



# Testing the Impact and Scalability of the Future Forward Literacy Program

*Curtis J. Jones and DongMei Li, University of Wisconsin-Milwaukee  
March 2021*

## **Testing the Impact and Scalability of the Future Forward Literacy Program Executive Summary**

Future Forward (FF) is an early elementary literacy program that pairs one-on-one tutoring with parent engagement. Under the name of SPARK, in 2011, FF/SPARK was funded by an Investing in Innovations grant to test the impact in Milwaukee. Two randomized studies found FF/SPARK to have positive impacts on literacy, reading achievement, and school attendance (Jones, 2018; Jones & Christian 2020). In 2017, Education Analytics (EA) received an Education Innovation and Research Mid-Phase grant from the U.S. Department of Education to expand FF to 14 schools in seven school districts across three states. The randomized control trial study of the 2019-20 FF program was designed to examine the effectiveness of its scale-up to 14 schools and its impact on reading achievement, social-emotional learning (SEL), and school attendance. Regarding scale-up, nearly all participating students received one-on-one tutoring as intended. Fewer students received the full implementation of family engagement (FE) however. In spring of 2020, schools were shut down nationwide because of the COVID-19 pandemic. The disruption on schools limited our study to testing the impact of FF on school attendance only. FF was found to have a statistically significant, positive impact on school attendance. The school attendance of FF participants was 1.3 percentage points higher than students who received Business-as-Usual (BAU) literacy instruction. Further, its impact was greater with Black students, male students, and students with a history of low school attendance. Students at the intersection of these groups (Black male students with low school attendance) were particularly affected, improving their school attendance by six percentage points and 4.2 fewer days absent from school.

# BACKGROUND

Future Forward (FF) is an early elementary literacy program, administered by Education Analytics, that leverages community agencies and combines one-on-one intensive tutoring with parent engagement to promote students' literacy development both at school and at home. Under the name of SPARK, in 2011, FF/SPARK was funded by an Investing in Innovations (i3) grant to test its impact in Milwaukee. Two randomized control trial (RCT) studies found the literary program to have positive impacts on literacy, reading achievement, and school attendance (Jones, 2018; Jones & Christian 2020). In 2017, Education Analytics (EA) received an Education Innovation and Research (EIR) Mid-Phase grant from the U.S. Department of Education to expand FF to 14 schools across seven school districts in three states during the 2018-19 school year, and then to test its impact in 2019-20 and 2020-21.

Within FF program sites, the site manager, who is typically a certified teacher, manages personalized one-on-one tutoring provided by paraprofessionals or volunteers. The family engagement (FE) coordinator, who is typically a community member or parent from the school, is responsible for family outreach and communication. These typically involve monthly family events and ongoing contacts. FF sites are expected to provide students a minimum of two 30-minute tutoring sessions per week and to communicate with participating families at least two times per month through in-person meetings, phone call conversations, or email or text conversations. Students participate for one school year. Adequate implementation means at least 80% of students within a program site meet these expected tutoring and family contact minimums. A fuller description of the Future Forward program model has been published elsewhere (Jones & Christian, 2020).

## Previous FF Research/Evaluation

Three impact studies of FF (previously named SPARK) have been conducted as part of the i3 and EIR grants (Table 1). Each of these three studies meets the Institute for Education Sciences What Works Clearinghouse (WWC) evidence standards *Without Reservations* (WWC, 2020). The first RCT study, funded by i3, was a pilot evaluation conducted as the program was still being developed in six Milwaukee schools during the 2011-12 and 2012-13 school years (Jones, 2018). While SPARK/FF had a small but significant impact on reading of 0.12 standard deviations, no significant impact was found on school attendance. However, because the FE component was being designed, and therefore was not fully implemented, the study did not test the impact of the FE component as it was designed (Jones, 2018).

The second RCT study, also funded by i3, took place in seven MPS campuses during the 2013-14 and 2014-15 school years. This study found positive and statistically significant impacts on literacy development and school attendance (Jones & Christian, 2020). Specifically, after two years of tutoring, participants' literacy assessment scores improved by 0.23 standard deviations. Further, SPARK/FF students were absent from school 4.5 fewer days than students in the BAU condition. While SPARK did not have a significant impact on reading achievement after two years of tutoring, the impact after one year, which was less affected by attrition, was statistically significant and positive ( $\beta = 0.11$ ,  $p = 0.048$ ).

# BACKGROUND

This study further found the impact of SPARK/FF was the greatest for students with the greatest need for literacy help. Related, implementation of FF was strong, with 96% of students receiving the intended amount of tutoring and 98% receiving the intended amount of family engagement. Thus, the results of the study are a truer test of the impact of SPARK/FF as it was designed.

Because of the positive results found in the i3 evaluations of FF/SPARK, EA was awarded an EIR grant to test the scalability of FF to a larger number of schools. From this grant, the third study of FF used a regression discontinuity (RD) design to measure its impact on student literacy during the first full year of the grant (2018-19 school year) in 12 schools across five school districts (Jones et al., 2020).<sup>1</sup> Although this study did not find a statistically significant positive impact ( $\beta = 0.16$ ,  $p = 0.394$ ), low statistical power and low fidelity of implementation limited the study's ability to measure an impact. During this year, only 34.7% of participants received the intended amount tutoring and 6.6% received the intended amount of FE.

**Table 1: Research and evaluation of SPARK/FF**

	Period	Design	Analytic Sample	Outcome	Impact Estimates	
Jones, 2018	2011-12	RCT	251 participants	Reading	0.12**	significant
	2012-13		245 control students	Absence	-3.33*	not significant
Jones & Christian, 2020	2013-14	RCT	286 participants	Literacy	0.23**	significant
	2014-15		290 control students	Absence	-4.53*	significant
			Reading	0.10**	not significant	
Jones et al., 2020	2018-19	RD	121 participants*** 141 control students	Literacy	0.16**	not significant

\*Attendance impact estimates represent the number of fewer days absent SPARK/FF students were from school. For instance, in the Jones & Christian study, SPARK/FF students missed 4.53 fewer days of school than BAU students.

\*\*Literacy and reading impact estimates are standardized.

\*\*\* The total sample for the RD study is 979 with 262 students included in the main RD analysis.

<sup>1</sup> Two schools did not participate in the evaluation.

The current evaluation, also part of the EIR grant, was again designed to test the scalability of FF. Scalability was assessed through an analysis of implementation and by measuring its impact on reading achievement, school attendance, and social-emotional learning (SEL) in 14 schools during the 2019-20 school year. However, in spring of 2020, schools were shut down nationwide because of the COVID-19 pandemic and instruction moved online. This prevented analyses of achievement and SEL. However, the study was still able to measure the impact of FF on school attendance.

## Research Questions

In the present study, we explore the implementation of FF to address the following question:

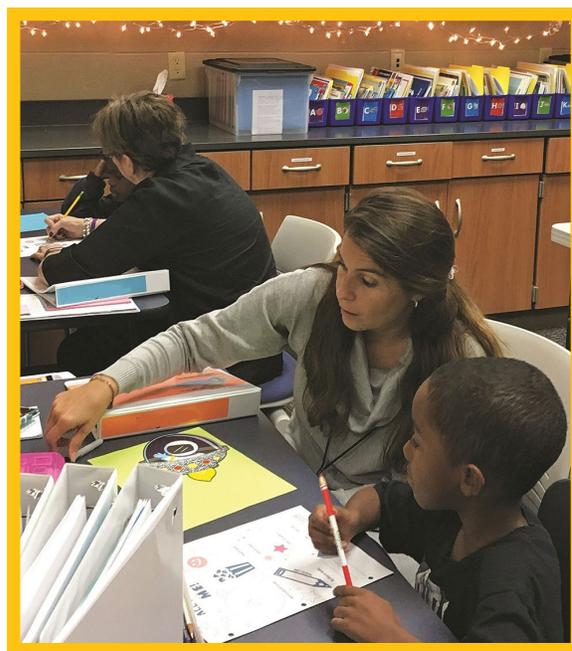
*How well was FF implementation scaled up to 14 schools across three states?*

We also use an RCT design, with students randomly assigned within blocks, to address the following confirmatory research question.

*What is the impact of one year of FF participation on regular-school-day attendance of lower primary students compared to students receiving BAU literacy instruction?*

And the following exploratory research question:

*Did FF have a differential impact on student subgroups?*



## Evaluation Methods

This evaluation study utilized an RCT design, with kindergarten, first grade, and second grade students randomly assigned to receive FF literacy programming or BAU literacy instruction. These students come from 14 schools across three states (Table 2): nine schools are in the state of Wisconsin, two are in Alabama, and the remaining three are in South Carolina. These 14 schools partnered with six local Boys & Girls Clubs. Seven schools are within large, urban districts. The remaining seven schools are located in small, rural communities. Participating schools had a history of overall literacy performance that placed them in the lowest 20% of schools in their state or had a history of large reading achievement gaps between races or economic groups. Overall, 58.7% of the study participants were Black and 84.3% were eligible for free or reduced lunch (Table 2). Eight of the study schools served a large number of Black students. The great majority of students in all but one school were eligible for free or reduced lunch.

**Table 2: Participating Schools**

	State	Community Type	Percent Black Students*	Percent White Students*	Free/Reduced Lunch Eligibility
School 1	WI	Urban	97.4%	0.0%	89.5%
School 2	WI	Urban	89.7%	7.7%	84.6%
School 3	WI	Rural	0.0%	86.7%	63.3%
School 4	WI	Rural	6.9%	82.8%	86.2%
School 5	AL	Urban	61.5%	23.1%	71.8%
School 6	SC	Rural	54.1%	40.5%	100%
School 7	WI	Rural	4.2%	91.7%	72.9%
School 8	AL	Urban	31.6%	50.0%	39.5%
School 9	WI	Rural	2.2%	88.9%	71.1%
School 10	WI	Urban	97.5%	0.0%	92.5%
School 11	WI	Urban	97.3%	0.0%	97.3%
School 12	WI	Urban	97.5%	0.0%	100%
School 13	SC	Rural	94.7%	5.3%	100%
School 14	SC	Rural	72.5%	26.1%	100%
<b>Overall</b>			<b>58.7%</b>	<b>35.3%</b>	<b>84.3%</b>

*\*The third racial group, "other," is omitted here.*

### ***Informed Consent***

In the fall semester of 2019, informed consent was obtained from parents for their students to participate in the study. Method of consent (passive vs. active) varied by school, with passive consenting used in half of the schools and active consenting in the remaining half. The consenting process lasted from September to November 2019.

### ***Study Eligibility***

Eligible students were in kindergarten, first, and second grade without an Individualized Education Plan (IEP) and were not English Language Learners (ELL). The specific numbers of students who were ineligible is not known because schools were instructed to not distribute consent forms to students who did not meet eligibility criteria.

## ***Random Assignment***

In the fall of 2019, 293 students were randomly assigned to the FF participant group and 294 to the BAU group. Assignment was done within blocks, defined as grade levels within schools (one grade per school is one block). Considering the three grade levels, 14 schools involved in the study, and that two schools did not have kindergarten students, the study included a total of 40 assignment blocks (3 grade levels x 12 schools + 2 grade levels x 2 schools = 40). Students within each block were ranked by their baseline reading assessment results, from lowest to highest. The number of study participants in each block was twice the capacity of the program to serve. Of these, half were randomly assigned to FF and the other to BAU reading instruction within each block. The number of study participants per block ranged from 6 to 33, with an average of 14 per block, assuming a fixed program effect, and 40% of the variance in outcomes explained by covariates, the current study, prior to attrition, had an 80% likelihood of detecting an impact of 0.184 standardized units.

## ***Instruments***

EA collected and shared program implementation data. Participating schools provided all other data directly to the research team.

**School Attendance.** School attendance was measured twice, during the program (from December to March) and prior to the start of the program (from September to December). Attendance rates were computed by dividing the total attended days by the total days of school during that time.

**Reading/Literacy Assessments.** Reading achievement assessments administered at baseline included the Phonological Awareness Literacy Screening (*PALS*), the MAP reading assessment for primary grade (*MPG*), the Dynamic Indicators of Basic Early Literacy Skills (*DIBELS*), and the Formative Assessment System for Teachers (FAST) - FastBridge reading assessments. Assessments were intended to be administered again at the end of the school year, but COVID-19 forced schools nationwide to be closed in March of 2020.

The *PALS*, used by seven of 14 schools, is a criterion-referenced, teacher-administered assessment of foundational literacy (Invernizzi et al., 2003). The assessment's internal reliabilities range from 0.76 to 0.83, inter-rater reliabilities are 0.92, and test-retest reliabilities are between 0.92 and 0.96 (Invernizzi et al., 2015). The assessment also has strong evidence of predictive validity (Invernizzi et al., 2004).

The *MPG* was used by three schools. MPG is a norm-referenced assessment of reading achievement and its measures of reliability and validity of the MAP test are high (NWEA, 2009). The reliability ranges from 0.70 to 0.90 and the predictive validity lies between 0.65 and 0.85.

The *DIBELS*, used by two schools, refers to five measures that assess the reading skills of K-8 students (Center on Teaching and Learning, 2018). The one-minute short measures have been "thoroughly researched and demonstrated to be reliable and valid indicators" (DIBELS, 2021). Three measures have

excellent alternate form reliability of 0.90+, and two measures have good reliability of 0.80+ (Center on Teaching and Learning, 2018). The composite score, which was used in the current study, combines the results from the individual assessments (University of Oregon, 2020). Composite score test-retest reliability is high, ranging from 0.70 to 0.93 depending on the grade level and form used. It also has high concurrent and predictive validity with the Iowa Test of Basic Skills (University of Oregon, 2018-2020).

The *FastBridge* reading assessment, used by two schools, adopts Computer-Adaptive Tests to screen student achievement and Curriculum-Based Measures to monitor growth over time. It uses norm-referenced standards, providing grade-level norms for the class, school, district and nation (Christ & Colleagues, 2015). *FastBridge* shows evidence of validity and “reliability coefficients that show promise for producing little test error” (Christ & Colleagues, 2015, p.20).

**SEL Assessments.** Baseline SEL was assessed with the Devereux Student Strengths Assessment Mini (*DESSA-Mini*) and Behavioral Emotional Social Traits (*b.e.s.t.*) tests. These were not administered at the end of the school year due to the nationwide school shutdowns in the spring of 2020.

Used by 13 schools, the *DESSA-Mini* is a strength-based eight-item behavior rating scale that can be completed in one minute by teachers (Shapiro et al., 2016). Its internal reliability exceeds 0.90. Classification consistency between the *DESSA-Mini* and the *DESSA* full assessment is 87%–94% in routine practice, with strong predictive validity of achievement (Shapiro et al., 2016) and school attendance (Kilpatrick, Maras, Brann, & Kilgus, 2018). Inter-rater reliability was reported as between 0.70 and 0.80 (Naglieri, et al., 2014, as cited in Shapiro et al., 2016).

The *b.e.s.t.*, used by one school, is an online universal screening platform for educators to identify and match interventions for externalizing (conduct) and internalizing (personality) behaviors (Hartwig & Hayes, 2019a). It possesses appropriate test-retest reliability with 0.90 correlation coefficient for the G-scale, or composite score, and high internal consistency of 0.95 (Hartwig & Hayes, 2019b). It also has predictive validity identifying students who later present with emotional disorders (Hartwig & Hayes, 2019b). Although no specific evidence is available that links *b.e.s.t.* scores to attendance, its relationship with emotional disorders suggests there likely is one.

## Modeling Strategy

Generalized linear models (GLM), which use maximum likelihood estimation, with linear error terms and an identity link function, were used to estimate the impact of FF on regular-school-day attendance. Both baseline and follow-up attendance, baseline reading achievement, and baseline SEL data were standardized within grade levels. The IBM SPSS 26.0 statistical software package was used to conduct analyses.

The outcome variable was modeled using the following linear regression equation:

$$(1) \quad Y_{ij} = \beta_0 + \beta_1(FF_{ij}) + \beta_2(S_{ij}) + \beta_3(M_{ij}) + \beta_4(A_{ij}) + \beta_{5a}(P_{ij} * PALS) + \beta_{5b}(P_{ij} * MPG) + \beta_{5c}(P_{ij} * DIBELS) + \beta_{5d}(P_{ij} * FAST) + \sum_{m=1}^M \beta_{6,m}X_{mij} + \sum_{j=1}^{J-1} \beta_{7,j}Block_j + \varepsilon_i$$

Where  $Y_{ij}$  is the outcome (attendance) for the  $i^{th}$  student in the  $j^{th}$  block;  $\beta_0$  is the intercept;  $\beta_1$  is the impact of FF;  $FF_{ij}$  is a binary indicator for FF participation;  $\beta_2$  is the effect of baseline SEL score;  $S_{ij}$  is a baseline SEL score for the  $i^{th}$  student in the  $j^{th}$  block set to 0 when missing;  $\beta_3$  is the effect of  $S_{ij}$  being missing;  $M_{ij}$  is a dummy variable indicating if  $S_{ij}$  is missing;  $\beta_4$  is the effect of baseline attendance;  $A_{ij}$  is baseline attendance (measured prior to the start of the program from September to December);  $\beta_{5a-d}$  are the effects of baseline reading achievement scores for each assessment (a – d) used;  $P_{ij}$  is a standardized achievement score for the  $i^{th}$  student in the  $j^{th}$  block.  $\beta_{6,m}$  are the effects of student covariates;  $X_{mij}$  is the  $m^{th}$  of  $M$  additional covariates representing demographic characteristics of student  $i$  in block  $j$  (e.g. gender, free/reduced lunch, or other student-level covariates);  $\beta_{7,j}$  is the effect of block (i.e., the difference in the intercept between block  $j$  and the reference block); within each block, all FF and BAU students received the same literacy assessment (a-d); and  $\varepsilon_i$  is the error term for student  $i$ .

In our model we use robust standard errors and fixed block effects (blocks are defined by grade levels within schools). We used fixed effects rather than random effects to control for any unobserved block-specific factors. We also conducted two robustness checks of the results. For the first simple model, we stripped out all model effects except block fixed effects. For the second cluster-robust standard error term model, we fitted the same full GLM equation, except we used clustered error terms instead of fixed block effects.

### Differential Effects

Using variations of equation 1, differential effects of FF on attendance were explored by race, gender, baseline attendance (low versus high), baseline achievement (low versus high), baseline SEL (low versus high), and the intersection of these groups.

## **Attrition, Missing Data, and Sample Characteristics**

At the start of the program, 587 students were participating in the study, 294 BAU and 293 FF. During the academic year, 20 students ( $20/587 = 3.4\%$ ) exited the study. These included five students who left the program, 14 students who changed schools, and one who left for unknown reasons. Eight BAU ( $8/294 = 2.7\%$ ) and 12 FF ( $12/293 = 4.1\%$ ) students attrited, respectively. We did not replace students who dropped from the study. Out of the remaining 567 students, 27 did not have SEL scores (16 BAU and 11 FF). Following the National Center for Education Evaluation technical methods, we made dummy variable adjustments for the missing SEL values of baseline SEL scores (Puma et al., 2009). The combination of overall and differential attrition is below conservative levels of acceptability as established by the WWC (2020).

In the following Tables 3 and 4, demographic and baseline characteristics of BAU and FF groups are presented. From these baseline data, we conclude that the BAU and FF groups were equivalent. Table 3 presents the demographic characteristics of students, broken down by assignment groups. Demographic groups were evenly distributed between the BAU and FF students. Table 4 presents baseline attendance, SEL and achievement. Students assigned to FF and to BAU conditions had comparable attendance, achievement, and SEL.

# CURRENT STUDY OF FF

**Table 3: Study Participant Characteristics**

	Analytic Sample		
		BAU	FF
Grade Level	KG	105 (36.7%)	104 (37%)
	1st	100 (35.0%)	100 (35.6%)
	2nd	81 (28.3%)	77 (27.4%)
School	School 1	20 (7.0%)	18 (6.4%)
	School 2	19 (6.6%)	20 (7.1%)
	School 3	15 (5.2%)	15 (5.3%)
	School 4	15 (5.2%)	14 (5.0%)
	School 5	21 (7.3%)	18 (6.4%)
	School 6	19 (6.6%)	18 (6.4%)
	School 7	25 (8.7%)	23 (8.2%)
	School 8	19 (6.6%)	19 (6.8%)
	School 9	23 (8.0%)	22 (7.8%)
	School 10	20 (7.0%)	20 (7.1%)
	School 11	18 (6.3%)	19 (6.8%)
	School 12	20 (7.0%)	20 (7.1%)
	School 13	18 (6.3%)	20 (7.1%)
	School 14	34 (11.9%)	35 (12.5%)
Race/Ethnicity	Black	164 (57.3%)	169 (60.1%)
	White	105 (36.7%)	95 (33.8%)
	Other	17 (5.9%)	17 (6.0%)
Gender	Female	153 (53.5%)	148 (52.7%)
	Male	133 (46.5%)	133 (47.3%)
F/R Lunch	No	44 (15.4%)	45 (16%)
	Yes	242 (84.6%)	236 (84%)
Baseline Achievement	High	139 (48.6%)	142 (50.5%)
	Low	147 (51.4%)	139 (49.5%)
Baseline Attendance	High	145 (50.7%)	148 (52.7%)
	Low	141 (49.3%)	133 (47.3%)
<b>Total</b>	<b>567</b>	<b>286 (100.0%)</b>	<b>281 (100.0%)</b>
Baseline SEL*	High	128 (47.4%)	139 (51.5%)
	Low	142 (52.6%)	131 (48.5%)
<b>Total</b>	<b>540</b>	<b>270 (100.0%)</b>	<b>270 (100.0%)</b>

# CURRENT STUDY OF FF

**Table 4: Baseline Attendance, Achievement, and SEL – All Study Participants**

	Grade	Attendance Rate (unstandardized)		Attendance Days		Absence Days		Achievement (standardized)		Students	SEL (standardized)		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	Students
<b>BAU</b>	KG	93.0%	8.0%	52.7	13.5	4.2	5.0	-0.9	0.71	81	0.1	1.05	77
	1st	93.9%	6.7%	56.7	12.4	3.8	4.4	-0.7	0.61	105	-0.1	0.99	99
	2nd	94.5%	5.9%	56.6	12.5	3.4	3.7	-0.8	0.73	100	-0.2	0.93	94
	<b>Total</b>	<b>93.8%</b>	<b>6.9%</b>	<b>55.5</b>	<b>12.8</b>	<b>3.8</b>	<b>4.4</b>	<b>-0.8</b>	<b>0.69</b>	<b>286</b>	<b>-0.1</b>	<b>1.00</b>	<b>270</b>
<b>FF</b>	KG	92.7%	7.0%	53.4	13.2	4.4	4.4	-0.8	0.66	77	0.2	0.97	73
	1st	93.8%	7.9%	57.2	12.4	3.7	4.4	-0.7	0.64	104	0.0	0.98	98
	2nd	94.8%	5.9%	57.4	12.9	3.2	3.6	-0.8	0.75	100	0.1	1.04	99
	<b>Total</b>	<b>93.8%</b>	<b>7.0%</b>	<b>56.2</b>	<b>12.9</b>	<b>3.7</b>	<b>4.2</b>	<b>-0.8</b>	<b>0.68</b>	<b>281</b>	<b>0.1</b>	<b>1.00</b>	<b>270</b>
<b>Total</b>	KG	92.8%	7.5%	53.1	13.3	4.3	4.7	-0.9	0.69	158	0.1	1.00	150
	1st	93.8%	7.3%	56.9	12.4	3.8	4.4	-0.7	0.62	209	0.0	0.98	197
	2nd	94.6%	5.9%	57.0	12.7	3.3	3.6	-0.8	0.74	200	-0.1	0.99	193
	<b>Total</b>	<b>93.8%</b>	<b>6.9%</b>	<b>55.9</b>	<b>12.8</b>	<b>3.8</b>	<b>4.3</b>	<b>-0.8</b>	<b>0.69</b>	<b>567</b>	<b>0.0</b>	<b>1.00</b>	<b>540</b>



## HOW WELL WAS FF SCALED UP?

FF students were scheduled to receive three tutoring sessions per week (30 minutes per session). For minimum fidelity, students needed to receive two per week, or 60 minutes of tutoring. For a school to meet minimum fidelity, 80% of students would need to meet this goal. If more than 75% of the schools reached the goal, the overall program meets fidelity requirements.

The average FF student across the schools received 2.6 tutoring sessions per week. Of the 281 FF participants, 252 (89.7%) received the full tutoring program (at least two sessions a week) (Table 5; Figure 1). Given that 78.6% of the schools (11 out of 14 sites) met the implementation benchmark of 80% of students receiving two or more sessions per week, EA successfully scaled up the tutoring component of FF (Figure 2).



**Table 5: Tutoring sessions per week (through March)**

	Mean	SD	Min	Max	Students Receiving 2+ Sessions Per Week		Total Students	Met Adequate Implementation (80% benchmark)
School 1	2.3	0.5	1.4	3.1	14	77.8%	18	No
School 2	2.3	0.6	1.2	3.3	16	80.0%	20	Yes
School 3	2.5	0.4	1.5	2.9	14	93.3%	15	Yes
School 4	2.6	0.4	2.0	3.2	14	100.0%	14	Yes
School 5	2.5	0.2	2.0	2.8	17	94.4%	18	Yes
School 6	2.8	0.6	1.8	3.9	15	83.3%	18	Yes
School 7	2.7	0.2	2.0	3.0	23	100.0%	23	Yes
School 8	2.9	0.2	2.7	3.2	19	100.0%	19	Yes
School 9	2.7	0.4	1.1	3.2	22	100.0%	22	Yes
School 10	2.7	0.2	2.1	3.0	20	100.0%	20	Yes
School 11	2.2	0.5	1.5	3.1	12	63.2%	19	No
School 12	2.4	0.5	1.6	3.2	15	75.0%	20	No
School 13	2.6	0.3	2.0	3.1	19	95.0%	20	Yes
School 14	2.8	0.6	1.7	4.6	32	91.4%	35	Yes
<b>Overall</b>	<b>2.6</b>	<b>0.5</b>	<b>1.1</b>	<b>4.6</b>	<b>252</b>	<b>89.7%</b>	<b>281</b>	<b>Yes</b>

# HOW WELL WAS FF SCALED UP?

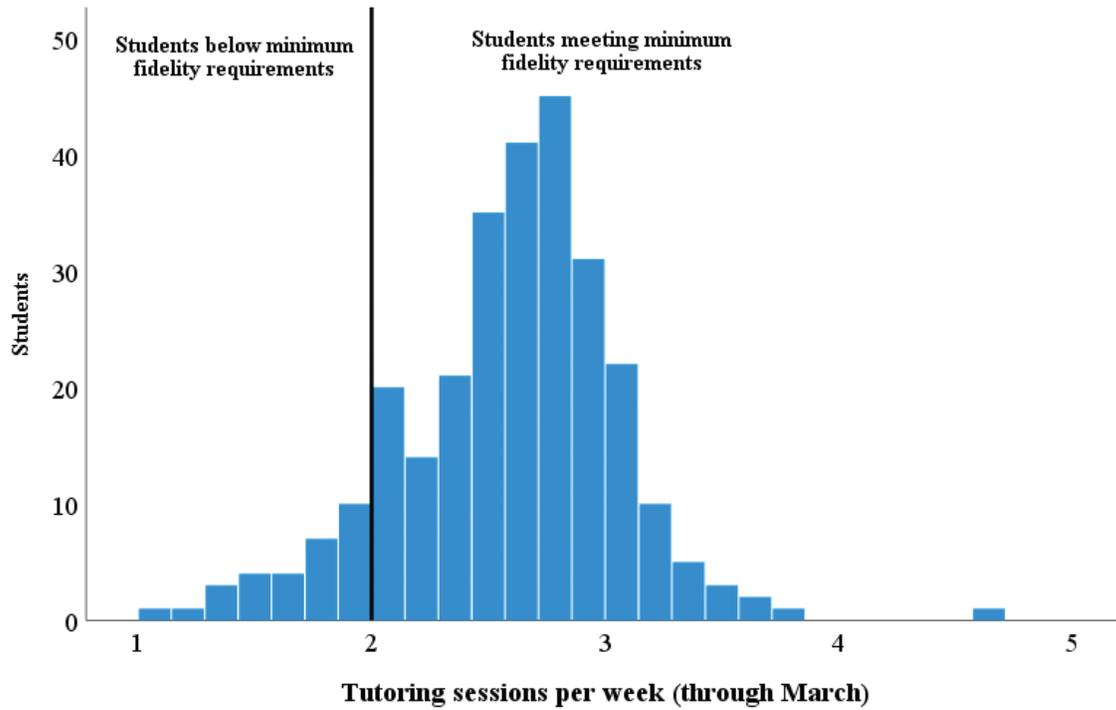


Figure 1: Distribution of the tutoring sessions per week for each FF participant

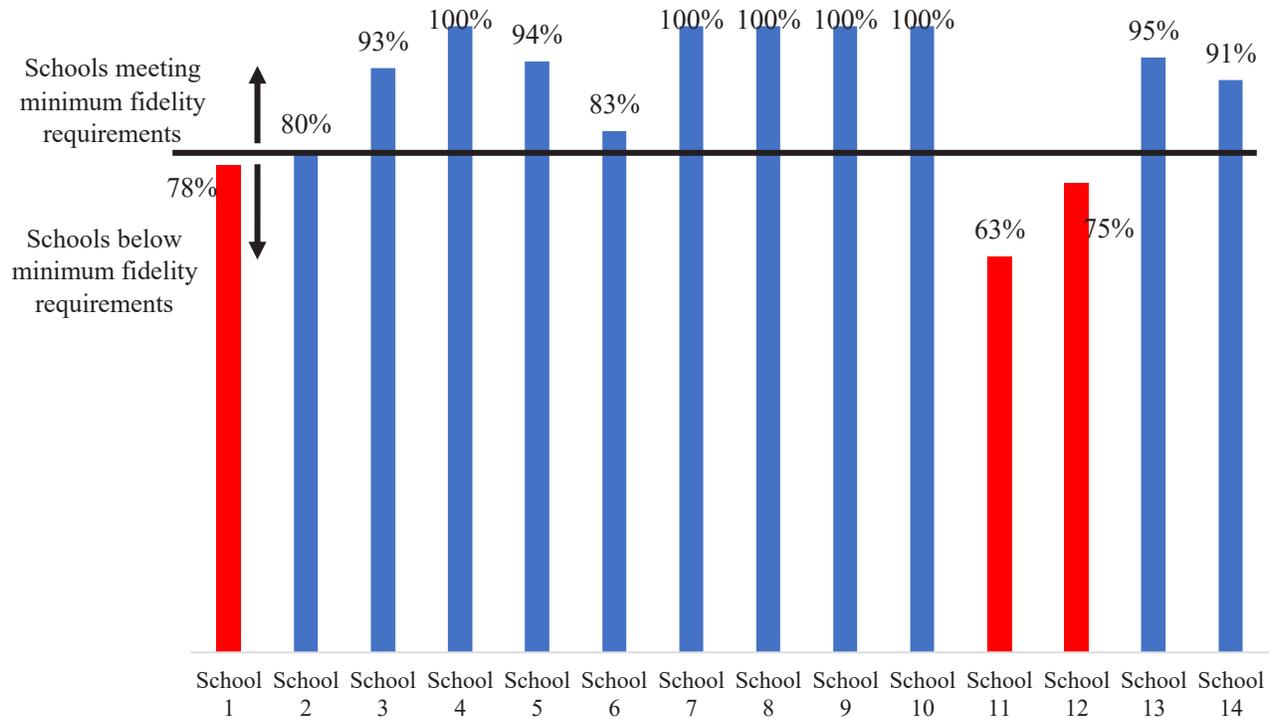


Figure 2: Percent of participants meeting tutoring fidelity (2+ tutoring sessions per week): By school

## HOW WELL WAS FF SCALED UP?

The FE component of FF involves substantive interactions or communications with participant family members. Successful contacts were defined as an exchange between FF staff and a student's family member. Two or more family contacts each month was required for a student to receive FE as intended. For a school to implement FE as intended, 80% of families/students need to meet this goal. Overall program fidelity requires more than 75% of the schools reaching the 80% objective.

The results suggest FF did not fully scale up the FE component. The typical student's family was contacted 1.6 times each month (Table 6). Only 30% of students' families were successfully contacted by FF staff at least two times per month (Table 6; Figure 3). None of the 14 sites met the 80% threshold requirement for full implementation (Figure 4).

**Table 6: FE contacts per month (through March)**

	Mean	SD	Min	Max	Students' Families Contacted 2+ Times Each Month		Total Students	Met Adequate Implementation (80% benchmark)
School 1	1.2	0.6	0.3	2.3	3	16.7%	18	No
School 2	1.4	1.0	0.0	3.8	6	30.0%	20	No
School 3	1.0	0.6	0.0	1.9	0	0.0%	15	No
School 4	2.5	1.9	0.6	7.5	5	35.7%	14	No
School 5	1.6	1.4	0.0	5.6	6	33.3%	18	No
School 6	0.7	0.5	0.0	1.8	0	0.0%	18	No
School 7	1.2	0.8	0.0	3.0	2	8.7%	23	No
School 8	1.9	1.1	0.0	3.8	8	42.1%	19	No
School 9	2.0	0.9	0.6	4.1	11	50.0%	22	No
School 10	1.9	0.7	0.7	3.2	10	50.0%	20	No
School 11	2.3	1.4	0.6	4.7	8	42.1%	19	No
School 12	1.5	1.0	0.6	3.4	6	30.0%	20	No
School 13	1.0	0.4	0.0	1.3	0	0.0%	20	No
School 14	2.1	0.4	1.2	3.0	20	57.1%	35	No
<b>Overall</b>	<b>1.6</b>	<b>1.1</b>	<b>0.0</b>	<b>7.5</b>	<b>85</b>	<b>30.2%</b>	<b>281</b>	<b>No</b>



# HOW WELL WAS FF SCALED UP?

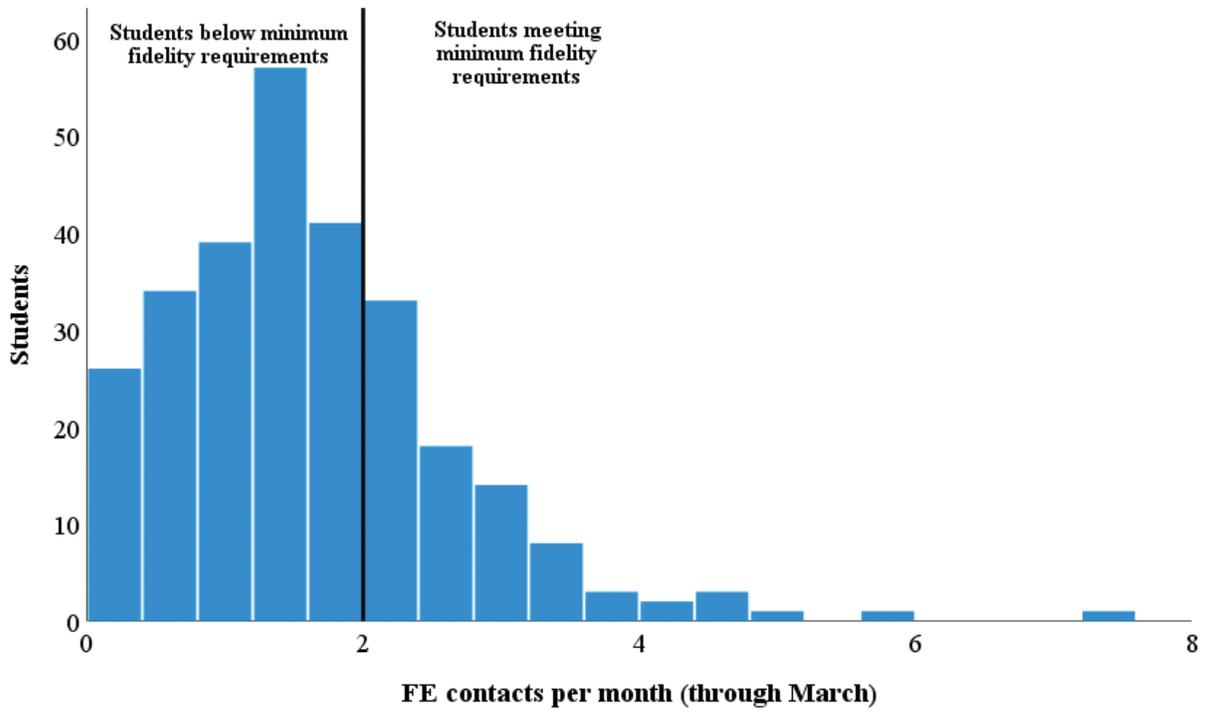


Figure 3: Distribution of the FE contacts per month for each FF participant

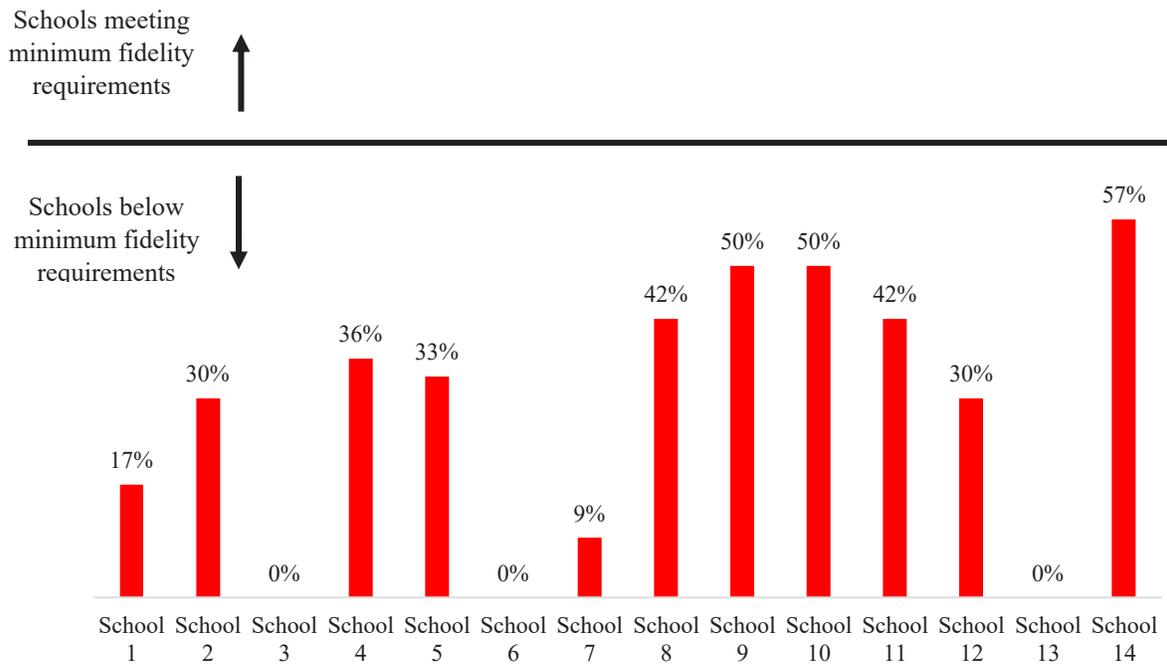


Figure 4: Percentage of participants meeting FE fidelity (2+ FE contacts per month): By school

# WHAT WAS THE IMPACT OF FF?

## Descriptive Results

As previously stated, students were not tested in the spring of 2020 due to the pandemic, making the study unable to assess the impact of FF on reading achievement and SEL. However, measuring its impact on school attendance was still possible. Table 7 presents the unadjusted follow-up attendance results for students retained in the study: Attendance was greater for FF students during the tutoring period by 1.4 percentage points ( $M = 91.2\%$  vs.  $M = 89.8\%$ ). On average, FF participants missed one fewer day of school.

**Table 7: Follow-up (during FF program) attendance – All study participants**

	Grade	Attendance Rate (unstandardized)		Attendance Days		Absence Days		Students
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<b>BAU</b>	KG	88.2%	13.4%	60.8	15.5	7.9	8.6	81
	1st	89.8%	12.3%	59.6	13.5	6.7	7.8	105
	2nd	90.1%	8.1%	60.7	12.8	5.8	5.0	100
	<b>Total</b>	<b>89.8%</b>	<b>11.4%</b>	<b>60.3</b>	<b>13.8</b>	<b>6.7</b>	<b>7.2</b>	<b>286</b>
<b>FF</b>	KG	89.9%	8.7%	61.5	13.9	6.8	6.0	77
	1st	91.2%	8.6%	60.2	11.9	5.7	5.5	104
	2nd	92.3%	7.8%	61.5	12.8	5.0	4.8	100
	<b>Total</b>	<b>91.2%</b>	<b>8.4%</b>	<b>61.0</b>	<b>12.8</b>	<b>5.7</b>	<b>5.4</b>	<b>281</b>
<b>Total</b>	KG	89.0%	7.5%	61.2	14.7	7.3	7.4	158
	1st	90.5%	10.6%	59.9	12.7	6.2	6.8	209
	2nd	91.6%	8.0%	61.1	12.8	5.4	4.9	200
	<b>Total</b>	<b>90.5%</b>	<b>10.0%</b>	<b>60.7</b>	<b>13.3</b>	<b>6.2</b>	<b>6.4</b>	<b>567</b>



# WHAT WAS THE IMPACT OF FF?

## Overall Impact

FF participation had a statistically significant, positive impact on attendance. After controlling for block fixed effects and covariates (Full GLM (1)), FF had a 0.13 standardized impact on student attendance ( $\beta = 0.13$ ,  $SE = 0.06$ ,  $p = 0.031$ ; Table 8). This standardized impact corresponds to a 1.3 percentage points higher attendance rate for FF than BAU students. This statistically significant impact was robust with a simple model that only controlled for fixed block effects ( $\beta = 0.14$ ,  $SE = 0.07$ ,  $p = 0.047$ ) and in a model that clustered error terms within blocks instead of fixed block effects ( $\beta = 0.14$ ,  $SE = 0.06$ ,  $p = 0.029$ ).

## Differential Impact

The impact of FF was larger with Black ( $\beta = 0.18$ ,  $SE = 0.09$ ,  $p = 0.047$ ) and male ( $\beta = 0.27$ ,  $SE = 0.10$ ,  $p = 0.009$ ) students (Table 8). The model-adjusted Black FF participants' attendance rate was 1.9 percentage points higher than BAU students. For male students, participating in FF resulted in a model-adjusted 2.8 percentage points better attendance rate. Across gender and race, students with low school attendance prior to the start of the program also benefited more from their participation ( $\beta = 0.25$ ,  $SE = 0.11$ ,  $p = 0.025$ ).

Examining the intersectionality of the impact of FF on students suggests Black male students were especially affected by their participation in FF (Table 8); their school attendance rate was a model-adjusted 4.3 percentage points higher ( $\beta = 0.42$ ,  $SE = 0.17$ ,  $p = .015$ ). Although not statistically significant, the impact of FF on Black male students with low school attendance at the start of the FF program was very large ( $\beta = 0.59$ ,  $SE = 0.33$ ,  $p = .074$ ). To put these impact estimates into context, the model-adjusted attendance rate for Black male students with low attendance prior to the start of the program was 86.2%. This compares to a model-adjusted attendance rate of 80.1% for Black male students with low attendance who received BAU reading instruction. This 6.0% attendance rate difference indicates FF participants in this group were absent from school 4.2 fewer days. Together, these results suggest FF had its greatest impact on students with the greatest risk for having difficulties in school (Figure 5; Figure 6).

# WHAT WAS THE IMPACT OF FF?

**Table 8: Impact of Future Forward school attendance**

		Standardized Impact ( $\beta$ )	SE	p	Unstandardized Impact (percentage points)
Overall	<b>Full GLM</b>	<b>0.13</b>	<b>0.06</b>	<b>0.031</b>	<b>1.3%</b>
	Robustness Model 1 – simple model	0.14	0.07	0.047	1.4%
	Robustness Model 2 – clustered standard error terms	0.14	0.06	0.029	1.4%
Subgroups	<b>Black students</b>	<b>0.18</b>	<b>0.09</b>	<b>0.047</b>	<b>1.9%</b>
	White students	0.01	0.08	0.879	0.1%
	Females	0.06	0.08	0.469	0.6%
	<b>Males</b>	<b>0.27</b>	<b>0.10</b>	<b>0.009</b>	<b>2.8%</b>
	<b>Students w/low attendance</b>	<b>0.25</b>	<b>0.11</b>	<b>0.025</b>	<b>2.5%</b>
	Students w/high attendance	0.04	0.05	0.470	0.4%
	Students w/low achievement	0.16	0.10	0.110	1.6%
	Students w/high achievement	0.11	0.07	0.134	1.1%
	Students w/low SEL	0.18	0.11	0.100	1.8%
	Students w/high SEL	0.10	0.08	0.213	1.0%
Intersectionality	Black students w/low attendance	0.28	0.16	0.080	2.9%
	Male students w/low attendance	0.36	0.19	0.058	3.7%
	Black students w/low achievement	0.16	0.15	0.276	1.6%
	Male students w/low achievement	0.37	0.17	0.026	3.8%
	Black students w/low SEL	0.26	0.16	0.099	2.6%
	Male students w/low SEL	0.18	0.14	0.197	1.9%
	<b>Black male students</b>	<b>0.42</b>	<b>0.17</b>	<b>0.015</b>	<b>4.3%</b>
	<b>Black male students w/low attendance</b>	<b>0.59</b>	<b>0.33</b>	<b>0.074</b>	<b>6.0%</b>
	Black male students w/low achievement	0.41	0.26	0.119	4.2%
	Black male students w/low SEL	0.27	0.24	0.269	2.8%

# WHAT WAS THE IMPACT OF FF?

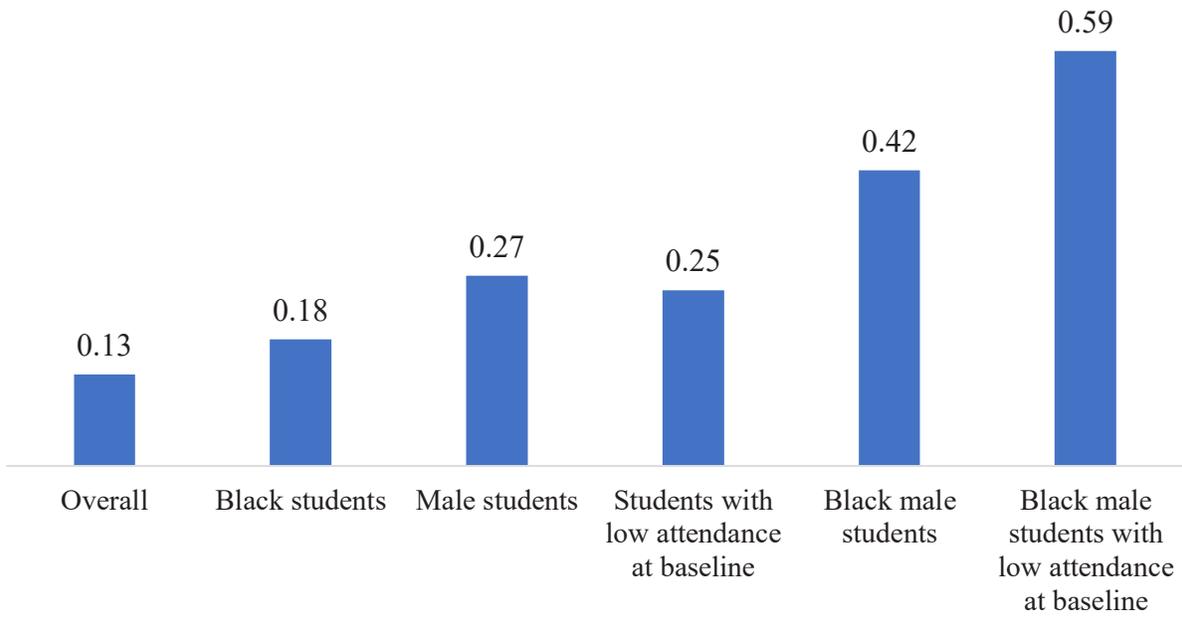


Figure 5: Standardized impact of FF on the attendance of student subgroups

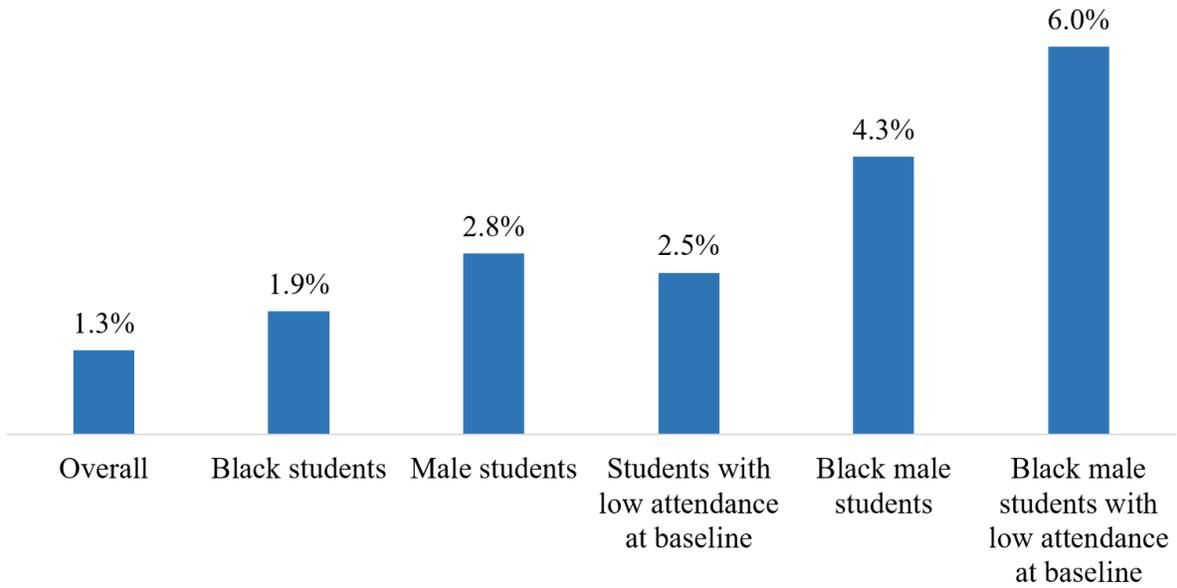


Figure 6: Percentage point change impact of FF on the attendance of student subgroups

## CONCLUSION AND DISCUSSION

The current EIR-funded RCT study from the 2019-20 school year adds to the growing body of evidence of the scalability and effectiveness of the FF program. Regarding scalability, most students (89.7%) across the 14 schools received one-on-one tutoring as intended. Most of these schools were new to FF in the 2018-19 school year. The evaluation from that year reported far fewer students (34.7%) received the amount of tutoring FF was designed to provide (Jones, et al., 2020). Further, 11 out of the 14 sites in the 2019-20 school year met the site-level threshold for tutoring fidelity (at least 80% of participants with two or more sessions per week) as opposed to only three sites in the 2018-19 school year. This suggests EA was able to scale up FF quickly, working with local Boys & Girls Clubs, to provide the tutoring component of the program effectively.

Regarding family engagement (FE), the current study found only a minority of students' families (30.2%) received FE as intended. Sites struggled to connect with participant families at the level that was expected. Even considering this, the current study's FE implementation was much improved from the implementation during the 2018-19 school year, when only 6.6% of students received the intended amount of FE (Jones et al., 2020). As FF continues in these schools in subsequent years, it will be important to continue to monitor FE implementation. As the EIR-funded evaluation of FF continues, it also may be necessary to adjust the program's expectations for the frequency of family contacts or for how long it takes a site to fully implement this critical component of FF.



## CONCLUSION AND DISCUSSION

Because in-person schooling stopped in the wake of the COVID-19 pandemic, our analyses were limited to school attendance. Regarding attendance, considering the challenges faced by sites to implement the FE component of FF, it is somewhat surprising that FF was still found to have a statistically significant, positive impact. This is the second study of FF to report a positive effect on attendance (Jones & Christian, 2020), which further validates the FF model design pairing tutoring with parent engagement to support the literacy development of students and families.

The current study went further than previous studies of FF to examine the differential effects of FF, i.e., who benefited the most from their participation. FF was found to have a greater impact on Black students, male students, and students with low school attendance prior to the start of the program. It was the impact on these student groups that explained FF's overall program impact. FF had its greatest impact on students at the intersection of these groups; Black male students starting the year with low attendance. It seems that FF had its greatest impact on students at greatest risk of facing difficulties in school.

The EIR-funded evaluation of FF will continue during the 2020-21 school year. However, given the continued impact of the COVID-19 pandemic on schools, FF has had to adapt. EA has designed a distance-learning version of FF so that students and families can continue to receive literacy supports during the pandemic. Because of the challenges schools and EA face trying to continue education during the COVID-19 crisis, the evaluation will only measure the impact of the modified FF program on student achievement. Many of the study schools have continued with distance learning during the 2020-21 school year, so it is unclear that an analysis of school attendance would be meaningful. The evaluation of the modified FF program represents an opportunity to test an alternate method for delivering FF tutoring and family engagement. If this version of FF is proven effective, it would lead to more flexibility for schools and families interested in participating in FF.

# REFERENCES

- Borman, G., & Hewes, G. (2003). Long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis, 24*(2), 243–266.
- Borman, G., Slavin, R., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Education Research Journal, 44*(3), 701–731.
- Bryan, J. (2005). Fostering educational resilience and achievement in urban schools through school-family-community partnerships. *Professional School Counseling, 8*(3), 219.
- Bryan, J., & Henry, L. (2012). A model for building school-family-community partnerships: Principles and process. *Journal of Counseling & development, 90*(4), 408-420.
- DIBELS. (2021.). *What are DIBELS?* Retrieved online at: <https://dibels.uoregon.edu/assessment/dibels#:~:text=The%20Dynamic%20Indicators%20of%20>
- TJCC and FastBridge Learning LLC. (2015). *Formative Assessment System for Teachers: Abbreviated Technical Manual for Iowa Version 2.0*. Minneapolis, MN Retrieved online at: [https://fastbridge.org/wp-content/uploads/2017/02/fastbridge\\_learning\\_abbreviated\\_iowa\\_technical\\_manual\\_version\\_2\\_0\\_final.pdf](https://fastbridge.org/wp-content/uploads/2017/02/fastbridge_learning_abbreviated_iowa_technical_manual_version_2_0_final.pdf)
- Hartwig, E. P., & Hayes, W. M. (2019a). *b.e.s.t. 3: Identifying Externalizing and Internalizing Behaviors*.
- Hartwig, E. P., & Hayes, W. M. (2019b). *b.e.s.t. 5: Validation and standardization procedures*.
- Invernizzi, M., Swank, L., Juel, C., & Meier, J. (2003). *Phonological Awareness Literacy Screening-Kindergarten*. Charlottesville, VA: University Printing.
- Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36*(4), 479-500.
- Invernizzi, M., Juel, C., Swank, L., & Meier, J. (2015). *Phonological Awareness Literacy Screening*. Charlottesville, VA: University of Virginia.
- Jones, C. J. (2018). SPARK Early literacy: Testing the impact of a family-school-community partnership literacy intervention. *School Community Journal, 28*, 247–264.
- Jones, C. J., & Christian, M. (2020). The results of a randomized control trial evaluation of the SPARK literacy program: An Innovative approach that pairs one-on-one tutoring with family engagement. *Journal of Education for Students Placed at Risk (JESPAR)*. Advance online publication. <https://doi.org/10.1080/10824669.2020.1809419>.
- Jones, C. J., Bower, J., Pyatigorsky, M., & Price, C. (2020, September). *The Results from a Multi-site Regression Discontinuity Impact Study of the Future Forward Literacy Program*. Retrieved online at: <https://uwm.edu/sreed/wp-content/uploads/sites/502/2020/10/Regression-Discontinuity-Study-of-Future-Forward.pdf>
- Kilpatrick, K. D., Maras, M. A., Brann, K. L., & Kilgus, S. P. (2018). Universal screening for social, emotional, and behavioral risk in students: DESSA-mini risk stability over time and its implications for screening procedures. *School Psychology Review, 47*(3), 244-257.
- McGee, L. M. (2006). Research on Reading Recovery: What is the impact on early literacy research? *Literacy,*

# REFERENCES

*Teaching, and Learning*, 10(2), 1–50.

- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011). Universal screening for social emotional competencies: A study of the reliability and validity of the DESSA-mini. *Psychology in the Schools*, 48(7), 660-671.
- Northwest Evaluation Association (2009). *Technical Manual for Measures of Academic Progress and Measures of Academic Progress for Primary Grades*. Lake Oswego, OR: Author.
- Pearson. (2014). *Assessment Instrument table: DRA2*. Retrieved online at: [https://www.cde.state.co.us/uip/assessment\\_instrument\\_description\\_dra2\\_3\\_20\\_14](https://www.cde.state.co.us/uip/assessment_instrument_description_dra2_3_20_14)
- Puma, M. J., Olsen, R. B., Bell, S. H., & Price. C. (2009). *What to do When Data are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved online at: <https://files.eric.ed.gov/fulltext/ED511781.pdf>
- Shapiro, V. B., Kim, B. K. E., Robitaille, J. L., & LeBuffe, P. A. (2016, October 13). Protective factor screening for prevention practice: Sensitivity and specificity of the DESSA-Mini. *School Psychology Quarterly*, 32(4), 449-464.
- University of Oregon. (2018). *Understanding the Research Behind DIBELS® 8th Edition* (Technical Report 1801). Eugene, OR: Author.
- University of Oregon. (2018-2020). *8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Technical Manual*. Eugene, OR: University of Oregon. Retrieved online at: <https://dibels.uoregon.edu>
- University of Oregon (2020). *Understanding DIBELS® 8th Edition Composite and Measure Scores*. Retrieved online at: <https://dibels.uoregon.edu/docs/UnderstandingDIBELS8CompositeScores.pdf>
- What Works Clearinghouse. (2020). *Standards Handbook, Version 4.1*. Retrieved online at: <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf>



This study was supported through a federal grant from the U.S. Department of Education. We appreciate the support and help of Kate Bauer-Jones, Kris Bischoff, Tia Hatchett, Anthony Milanowski, and Cristopher Price.

Curtis Jones is the Director of the Office of Socially Responsible Evaluation in Education at UWM.

DongMei Li is a Senior Research/Evaluation Associate in the Office of Socially Responsible Evaluation in Education at UWM.

**For more information about this report or about the evaluation of the Future Forward Literacy Program, please contact Curtis Jones at [jones554@uwm.edu](mailto:jones554@uwm.edu) or visit [uwm.edu/sreed](http://uwm.edu/sreed).**