Future Forward is an early elementary school literacy program that pairs one-on-one tutoring with family engagement.

The Results from a Multi-Site Regression Discontinuity Impact Study of the Future Forward Literacy Program

September 2020

Curtis J. Jones and John Bowser, University of Wisconsin Milwaukee
Mikhail Pyatigorsky, Pyatigorsky Consulting
Cristofer Price, Abt Associates

Future Forward (FF) is an early elementary literacy program that pairs one-on-one tutoring, provided by volunteers, with parent engagement. Previously, FF, under the name of SPARK, was awarded an Investing in Innovations (i3) grant to test its impact in seven Milwaukee schools. Two randomized studies of FF/SPARK conducted as part of the i3 grant found FF/SPARK to have positive impacts on literacy, reading achievement, and school attendance. In 2017, Education Analytics received an Education Innovation and Research (EIR) Mid-Phase grant to establish FF in 14 additional schools across three states. During the 2018-19 school year, 12 schools participated in a multi-site regression discontinuity (RD) study of its impact on the literacy development of 222 students. While we did not find a statistically significant impact of FF, the magnitude of the impact estimate was similar to what was reported in the i3 studies. Limitations inherent to RD studies, and implementation challenges, hindered our ability to reliably measure the impact of FF.

The Results from a Multi-Site Regression Discontinuity Impact Study of the Future Forward Literacy Program

Future Forward (FF) works to improve the literacy skills of early elementary children through a combination of one-one-one tutoring and family engagement. Administered by Education Analytics, in partnership with local community organizations and schools, FF leverages a school-community-family partnership strategy (Bryan & Henry, 2012) to address the literacy needs of students and families. In 2017, EA received an Education Innovation and Research (EIR) Mid-Phase grant to establish FF in 14 schools during the 2018-19 school year, and then to test its impact in 2019-20 and 2020-21. Although the 14 schools were learning to effectively implement FF during the 2018-19 school year, the external evaluators at the Office of Socially Responsible Evaluation in Education (SREED) in the University of Wisconsin in Milwaukee (UWM) conducted a quasi-experimental, regression discontinuity (RD) study of its impact. Through this RD study, we compare the literacy development of students who were close to eligible to receive FF, but did not receive it, to that of students who were barely eligible to receive FF, and did. Well-executed RD studies provide rigorous estimates of the impact and causal effects of interventions, and similar to randomized control trials, meet the Institute for Education Sciences What Works Clearinghouse (WWC) evidence standards *without reservation*.

The importance that early elementary students learn to read.

The successful development of literacy in elementary school is a strong predictor of future academic success (Rabiner, Godwin & Dodge, 2016). Further, dropping out of school is predicted with 70% accuracy by the 3rd grade based on reading ability and prior retention (Hernandez, 2012). Outside of the classroom, literacy also predicts long-term economic and health outcomes (Berkman, Sheridan, Donahue, Halpern & Crotty, 2011; DeWalt, Berkman, Sheridan, Lohr & Pignone, 2004). The racial/ethnic and economic disparities in reading achievement among children are reflected in disparities among health outcomes among adults (Sudano & Baker, 2006) and children (Mehta, Lee & Ylitalo, 2013). Thus, the importance of developing emergent literacy skills at a young age, for the short and long-term benefits it is associated with, cannot be overstated.

Considering the well-understood importance of reading and literacy skill development, it is discouraging that, nationally, only 36% of students at both the fourth and eighth grade levels are

proficient in reading (U.S. Department of Education, 2017). While overall literacy rates are already low, they are even lower for students who are low-income (20%). Considering the low reading proficiency rates and the impact that early literacy has on students' lives, it is critical that new effective reading interventions are developed that can reach the large numbers of early elementary students behind in their literacy development. The current study tests the impact of one such program, FF, which could help reduce the societal gap between the literacy development needs of students and the supports available to them.

The challenge of supporting the literacy development of all students who need support.

The Evidence for Every Student Succeeds Act (ESSA) website (https://www.evidenceforessa.org/) collects reading and math programs that meet the various levels of evidence of effectiveness as defined by ESSA (2015). Programs with at least one well-designed randomized control study showing a positive impact are categorized as having Strong Evidence of effectiveness. Only 11 reading programs that focus on struggling readers meet the criteria for having strong evidence of effectiveness.

Of the 11 reading programs, Reading Recovery may be the most widely known and studied. Reading Recovery uses trained, certified teachers to provide one-on-one tutoring to students in need of reading support. This approach has proven effective across a number of studies (D'Agostino, & Murphy, 2004; D'Agostino, & Harmey, 2016), with a recent large randomized study of 6,888 participants across 1,222 schools in the United States finding it had an impact of .48 standardized units on the Iowa Test of Basic Skills reading assessment (Sirinides, Gray, & May, 2018). Reading Recovery has also proven effective in international schools (D'Agostino & Harmey, 2016). While Reading Recovery might be an effective option for some districts, the staffing resources needed to use certified teachers to tutor all students who need additional support make it difficult to implement in districts already facing staffing shortages with large numbers of students behind in their literacy development.

In districts consistently facing teacher shortages with large numbers of students who need literacy support, one-on-one tutoring provided by volunteers or paraprofessionals may be a more viable option for helping bridge the gap between the literacy needs of students and the resources available to support them. While this approach may not be as effective as one-on-one tutoring provided by a certified teacher, it has proven to be effective (Inns, Lake, Pellegrini, & Slavin,

2018). In fact, of the 11 programs with strong evidence of effectiveness included on the Evidence for ESSA website, seven use paraprofessionals or volunteers as tutors. A meta-analysis of 21 tutoring program studies with randomized evaluation designs found that students tutored by volunteers demonstrate greater oral fluency and writing scores compared to control students (Ritter, Barnett, Denny, & Albin, 2009). One recent evaluation found that, even with "minimally trained" college students from non-education majors serving as tutors, tutored students displayed significantly more growth than non-tutored students in key literacy outcomes (Lindo, Weiser, Cheatham, & Allor, 2017). Another evaluation suggests that positive reading outcomes can be achieved by culturally diverse students who participate in a structured volunteer tutoring program (Moore-Hart & Karabenick, 2009). In her review of effective volunteer tutoring programs, Wasik (1998) found that the most successful volunteer tutoring programs are highly structured, have quality materials, and provide strong professional development and supervision to tutors. Effective programs provide an intensive student experience of at least 90 minutes per week, are well coordinated with classroom instruction, and use ongoing, regular assessments to track student progress. While it may not be possible for many districts to provide students one-on-one tutoring from a certified teacher, a well-designed program that effectively uses volunteers or paraprofessionals can still have a significant impact on students.

Leveraging families to improve student literacy.

In addition to skill-based strategies, parent engagement also could be leveraged to address student difficulties developing foundational literacy skills. Traditional efforts to encourage family involvement in educational programs focus on families attending events, receiving information from staff, volunteering, and generally exhibiting "good parent" behaviors (Epstein, 2001; Li, 2010). However, fully engaging parents in the education of students involves a much deeper level of involvement that ultimately yields a greater impact on student success. A meta-analysis of 51 studies found that a number of parent activities are associated with student academic outcomes, such as parents and their children reading together, checking homework, and parent/teacher communication and partnerships (Jeynes, 2012). Engaging families in tutoring programs can improve children's academic knowledge, skills, and confidence (Bryan, 2005; Little, 2009). Getting to know families and the ways that their lives are structured outside of the educational setting may lead to a reciprocal relationship that can increase their involvement (Graue & Hawkins, 2010). While outcomes for all students improve with additional family

engagement, the demonstrated positive working relationship between the home and school is shown to have an added literacy benefit for low-income children with less-educated parents (Dearing, Kreider, Simpkins, & Weiss, 2006; Lin, 2003). Increased family engagement leads to increased positive feelings about literacy, which in turn improves literacy performance (Dearing et. al, 2006). Literacy benefits result when parents believe reading is enjoyable and children observe them enjoying reading to them (Sonnenschein, Baker, Serpell, & Schmidt, 2017).

Family involvement is also closely connected to student attendance. Chronic absenteeism (missing more than 10% of school days) is a problem that disproportionally impacts poor students and students of color and predicts both academic and social problems later in students' educations (Chang & Romero, 2008). A U.S. Department of Education (2016) report on chronic absenteeism revealed that one in seven students missed 15 or more days of school in 2013-14. Research has shown that school, family, and community partnership strategies can reduce chronic absenteeism (Sheldon & Epstein, 2004). Ultimately, for literacy instruction to work and for student literacy to improve, children first need to be in school to receive instruction. Thus, family engagement has the potential to improve student literacy both through changes in parent attitudes about literacy and through increased student school attendance.

The Future Forward Early Literacy Program

A weakness common to many tutoring programs is that learning gains made during the early elementary years have been shown to diminish by late elementary grades (D'Agostino, Lose, & Kelly, 2017; Hurry & Sylva, 2007). A reason for this diminished effect is that when students complete a program, they remain influenced, directly and indirectly, by the same factors (school, family and community) that contributed to their deficit. A way to promote a more stable effect is to work to develop school, family, and community systems around a student that can support their literacy development beyond program participation. This was the motivation behind Future Forward combining tutoring with family engagement.

Future Forward (FF) is an early literacy program, administered by Education Analytics for early primary grade students. EA partners with local community agencies to provide struggling readers intensive one-on-one tutoring and family engagement. FF leverages a School-Community-Family Partnership approach (Bryan, & Henry, 2012) to building literacy skills in students. The partnership structure can assist the student and family by addressing academic and even

logistical barriers towards fully addressing educational needs that exist prior to, and during, FF participation. Furthermore, the inclusion of a community partner into the school and school-family relationship has the potential to add an extra level of caring and empathy into the child's educational experience, in particular when families are working with larger school districts (Toffler & Toffler, 1995).

Within FF program sites, each building has an assigned site manager, who is typically a certified teacher, and a family engagement coordinator (FEC). The site manager manages the tutoring activities, which includes the hiring of, generally paraprofessional, tutors to conduct one on one tutoring and coordinates with the school and teachers to schedule sessions around core instruction classes. The FEC is responsible for family outreach and communication, which involves monthly family events and ongoing contacts to help facilitate literacy development outside of school.

The individual tutoring sessions vary based on the skills and interests of the students. Areas of focus during the sessions can include a focus on letter sounds, phonological awareness, shared readings, tutor read aloud, leveled/instructional readings, and writing activities. FF sites are expected to provide students with a minimum of two 30-minute tutoring sessions per week and to communicate with families at least two times per month. Family contacts should be mutual (i.e. in-person meeting, phone call conversation, email conversation). At least 80% of students within a program site should meet tutoring and family contact minimums for FF to be considered fully implemented. A more full description of the Future Forward program has been published elsewhere (Jones & Christian, 2020).

What does previous research tell us about the Future Forward program?

Under the name SPARK, FF was previously awarded an Investing in Innovations (i3) grant to test its impact within the Milwaukee Public Schools (MPS). Two separate impact studies were conducted as part of this grant. Each of these studies has been reviewed and included in the WWC as studies that meet their design standards "without reservations".

The first, over the course of the 2011-12 and 2012-13 school years, represented a pilot evaluation conducted as SPARK/FF was still being developed. This randomized control trial study of 251 participants and 245 control students took place in six MPS buildings and found a small, but

positive effect (Hedges's $g = 0.12$) of two years of SPARK/FF programming on the Measures of Academic Progress (MAP) reading achievement test (NWEA, 2009) (Jones, 2018).

A second randomized control study of 286 SPARK/FF participants and 290 control students took place over the course of the 2013-14 and 2014-15 school years in seven MPS schools. The results of this study found positive and statistically significant impacts on literacy development, reading achievement, and school attendance (Jones & Christian, 2020). Specifically, two years of SPARK/FF had a statistically significant impact of 0.23 standard deviations on foundational literacy ($\beta = 0.23$, $p = 0.001$). Although the standardized effect of two years of SPARK/FF on reading achievement was not statistically significant ($\beta = 0.10$, $p = 0.125$), it was found to have a statistically significant positive impact after just one year ($\beta = 0.11$, $p = 0.048$). The impact of SPARK/FF on foundational literacy was found to be strongest among struggling readers ($\beta = 0.52$, $p < 0.001$). Finally, and validating of the family engagement focus of the program, SPARK/FF was found to have a positive, statistically significant impact on school attendance. After one year, SPARK/FF students were absent 3.4 fewer days than control students ($\beta = -3.36$, $p = 0.001$; Table 7), and after two years, 4.5 fewer days ($\beta = -4.53$, $p = 0.043$).

Current Research/Evaluation Study

In 2017 FF was awarded a Mid-Phase/Tier 2 Education Innovation and Research (EIR) Grant from the U.S. Department of Education to further test its impact on students in 14 schools across seven school districts in three states. Based on the finding that most of the benefit of SPARK/FF was realized after one year of programming, program participation duration was changed from two years to one. In addition, based on the finding that students with the lowest baseline literacy scores realized the most benefit from their participation, the current study focuses on struggling readers. The independent evaluation, again led by SREED, uses a regression discontinuity (RD) design to measure impact during the 2018-19 program year and a randomized control design to measure impact during the 2019-20 and 2020-21 program years. The current report presents the results of the 2018-19 RD study.

In the current study, our primary research question is:

> *What is the impact of one year of Future Forward participation on the literacy development of lower primary students compared to students receiving business as usual literacy instruction?*

What is a regression discontinuity study?

H In a regression discontinuity design (Thistlethwaite & Campbell, 1960), the impact of a program is determined by comparing the outcomes of treatment and comparison groups with similar baseline abilities. Assignment to the treatment or comparison group is determined by whether students are above or below an eligibility cut point on a continuous "forcing variable" or "assignment variable" (Bajari, Hong, Park, & Town, 2011). The difference between the outcomes of students on either side of the eligibility cut point represents the program's impact. As will be described in more detail below, the analysis does not include all students in the study. Rather it includes only those who are near the eligibility cut point. In the current study of FF, the forcing variable was derived from baseline literacy assessment results.

Which literacy assessments were used to determine eligibility and impact?

The literacy assessments, which varied across the 12 schools, included: (1) Phonological Awareness Literacy Screening *PALS* (Invernizzi, Swank, Juel & Meier, 2003) (2) MAP Reading Fluency (NWEA, 2009) and (3) AIMSweb Plus (Pearson Education, 2008)

The *Phonological Awareness Literacy Screening (PALS),* used by eight of 12 schools, is a criterion-referenced, teacher administered, assessment of foundational literacy (Invernizzi, Swank, Juel, & Meier, 2003). The PALS has reliability coefficients ranging above .80 and has strong evidence of predictive validity (Invernizzi, Justice, Landrum & Booker, 2004).

The *Measures of Academic Progress (MAP) Reading* (NWEA, 2009) assessment, used by three schools, is a norm-referenced assessment of reading achievement. Measures of reliability and validity of the MAP test, as reported by NWEA, are high with reliability estimates between .70 and .90 and predictive validity estimates of future reading ability ranging from .65 to .85 (NWEA, 2009).

The AIMS Web Plus assessment, used by one school, is a norm-referenced assessment of early literacy. The AIMSweb test has test-retest reliability estimate of .94 and has strong predictive validity for scores on the Iowa Tests of Basic Skills (Pearson, 2012).

Which schools participated?

This study includes 12 schools across two states. Two additional project schools did not participate in the 2018-19 evaluation. Among the 12 schools, nine are in the state of Wisconsin,

while the remaining three are in South Carolina. These 12 schools partnered with five local Boys & Girls Clubs. Five schools are within a large, urban district. The remaining seven schools are located in small, rural, communities. Schools selected for the study 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.

How were students consented?

Students in Kindergarten[1], first, or second grade in the 12 schools were invited to participate in the study. Students with an Individualized Education Plan (IEP) or who were designated as English Language Learners (ELL) were not eligible. Method of consent (passive vs. active) varied by school. Passive consent procedures involved a total of five outreaches to families; four through weekly letters home and at least one additional outreach method, which varied at each school. A total of 1,361 students were ultimately consented, by way of passive or active consent. Of these, 979 were determined to be eligible for the RD study.

How were students assigned to FF or business as usual literacy instruction?

Consented students within each grade level within each school were ranked, from lowest to highest, with regards to their fall (baseline) literacy assessment results. Within each grade level within each school, students with the lowest baseline assessment scores were assigned to receive FF, up to the capacity of the site to provide tutoring. Thus, each grade level within each school had a unique eligibility cut score; students at or below this score received FF while those above did not. The evaluation team identified the eligibility scores for each grade level within each school independent of the program and school. When more students were tied with baseline literacy scores at the eligibility cut off than the program could accommodate, they were randomly assigned to groups. For later analytic purposes, the scores of students randomly assigned to business as usual were modified up +.001 standard deviations. This way their baseline literacy score would place them in the business as usual student group.

Was there any evidence of score manipulation at the eligibility cut point?

The cut point is a crucial part of the RD design. It is based on the premise that those near it, but on opposite sides, are comparable. Given that the cut-line is so important both for the program

[1] Two of the three South Carolina schools did not have Kindergarten students.

implementation and analysis, it is important to ensure that manipulation of participant scoring and placement around this cut-line has not occurred. Score manipulation may manifest as a disproportionate number of students with scores on the forcing variable that barely qualified for participation. Manipulation was not possible in the current study, however. Literacy assessments were conducted within normal testing procedures and environments by schools. FF staff were not involved in the assessment process. Further, eligibility cut points were not known until after assessments had been completed.

Along with contextual evidence, the presence of score manipulation can be tested statistically (McCrary, 2008). To test for this, a density test of discontinuity was conducted using the Stata command, rddensity. This test, also known as the manipulation test, looks for evidence that assignment involved some sort of self-selection or falsification of a participant's true status based on baseline scores. The recommended testing method is to use a robust, bias-correction method, as opposed to a conventional method (Calonico, Cattaneo, & Titiunik, 2018). Results indicate the absence of data manipulation ($t = 0.734$; $p = 0.463$).

Finally, a histogram reflecting the distribution of students by group is shown below in Figure 1, with those below the "0" point on the forcing variable (baseline literacy score) receiving FF programming, and those above receiving business as usual literacy instruction. Again, there is no evidence that a disproportional number of students' literacy scores made them barely eligible for participation in FF than barely ineligible.

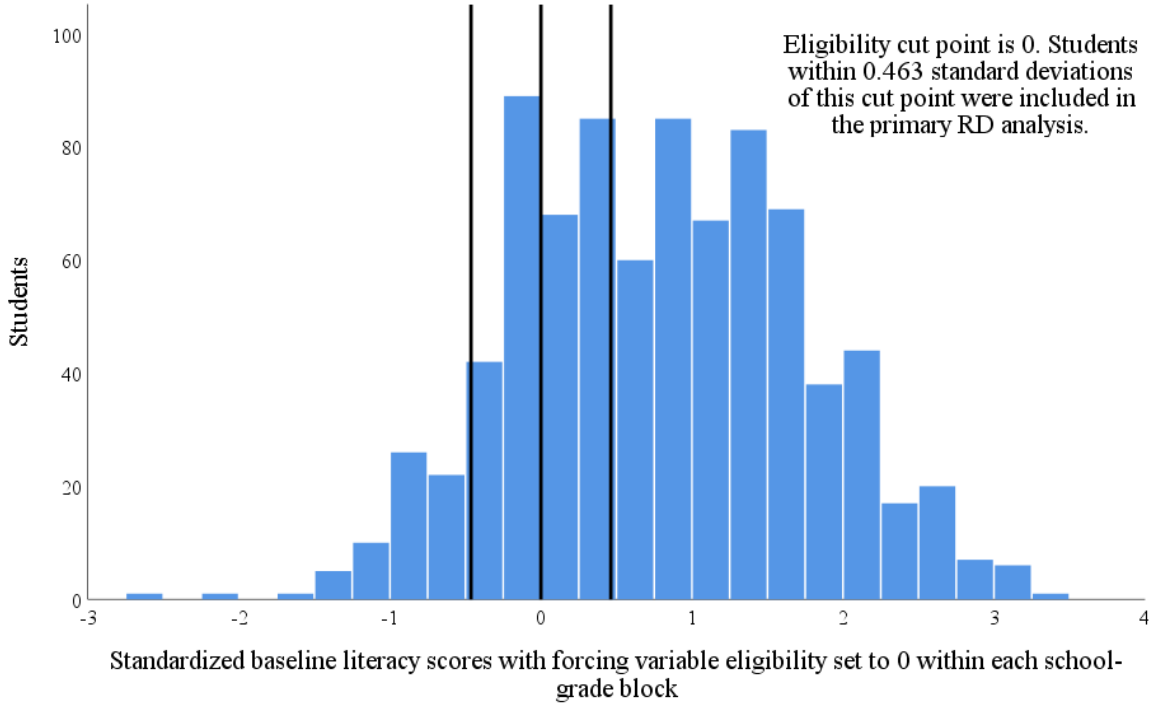

Figure 1: Distribution of standardized literacy scores relative to the eligibility cut point.

What are the characteristics of students included in the study?

The study sample consisted of mostly students of color (57.3%) and students eligible for free or reduced lunch (80.4%) (Table 1). Within the total study sample, more FF participants were eligible for free/reduced lunch (86.3%) than were business as usual students (78.7%). However, nearer the eligibility cut point, both groups had the same free/reduced lunch eligibility rate (85.1%).

Table 1: Study participant characteristics

	Total Sample			*Analytic Sample (Within 0.463 Standard Deviations of the Eligibility Cut Point)*		
Group % by category	Overall	FF Participant	Business as Usual	Overall	FF Participant	Business as Usual
Consented students	979	222	757	262	121	141
Female	53.5%	50.4%	54.4%	50.8%	48.8%	52.5%
Black	51.3%	59.9%	48.8%	53.4%	57.0%	50.4%
White	42.3%	34.7%	44.6%	40.8%	41.3%	40.4%
Free/Reduced Lunch	80.4%	86.3%	78.7%	85.1%	85.1%	85.1%
Kindergarten	28.9%	29.7%	28.7%	23.3%	28.1%	19.1%
1st Grade	34.4%	37.4%	33.6%	37.0%	38.8%	35.5%
2nd Grade	36.7%	32.9%	37.8%	39.7%	33.1%	45.4%

The RD analytic sample is restricted to students who are near the eligibility cut point on the forcing variable (baseline literacy assessment in the current study of FF). Identifying the analytic sample requires the designation of a "bandwidth" of proximity to the eligibility cut point. The bandwidth is the distance from the eligibility cut point, centered at 0, within each grade level within each school. Students within the bandwidth are included in the analysis while those outside the bandwidth are not included.

For bandwidth selection, the rdrobust package in R was utilized. This package utilizes a data-driven bandwidth selection process. Several options are available for the determination of the bandwidth. For this analysis, the default, mean squared error (MSE) bandwidth, was used. The development of the package and the data-driven process by which the bandwidth is determined is based broadly on the work of Imbens and Kalyanaraman (2012), with specific application to R through the subsequent work of Calonico, Cattaneo and Titiunik (2014).

Results of the rdrobust procedure indicated an optimal analytic bandwidth of 0.463 standard deviations on either side of the eligibility cut point. Two hundred sixty two students were within this bandwidth (121 FF participants below the eligibility cut point; 141 business as usual students above the cut point). Within this bandwidth, the overall attrition rate was 8.7% with a differential attrition rate of 2.0 percentage points (Table 3). Among the 25 students who attrited, six were documented to have moved and/or transferred schools during the course of the school year. The

remaining students did not have spring scores included in the data pulls from their respective districts. These low attrition numbers suggest the study's internal validity was not likely impacted by attrition.

Table 2: Attrition within the 0.463 standard deviations bandwidth

	Total	Participant	Comparison
Total Study Participants	287	131	156
Final Sample	262	121	141
Attrition Rate	8.7%	7.6%	9.6%

What is the functional form of the relationship between baseline and post literacy scores?

A critical part of RD analysis is to ensure models include the correct functional form between the baseline (forcing) variable and the outcome. A visual inspection of the relationship suggests the functional form of the relationship between baseline and post literacy scores, within the bandwidth, was linear (Figure 2). Thus, in our impact analysis, we include the linear relationship of baseline and post literacy scores.

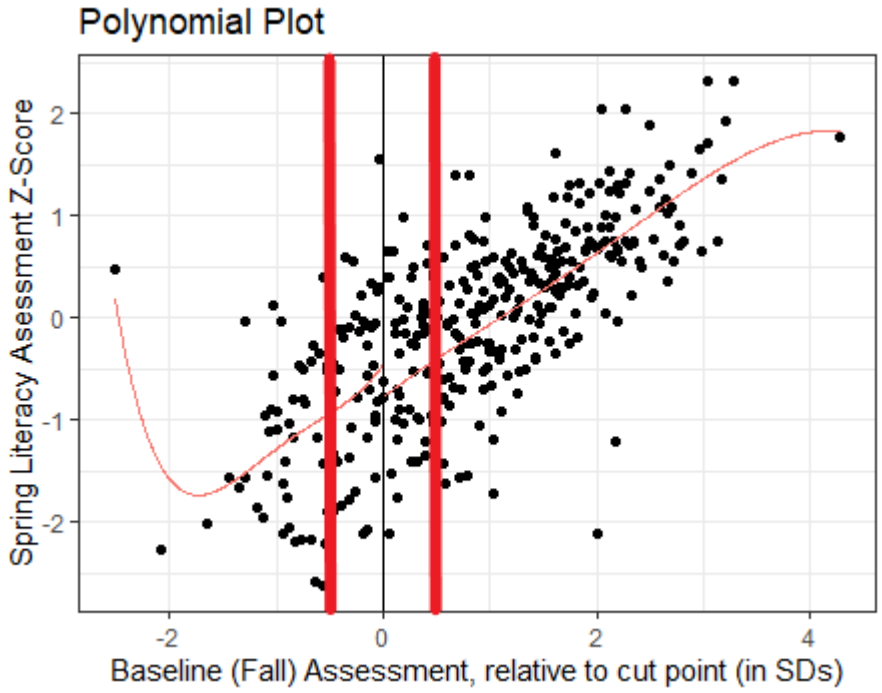

Figure 2: Polynomial Plot (no X's): Results of the rdrobust processes suggested including students within 0.463 standard deviations of the eligibility cut point.

How were data analyzed to determine impact?

We used the *rdrobust* procedure in R to measure the impact of FF. rdrobust uses local linear regression to estimate two separate regression lines, one for FF participants and one for comparison students. These lines are compared to determine impact (Hahn, Todd, & Van der Klaauw, 2011). The primary statistical model used for the analysis relies on controlling for school-grade fixed effects (blocks), which is where assignment was made. Both baseline literacy and post literacy scores were standardized within grade levels according to regional norms. Estimating the following model is equivalent to the results of local linear regression:

$$Y_{ij} = \beta_0 + \beta_1(FF_{ij}) + \beta_2(forcingvariable_{ij}) + \beta_3(forcingvariable_{ij}xFF_{ij}) + \sum_{j=1}^{J-1} \beta_{4.j} Block_j + \varepsilon_i \quad (1)$$

Where Y_{ij} is the literacy outcome score for student *i* in block *j*; β_0 is an intercept term; β_1 is the impact of participation of FF; FF_{ij} is an indicator of whether the student was in the FF condition (1) or business as usual (0); β_2 is the relationship of the baseline literacy score and the literacy outcome score for students in the business as usual condition; β_3 allows for a different relationship between baseline literacy scores and literacy outcomes for students in and not in FF; $\beta_{4.j}$ controls for the fixed effect of school-by-grade assignment block; ε_i is the error term.

Study Results

How much FF did students receive?

Each FF participant received an average of 1.57 tutoring sessions per week, with 34.7% meeting the intended program intensity of two or more sessions per week. The average family received 0.69 contacts per month, with 6.6% meeting the threshold of two or more contacts per month. School-level participation characteristics (Table 3) illustrate that sites, on average, provided each student 1.61 tutoring sessions each week and 0.64 parent contacts each month. Three sites met the site-level threshold for tutoring fidelity (at least 80% of participants with 2+ session per week), while no sites met site-level criteria for family engagement fidelity.

Table 3: Average school-level FF participation

	Mean	*Min.*	*Max.*
Weeks of programming provided (2018-19)	28.5	12	40
Tutoring (sessions/week)	1.61	0.89	2.22
Family engagement (contacts/month)	0.64	0.14	1.76
% of participants meeting tutoring fidelity threshold [a]	37%	0%	94%
% of participants meeting family engagement fidelity threshold [b]	5%	0%	31%

[a] Fidelity for tutoring is defined as a minimum of 2 sessions/week

[b] Fidelity for family engagement is defined as a minimum of 2 contacts/month

What is the impact of one year of Future Forward participation on the literacy development of lower primary students compared to students receiving business as usual literacy instruction?

The results indicate FF resulted in a non-statistically significant, but positive, impact of 0.164 standard deviations ($p = 0.394$) (Table 4). Robustness checks were conducted with two variations of this primary model. One model clustered the error terms by school-grade, as opposed to including them as fixed effects, and also included a set of demographic covariates. The results of this model showed a non-significant impact of 0.249 standard deviations ($p = 0.293$). A second robustness check involved pooling the results of separate regression equations for students within each of 22 school-grade blocks. This resulted in a non-significant impact of 0.146 standard deviations ($p = 0.332$). An additional check was conducted by comparing the primary model results using the original bandwidth selection method, mean squared error (MSE), with a second, data-driven bandwidth selection method, coverage error rate (CER), using the same school-grade fixed effects modeling approach. Results using the CER method, with a bandwidth of 0.331 SDs, again showed a positive, non-significant impact of 0.108 standard deviations ($p = 0.630$).

Another step in analyzing the validity of a RD study's results involves checking for the presence of other discontinuities along the forcing variable distribution. To check for the presence of other discontinuities, we adjusted the eligibility cut point up and down by 0.1, 0.2, 0.3 and 0.4 standard deviations. Discontinuities were checked using the same primary model (equation 1) as was used in the main analysis. We did not find any evidence of additional discontinuities (Table 5). No impact estimates at the other eight cut points were statistically significant.

Table 4: RD results for the primary and robustness check models

Model Structure	Bandwidth Selection Method	Bandwidth (SDs)	Students below cut point within bandwidth	Students above cut point within bandwidth	β	Standard Error	$p > z$
The primary modeling strategy used Local Linear Regression with school-grade fixed effects.	MSE	+-0.463	121	141	0.164	0.192	0.394
This robustness check replicated the primary modeling but changed the bandwidth selection method to CER.	CER	+-0.331	107	92	0.108	0.224	0.630
This robustness check used Local Linear Regression with school-grade clustered error terms and a vector of student demographics.	MSE	+-0.525	134	158	0.249	0.237	0.293
This robustness check involved fitting separate Regression models within each school-grade block. Results were pooled and weighted by the inverse of the error term.	Manual	+-0.500	131	153	0.146	0.150	0.332

Table 5: Results of models checking for other discontinuities with forcing variable cut point adjusted…

	MSE Bandwidth (SDs)	Students below cut point within bandwidth	Students above cut point within bandwidth	β	Standard Error	z	$p > z$
down 0.1 standard deviations.	+-0.662	97	229	-0.219	0.195	-1.125	0.261
down 0.2 standard deviations.	+-0.515	73	164	-0.029	0.217	-1.132	0.895
down 0.3 standard deviations.	+-0.402	52	120	0.030	0.262	0.907	0.907
down 0.4 standard deviations.	+-0.370	45	75	0.316	0.325	0.971	0.332
up 0.1 standard deviations.	+-0.407	120	134	-0.138	0.226	-0.612	0.541
up 0.2 standard deviations.	+-0.474	153	138	-0.018	0.204	-0.090	0.928
up 0.3 standard deviations.	+-0.500	159	150	0.152	0.193	0.786	0.432
up 0.4 standard deviations.	+-0.533	176	156	0.281	0.154	1.823	0.068

Discussion

In this study we used a multi-site, blocked RD design to test the impact of FF on the literacy development of kindergarten, first, and second grade students in 12 schools. Although in the 2018-19 school year participating schools were still learning how to implement FF, the results suggest there was still potentially a positive impact of FF on student literacy. However, our ability to determine the actual impact of FF with confidence was limited. One limitation was a lack of statistical power. RD studies require a much larger sample to measure impact than do randomized control studies. This is because RD studies only include participants whose forcing variable values are close to the eligibility cut point. As a result, most students are excluded from analyses. In the current study, the primary analysis only included 262 out of 848 study participants. With the small sample, and the corresponding larger standard errors estimates, the minimum detectable effect size (MDES)[2] is approximately 0.58 standard deviations, much larger than was expected. To put it in another way, given the 0.164 standardized impact found in the current study, the analysis would need to include 540 students to have an 80% chance of detecting an impact of that size.

Second, FF participants included in the analyses (within the bandwidth) represented the students with the strongest baseline literacy skills among all the FF participants. We learned in our previous studies of FF (Jones & Christian, 2020) that students with the greatest need for literacy support benefit the most from FF. However, these students were mostly excluded from the impact analyses. This is another consequence of using an RD design to measure impact. Again, only the students near the eligibility cut off were included in the analyses. These students did not have the greatest need for literacy support among the students in the study.

Another limitation of the current study is that most FF participants only received a fraction of the program in the 2018-19 school year. This was because many of the sites were still learning how to implement FF. This was especially apparent in the parent engagement component of FF, as only a small percentage of families received the intended amount of parent engagement. Even considering these limitations, the magnitude of the positive effect was consistent with the overall effects found in previous studies of FF/SPARK from the previous i3 grant, which is promising.

[2] The smallest statistically significant impact for a two-tailed test and 80% power.

These limitations will factor less into the 2019-20 and 2020-21 studies of FF, both of which will use a randomized control design (RCT). These studies will include all students in impact analyses, increasing the statistical power to reliably measure impacts and capturing many more struggling readers. Furthermore, the 2018-19 program year allowed sites to gain a better understanding of how to achieve levels of fidelity in tutoring and family engagement. This development will help sites increase their effectiveness in the 2019-20 and 2020-21 school years. Because of these improvements to both the program and the study methods, it is expected that the next FF studies will provide a more reliable and accurate estimate of its impact. Through these studies, the potential of FF for helping address the large gaps between the literacy needs of U.S. students and the literacy supports available to them will be better understood.

References

Bajari, P., Hong, H., Park, M., & Town, R. (2011). *Regression discontinuity designs with an endogenous forcing variable and an application to contracting in health care* (No. w17643). National Bureau of Economic Research.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: an updated systematic review. *Annals of Internal Medicine, 155*(2), 97-107.

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.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal, 14*(4), 909-946.

Chang, H. & Romero, M. (September, 2008). *Present, Engaged, and Accounted For: The Critical Importance of Addressing Chronic Absence in the Early Grades*. National Center for Children in Poverty.

D'Agostino, J. V., Lose, M. K., & Kelly, R. H. (2017). Examining the sustained effects of Reading Recovery. *Journal of Education for Students Placed at Risk*, *22*(2), 116-127.

D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis, 26*(1), 23-28.

D'Agostino, J. V., & Harmey, S. J. (2016). An international meta-analysis of Reading Recovery. *Journal of Education for Students Placed at Risk*, *21*(1), 29-46.

Dearing, E., Kreider, H., Simpkins, S., & Weiss, H. B. (2006). Family involvement in school and low-income children's literacy: Longitudinal associations between and within families. *Journal of Educational Psychology, 98*(4), 653.

DeWalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes. *Journal of General Internal Medicine, 19*(12), 1228-1239.

ESSA (2015). *Every Student Succeeds Act of 2015*, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Graue, E. & Hawkins, M. (2010). I always feel like they don't know anything about us: Diverse families talk about their relations with school. In Miller-Marsh, M. and Turner-Vorbeck,

T. (Eds.) *(Mis) Understanding Families: Learning from Real Families in our Schools* (pp.109-125) New York, NY: Teachers College Press.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading, 30*(3), 227-248.

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

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies, 79*(3), 933-959.

Inns, A., Lake, C., Pellegrini, M., & Slavin, R. (2018). *A synthesis of quantitative research on programs for struggling readers in elementary schools*. Available at www.bestevidence.org. Manuscript submitted for publication.

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.

Jeynes, W. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban Education, 47*(4), 706-742.

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*, DOI: 10.1080/10824669.2020.1809419

Lin, Q. (2003). *Parent Involvement and Early Literacy*. Harvard Family Research Project.

Lindo, E. J., Weiser, B., Cheatham, J. P., & Allor, J. H. (2018). Benefits of structured after-school literacy tutoring by university students for struggling elementary readers. *Reading & Writing Quarterly*, *34*(2), 117-131.

Little, P. M. (2009). *Supporting Student Outcomes through Expanded Learning Opportunities*.

Harvard Family Research Project. Retrieved from: http://www.hfrp.org/out-of-school-time/publications-resources/supporting-student-outcomes-through-expanded-learning-opportunities

Mehta, N. K., Lee, H., & Ylitalo, K. R. (2013). Child health in the United States: recent trends in racial/ethnic disparities. *Social Science & Medicine, 95*, 6-15.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698-714.

Moore-Hart, M., & Karabenick, S. A. (2009). Becoming successful readers: A volunteer tutoring program for culturally diverse students. *Literacy Research and Instruction*, *48*(2), 149-171.

Northwest Evaluation Association. (2009). *Technical Manual for Measures of Academic Progress and Measures of Academic Progress for Primary Grades.* Lake Oswego, OR.

Pearson Education. (2008). AIMSweb. San Antonio, TX: Author

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review, 45*(2), 250-267.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79*(1), 3-38.

Sheldon, S. B. & Epstein, J. L. (2004). Getting students to school: Using family and community involvement to reduce chronic absenteeism. *School Community Journal 14*, 39.

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

Sonnenschein, S., Baker, L., Serpell, R., & Schmidt, D. (2017). Reading Is a Source of Entertainment: The Importance of the Home Perspective for Children's Literacy Development. In Roskos, K. A. (Ed.), *Play and literacy in early childhood: Research from multiple perspectives*. Routledge.

Sudano, J. J., & Baker, D. W. (2006). Explaining US racial/ethnic disparities in health declines and mortality in late middle age: the roles of socioeconomic status, health behaviors, and health insurance. *Social science & medicine, 62*(4), 909-922.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology, 51*(6), 309.

Toffler, A., & Toffler, H. (1995). Getting set for the coming millennium. *The Futurist, 29*(2), 10.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2017). Reading Performance - NCES. Retrieved from:.https://nces.ed.gov/programs/coe/pdf/coe_cnb.pdf

U.S. Department of Education. (2016). *Chronic absenteeism in the nation's schools: An unprecedented look at a hidden educational crisis.* Retrieved from: http://www2.ed.gov/datastory/chronicabsenteeism.html

Wasik, B. A. (1998). Using volunteers as reading tutors: Guidelines for successful practices. *The Reading Teacher, 51*(7), 562-570.

This work was supported by an Education Innovation and Research (EIR) grant by the U.S. Department of Education. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

The Office of Socially Responsible Evaluation in Education conducts rigorous evaluations and research on issues relevant to providing students from all backgrounds with equitable education opportunities.

Curtis Jones is the Director of the Office of Socially Responsible Evaluation in Education at the University of Wisconsin Milwaukee and Principal Investigator of the Future Forward evaluation project.

John Bowser is a Senior Research Associate in the Office of Socially Responsible Evaluation in Education at the University of Wisconsin Milwaukee.

Mikhail Pyatigorsky is the President of Pyatigorsky Consulting.

Cristofer Price is a Principal Associate in Social & Economic Policy at Abt Associates.

For more information about this report or the evaluation of Future Forward, please contact Curtis Jones at jones554@uwm.edu or visit www.uwm.edu/sreed.

For more information about the Future Forward program, please contact Kate Bauer-Jones at kbauerjones@edanalytics.org or visit https://futureforwardliteracy.org/.

UWM
UNIVERSITY of WISCONSIN
MILWAUKEE