089	0.065	0.094
	(0.131)	(0.120)	(0.121)	(0.101)	(0.302)	(0.332)	(0.335)	(0.238)
Female			-0.091	-0.243*			-0.137	-0.354+
			(0.124)	(0.105)			(0.250)	(0.187)
Black			-0.356	-0.076			-0.556	0.520
			(0.261)	(0.217)			(0.583)	(0.443)
Hispanic			-0.391	0.052			-0.153	0.429
			(0.283)	(0.276)			(0.586)	(0.479)
ELL				0.123				0.302
				(0.211)				(0.276)
Special Ed.				-0.401*				-0.123
				(0.169)				(0.232)
Beginning of Year DIBELS				0.465***				0.756***
				(0.078)				(0.127)
Beginning of Year Quadratic				-0.098				-0.064
				(0.080)				(0.153)
Minimum DIBELS Scorer				-0.151				-0.151
				(0.201)				(0.439)
R-squared	0.002	0.344	0.354	0.576	0.008	0.233	0.272	0.687
Comparison Mean	0.000	0.023	0.010	0.025	0.000	-0.058	-0.054	-0.059
Observations	207	207	207	207	63	63	63	63
Classroom FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Standard error in parentheses. Constant omitted from display. Reference category for race/ethnicity is White/Other Race. DIBELS is standardized within grade and time of year administered within sample. Minimum BOY DIBELS Scorer is an indicator that equals one if a student scored the minimum possible score on the BOY DIBELS.

+ p<0.10 * p<0.05 ** p<0.010 *** p<0.001.

Appendix Table 3: Program Effect on End-of-Year DIBELS Score (Standardized) by Baseline DIBELS Level

	Well Below Benchmark				Below Benchmark			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program Effect	0.123	0.109	0.096	0.111	-0.064	-0.091	-0.066	-0.103
	(0.136)	(0.136)	(0.136)	(0.115)	(0.167)	(0.172)	(0.158)	(0.159)
Female			0.003	-0.209+			-0.689***	-0.548**
			(0.134)	(0.115)			(0.182)	(0.184)
Black			-0.467	-0.014			0.345	0.263
			(0.347)	(0.296)			(0.263)	(0.271)
Hispanic			-0.162	-0.092			0.506	0.496
			(0.356)	(0.331)			(0.315)	(0.377)
ELL				0.379+				-0.139
				(0.217)				(0.319)
Special Ed.				-0.444**				0.701+
				(0.157)				(0.386)
Beginning of Year DIBELS				0.452**				2.595
				(0.170)				(1.703)
Beginning of Year Quadratic				-0.307				-0.886
				(0.222)				(0.700)
Minimum DIBELS Scorer				-0.086				0.000
				(0.217)				(.)
R-squared	0.004	0.377	0.392	0.592	0.002	0.527	0.659	0.728
Comparison Mean	-0.257	-0.252	-0.247	-0.252	0.622	0.631	0.623	0.636
Observations	193	193	193	193	77	77	77	77
Classroom FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Standard errors in parentheses. Constant omitted from display. Reference category for race/ethnicity is White/Other Race. DIBELS is standardized within grade and time of year administered within sample. Minimum BOY DIBELS Scorer is an indicator that equals one if a student scored the minimum possible score on the BOY DIBELS.

+ $p<0.10$ * $p<0.05$ ** $p<0.010$ *** $p<0.001$.

Appendix Table 4: Program Effect on End-of-Year DIBELS Score (Standardized) by Gender

	Male Students				Female Students			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program Effect	0.134	0.214	0.220	0.181	-0.013	-0.095	-0.069	-0.069
	(0.194)	(0.187)	(0.189)	(0.139)	(0.135)	(0.151)	(0.155)	(0.132)
Black			-0.098	0.308			-0.530	-0.290
			(0.346)	(0.255)			(0.380)	(0.321)
Hispanic			0.010	0.464			-0.332	-0.071
			(0.371)	(0.314)			(0.399)	(0.398)
ELL				-0.044				0.113
				(0.230)				(0.301)
Special Ed.				-0.653***				0.089
				(0.184)				(0.234)
Beginning of Year DIBELS				0.569***				0.469***
				(0.095)				(0.099)
Beginning of Year Quadratic				-0.111				-0.077
				(0.100)				(0.115)
Minimum DIBELS Scorer				-0.163				-0.170
				(0.245)				(0.310)
R-squared	0.004	0.469	0.470	0.738	0.000	0.324	0.342	0.575
Comparison Mean	0.006	-0.021	-0.023	-0.010	-0.006	0.026	0.016	0.016
Observations	136	136	136	136	134	134	134	134
Classroom FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Standard eros in parentheses. Constant omitted from display. Reference category for race/ethnicity is White/Other Race. DIBELS is standardized within grade and time of year administered within sample. Minimum BOY DIBELS Scorer is an indicator that equals one if a student scored the minimum possible score on the BOY DIBELS.

+ p<0.10 * p<0.05 ** p<0.010 *** p<0.001.

Appendix Table 5: Program Effect on End-of-Year DIBELS Score (Standardized) by School

	School A	School B	School C	School D	School E	School F	School G	School H	School I	School J	School K	School L	School M
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Program	1.736	0.208	0.079	0.055	1.517	-0.023	-0.454	-0.315	-0.053	0.403+	-0.077	-0.08	-0.732
	(.)	(0.676)	(0.366)	(0.264)	(2.558)	(0.442)	(0.744)	(0.367)	(0.293)	(0.214)	(0.298)	(0.158)	(.)
R-squared	1	0.585	0.571	0.645	0.379	0.559	0.851	0.556	0.656	0.675	0.796	0.848	1
Observations	3	18	33	23	10	21	15	12	65	22	14	27	7

Notes: Standard errors in parentheses. All other variables aside from program effect omitted from display. In the case of two schools, control variables included in main analyses (e.g., gender, race/ethnicity indicators) were automatically dropped from analyses due to small sample size and made the model a perfect fit, resulting in the lack of standard errors.

+ $p < 0.10$ * $p < 0.05$ ** $p < 0.010$ *** $p < 0.001$.