2019 Impacts
of the Kids Read Now Summer Reading Program

A QUASI-EXPERIMENTAL IMPACT STUDY

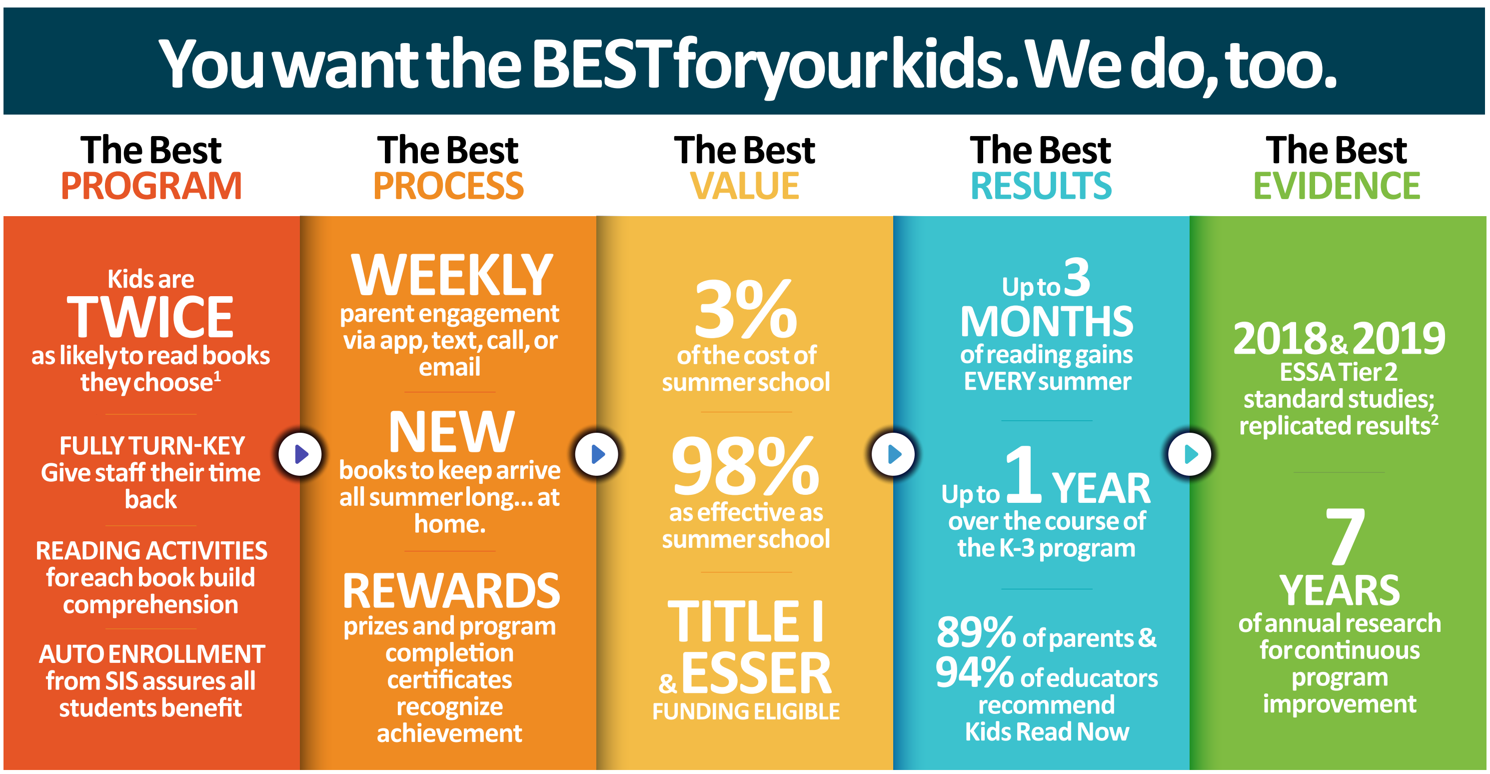
Abstract

Drawing on administrative data and reading achievement data provided by two Midwestern school districts for three participating Kids Read Now schools, this successive year study provides the opportunity to study the reading outcomes of Kids Read Now students. Relying on data from the three schools, we contrast the reading outcomes for KRN student participants and a matched control group of non-participants with replicable results.

Geoffrey D. Borman
Measured Decisions, Inc.
Hyunwoo Yang
University of Wisconsin—Madison

Significant Finding Summary

Drawing on administrative data and reading achievement data provided by two Midwestern school districts for five schools, we analyze the literacy impacts of a replicable summer reading program, Kids Read Now (KRN). The program includes both school-based and home-based components that together encourage students to remain engaged in reading high quality books over the summer months. We apply propensity score matching methods to match participating KRN students with similar comparison students.



[1] Self-selected reading is twice as powerful as teacher-selected reading in developing motivation and comprehension (Guthrie & Humenick, 2004).
[2] Less than 5% of reading studies are replicable. Kids Read Now is one of them.

"KRN offers consistent and replicable positive impacts, and it is a replicable model that can be scaled. This is highly important."

– **Geoffrey D. Borman, Ph.D.**
Foundation Professor of Quantitative Methods and Education Policy
Arizona State University, Mary Lou Fulton Teachers College

Our results suggest that KRN participants outperformed comparison group students, with a mean effect size of d = .15. Additional model estimates of the impacts for those students who read more of the books provided by KRN revealed that those who received all 9 books realized an effect size of d = .21 relative to the outcomes for matched comparison students. Supplemental analyses revealed larger impacts in Battle Creek d = .20 relative to Troy City, d= .06. Finally, the estimated effects seemed particularly strong for first grade students, d = .27.

In addition to the statistical significance of these results, their practical significance is considerable. Applying the widely-used criteria from Bloom et al. (2008), the average effect of KRN equaled over 2 months of learning, or greater than 23% of the learning that takes place over a typical 9-month school year. Considering the full impact of KRN, we find that those students who received all 9 books attained the equivalent of over 3 months of learning, or approximately one-third of the learning taking place over the school year. Similar to the 2018 results for KRN reported by Borman, Yang, & Xie (in press), when students and parents commit to the full KRN program, these results suggest that the impact of KRN can more than compensate for the 2 months of summer learning loss typically experienced by low-income students.

Drawing on administrative data and reading achievement data provided by two Midwestern school districts for five schools, we analyze the literacy impacts of a replicable summer reading program, Kids Read Now (KRN). The program includes both school-based and home-based components that together encourage students to remain engaged in reading high-quality books over the summer months. We apply propensity score matching methods to match participating KRN students with similar comparison students.

In the following pages of this report, we describe our methodology, including the student and school samples, information about how KRN was implemented in the two participating districts, and the measures that we used for matching and analysis of program impacts. Next, we present the analytical results, which first demonstrate that we achieved baseline equivalence between the KRN and non-KRN samples on all measures. Because treatment and control students were within 0.16 *SD*s on all three pretest measures, the analytical sample clearly meets the baseline equivalence criterion of 0.25 *SD*s for quasi-experimental studies established by the What Works Clearinghouse (2020). Our presentation of our estimates of the overall impacts of KRN and the estimates of the full impacts of the program follow. Finally, in our Discussion section, we briefly contextualize the practical significance of the results.

METHOD

Sample

We employed data provided by the Troy City School District in Ohio and by the Battle Creek School District in Michigan to evaluate the effects of the Kids Read Now (KRN) program on students' reading achievement. The Battle Creek Public School system, which has three schools participating, is an urban school district and the Troy City School District, which has two schools participating in KRN, is located in a suburban setting. The three Battle Creek schools include Verona Elementary School, which serves 308 students in grades PK-6, 59% of whom are minority students (16% Hispanic, 30% Black, and 13% multiracial) and 84% of whom are eligible for free or reduced-price lunch. Valley View Elementary serves 560 students in grade PreK-5, 55% of whom are minority students (22% Black, 13% Asian, and 11% multiracial) and 87% of whom are eligible for free and reduced-price lunch. Finally, Dudley Elementary school serves 391 students in grade 2-4, 17% of whom are minority students (10% Hispanic, 3% Black, and 4% multiracial) and 34% of whom are eligible for free and reduced-price lunch.

From Troy City, Hook Elementary School serves 248 students in grades K-5, 16.1% of whom are minority students (8.5% multiracial, 4.4% African American, 2% Hispanic, 1.2% Asian) and 40% of whom are eligible for free and reduced-price lunch. The second Troy City school, Kyle Elementary School, serves 212 students in grades K-5, 20.3% of whom are minority students (11.3% multiracial, 5.7% Hispanic, and 3.3% African American) and 59% of whom are eligible for free and reduced-price lunch. This sample, thus, offers some variability in terms of district and school context, and in terms of student characteristics. The overall student and school sample sizes are summarized in Table 1.

Table 1. Student Sample Sizes by School, Grade Level, and Kids Read Now Participation Status.

Grade	1		2		3		4	
	Non-KRN	KRN	Non-KRN	KRN	Non-KRN	KRN	Non-KRN	KRN
Battle Creek District								
Dudley Elementary	35	6	33	8				
Valley View Elem.	71	14			67	21	66	24
Verona Elementary							71	11
Troy City District								
Hook Elementary			22	21	28	27		
Kyle Elementary	23	11	24	18				
Total	129	31	79	47	95	48	137	35

KRN Implementation in Troy City and Battle Creek

In both districts, educators engaged in efforts to convince children and their parents to participate rather than simply accepting only those who immediately came forward. Ultimately, though, the participating students self-selected into KRN. School principals from Troy City and Battle Creek applied different strategies to encourage students to enroll. In Troy City, the school principal focused program enrollment on students with low reading scores, encouraging students with the greatest needs to participate. On the other hand, Battle Creek invited all students, without regard to pre-summer reading scores. Our quasi-experimental methods account for these and any other implementation differences across schools by restricting both matching of students and analyses of their outcomes to *within-school* comparisons.

In all five schools, teachers received a book wish list (catalog) of 120 books from which they could help their students select their nine books. Students selected their books with the help of their teacher, received parent approval to participate, and were told they would be eligible to receive up to nine new, free-to-keep books over the summer and would win a prize for reporting reading their books. The schools hosted a Family Reading Event where students received their first three books and parents received reading tips. Students were encouraged to read the books and respond to reading comprehension activities specific to each book, and at the Lexile level of the book, printed on a sticker affixed to the inside of each book. Weekly calls, text messages, emails and smart phone app messages from KRN asked parents to respond after each book was completed, and at least one comprehension activity was completed. After parents reported to KRN that their child had read a book, the KRN staff mailed a new book from those originally self-selected by the child, directly to the student's home. Students who read all nine books got a prize and a certificate of recognition when they returned to school in the fall.

Measures

Dependent variable. We used students' test scores in fall 2019 as a post-treatment measure of impact. For the Ohio schools, we used aimswebPlus test scores, which comprehensively measure children's early literacy abilities, including reading and vocabulary skills as well as silent reading fluency (aimswebPlus, n.d.). For the students in Michigan, we used the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP)

Reading Fluency scores (NWEA, n.d.), which assess students' oral reading fluency, comprehension, and foundational reading skills. Given that the two districts used different tests but measured overall students' academic ability within the same general domain of literacy, we standardized the test scores within each district and grade level to have the same mean and standard deviation across the districts, as suggested by the work of May, Perez-Johnson, Haimson, Sattar, & Gleason (2009).

Independent variables. For students' academic background information, we used three aimswebPlus or NWEA scores as pretest scores from fall 2018, winter 2019, and spring 2019. In the same way as the fall 2018 dependent variable, these pretest scores were also standardized within district and grade level. These three scores served as pretest measures of students' reading achievement before implementation of the KRN program during summer 2019. In addition to the pretest measures, the districts provided indicators of students' gender, race/ethnicity, economically disadvantaged status, and indicators of each student's school and grade level. Specifically, gender was a binary code (1=female, 0=male), and the student race/ethnicity indicator was coded as a series of binary variables to indicate five possible racial/ethnic groups: Asian, White, Hispanic, Black, or multiracial. Economically Disadvantaged Status (EDS), an indicator of family poverty, was determined by whether a student was eligible for free or reduced-price lunch, was coded 1 = EDS, 0 = non-EDS.

For treatment students, KRN provided data indicating how many books were requested and delivered to each student during the summer. All KRN students received 3 books initially from their schools, but students could obtain up to 6 additional books during the summer months. On average, KRN students received 6.39 (*SD*=2.36) books. We used these data on the number of books each student received in analytic models, which assess potential "dosage effects" of receiving more or less books from KRN. Finally, we included a dummy indicator of each student's grade and school for the propensity score matching, as the applied literature from quasi-experimental studies suggests that bias is lower when the comparison group is locally matched to treatment (Glazerman, Levy, & Myers, 2002).

ANALYTICAL STRATEGY

Propensity Score Matching Methods

In that schools and students voluntarily participated in the KRN program, estimating the treatment effect by simply comparing the posttest outcomes is likely to lead to a biased estimate of the impact of KRN. To attenuate possible selection bias, we exploited the propensity score matching (PSM) technique. We used PSM to match the treatment students and comparison students based on the baseline information described above, including the three pretest scores, demographic information, and indicators of students' schools and grade levels, which enabled us to produce treatment and control groups that should be equivalent, in expectation (Rubin, 2001). Since pretests play a key role relative to other covariates in composing comparable groups (Cook & Steiner, 2010), and because any test score has some measurement error, including the three pretests for matching and for the analytic models can effectively attenuate both selection bias and measurement error from a single test.

The literature on propensity score matching suggests that impact estimation is most efficient and effective in situations with more non-treated comparison than treated subjects (Stuart, 2010; Pirracchio et al., 2016). The situation in the current study did provide a somewhat larger pool of comparison than treatment students. Specifically, we identified a total of 161 students as KRN participants in 1st grade through 4th grade, who had complete data from the two districts. A total of 440 students, who were enrolled at the five schools but did not participate in the KRN program was identified as the comparison group pool.

We used one-to-three matching with replacement. We chose one-to-three matching, a ratio matching method, rather than one-to-one nearest neighbor matching because we detected the comparison group's propensity scores exhibited a highly right-skewed distribution. Due to this outcome, conventional one-to-one matching would result in suboptimal matches, with dissimilar non-KRN students being matched with many KRN students with propensity scores close to a value 1. As an example, when using a one-to-one nearest matching algorithm with replacement a single comparison student was matched to 19 of the 41 treatment students from a grade-by-school block. Because the 1:k ratio matching method can provide more flexibility and decreased variance when comparison cases are concentrated toward the tail of the distribution (Stuart, 2010; Lanza, Moore, & Butera, 2013), we employed one-to-three matching. We found that one-to-two matching did not solve the concentration problem and did not generate covariate balance while one-to-four matching produced results similar to those found with the one-to-three procedure used here. We matched each treatment student with three control students who had the nearest propensity score and allowed control students to be matched with multiple treatment students if their propensity scores were nearest with replacement (Caliendo & Kopeinig, 2008). To compensate for this in our subsequent analytical models, each control student is weighted by 1/3.

To calculate the conditional probability of receiving the treatment based on the predetermined covariates, which are called propensity scores, we used logistic regression. The logistic regression model included the covariates mentioned above and the statistical interactions

[1]See http://toptierevidence.org/programs-reviewed/annual-book-fairs-in-high-poverty-elementary-schools

between economic status and gender, race/ethnicity and gender, race/ethnicity and economic status race/ethnicity. Our matching model was specified as follows:

$$\begin{aligned}\text{logit}(Treatment) &= \alpha + \beta_1 Spring2019_i + \beta_2 Winter2019_i + \beta_3 Fall2018_i + \beta_4 Female_i \\ &+ \beta_5 EconDis_i + \beta_6 EconDis_i \times Female_i \\ &+ \sum \gamma_1 Race_i + \sum \gamma_2 Race_i \times Female_i + \sum \gamma_3 Race_i \times EconDis_i \\ &+ \sum \gamma School/grade_i + \varepsilon_i\end{aligned}$$

Specifically, we estimate the log odds of participation in the KRN treatment as a function of students' pre-treatment spring, winter, and fall test scores, gender, economically disadvantaged status, a vector of race/ethnicity indicators, the above-mentioned interaction terms, a vector of school-by-grade level indicators, and a student-specific error term, ε_i.

KRN Quasiexperimental Impact Estimates. After constructing comparable groups through matching, we formulated two main models to gain estimates of the treatment effect of the KRN program on academic achievement. The first was a doubly-robust regression model to estimate the quasi-experimental intent-to-treat (ITT) effect estimate, as follows:

$$\begin{aligned}\mathrm{E}(Y_{isg}|\mathrm{PS}) = \alpha &+ \beta_1 Treatment_{isg} + \beta_2 Spring2018_{isg} + \beta_3 Winter2018_{isg} \\ &+ \beta_4 Fall2017_{isg} + \beta_5 Female_{isg} + \beta_6 EconDis_{isg} \\ &+ \beta_7 EconDis_i \times Female_i \\ &+ \sum \gamma_1 Race_i + \sum \gamma_2 Race_i \times Female_i + \sum \gamma_3 Race_i \times EconDis_i \\ &+ \sum \gamma School/grade_i + \pi_{sg} + \varepsilon_i\end{aligned}$$

Doubly robust estimation applies the matching model and the weighted OLS model simultaneously in estimating the causal effect of treatment exposure on the outcome, $(Y_{isg}|\mathrm{PS})$, and produces a consistent estimate of this parameter if one of the two models is correctly specified (Funk, et al., 2011; Kang & Schafer, 2007; Tan, 2010). The doubly-robust estimation method in this case combines a form of outcome regression with a propensity score model for the probability of treatment exposure, which enables us to control for the remaining bias (Funk, et al., 2011; Linden, 2014; Robins et al., 2007). To improve precision, we include the full set of covariates used in the matching procedure, including the three pretest scores, indicators of gender and economically disadvantaged status, the interaction terms, and vectors of both race/ethnicity indicators and school-by-grade level indicators.

A second model that we apply is a two-stage least squares (2SLS) regression analysis to estimate the treatment-on-the-treated (TOT) effect (Angrist & Imbens, 1995; Ichimura & Taber, 2001). For our purposes, the 2SLS regression is particularly useful in two respects. First, this approach addresses the issue of participant non-compliance, in that there is often variation in the uptake or "dosage level" of the treatment. In this case, some students may read few or no books after receiving the initial three books prior to the summer, and other students may request and receive the 6 additional books offered by KRN. In this way, the 2SLS regression model can inform a more detailed estimate of the potential "dosage" effect of the treatment, as the model estimates the causal effect of each additional book that the students received as a result of participation in KRN.

To perform this analysis, the following models were estimated using a weighted 2SLS regression analysis:

$$\#Books = \alpha_0 + \alpha_1 Treatment + \alpha \boldsymbol{X} + \delta$$

$$Posttest_{Fall\ 2019} = \beta_0 + \beta_1 \#Books + \beta \boldsymbol{X} + \varepsilon$$

In this formulation, the endogenous explanatory variable (i.e., the number of books received) is regressed on the instrumental variable (i.e., treatment) in the first stage. The predicted values of the endogenous variable from the first stage are then used in place of the actual values to predict the outcome variable, $Posttest_{(Fall\ 2019)}$, in the second stage. The model in the second stage addresses our questions concerning potential dosage effects. The coefficient β_1 is the average treatment effect associated with each additional book received from KRN on the fall 2019 reading achievement outcome, or the effect of the "treatment on the treated" (TOT). The overall set of covariates in the two stages of the model, noted as αX and βX, are the same set used in the matching and doubly-robust regression analyses, and are included to improve precision.

With the 2SLS regression, we assume that the number of books requested from KRN is endogenous to treatment because various factors that may influence a student's request for more books, such as his or her motivation, aptitude, or desire to read, which may also contribute to the student's posttest achievement, independent of KRN enrollment. By using KRN enrollment as an instrument for the number of KRN books received, though, we can gain some leverage on identifying the portion of the variation that the key ingredient of the program (i.e., the books delivered to each student) that is exogenous. Though treatment was not randomly assigned, our matching methods suggest that KRN participation is unrelated to any measurable student characteristics, as treatment and control students were statistically equivalent with respect to all baseline characteristics. Similar to our quasi-experimental ITT analyses, to the extent that unmeasurable characteristics may be related to KRN enrollment, our TOT analyses also may be biased. Nevertheless, this estimation strategy is likely to be more informative than, for instance, simply estimating the linear relationship between books received and the achievement outcome. If KRN impacts student achievement only through the receipt of the student-selected books, which seems a plausible assumption Because we have no reason to believe that KRN would have an impact on students' fall posttest scores beyond the students' receipt of the books from KRN, this method can provide an estimate of the dosage effect of the program.

RESULTS

Descriptive Statistics and Balance Checks

Table 2 provides comparisons between treatment and control students on all baseline covariates prior to matching and after matching. The tabulated information includes the three pretest scores and all demographic information for the control and treatment groups. We evaluated the statistical significance of any treatment-control mean differences using a *t*-test for the three pretests and a Chi-Square test for all other dichotomous covariates. Before matching, there were no statistically significant differences between treatment and control for all three pretests. Specifically, on the three pretests, treatment students' scores were between 0.09 and 0.14 standard deviations higher than those for control students. However, as Table 2 shows, statistically significant differences were found between KRN and comparison students for many of the demographic characteristics and their corresponding interaction terms.

The right panel of Table 2 shows the same descriptive information after PSM. Although some small differences remained, none of these, including the set of pretests and demographic controls, mean differences was statistically significant. The final matched sample included 156 of the 440 control students and 110 of the 116 treatment students. The final sample was determined through several iterations. First, 142 students in the comparison group and 4 students in the treatment group were excluded because they did not have covariate differences within treatment status (e.g., all Hispanic students in a grade-by-school block were KRN recipients). Among the remaining samples, 298 controls and 112 treatment, two treatment students were

Table 2. Comparison of Baseline Student Control and Treatment Characteristics Before and After Matching.

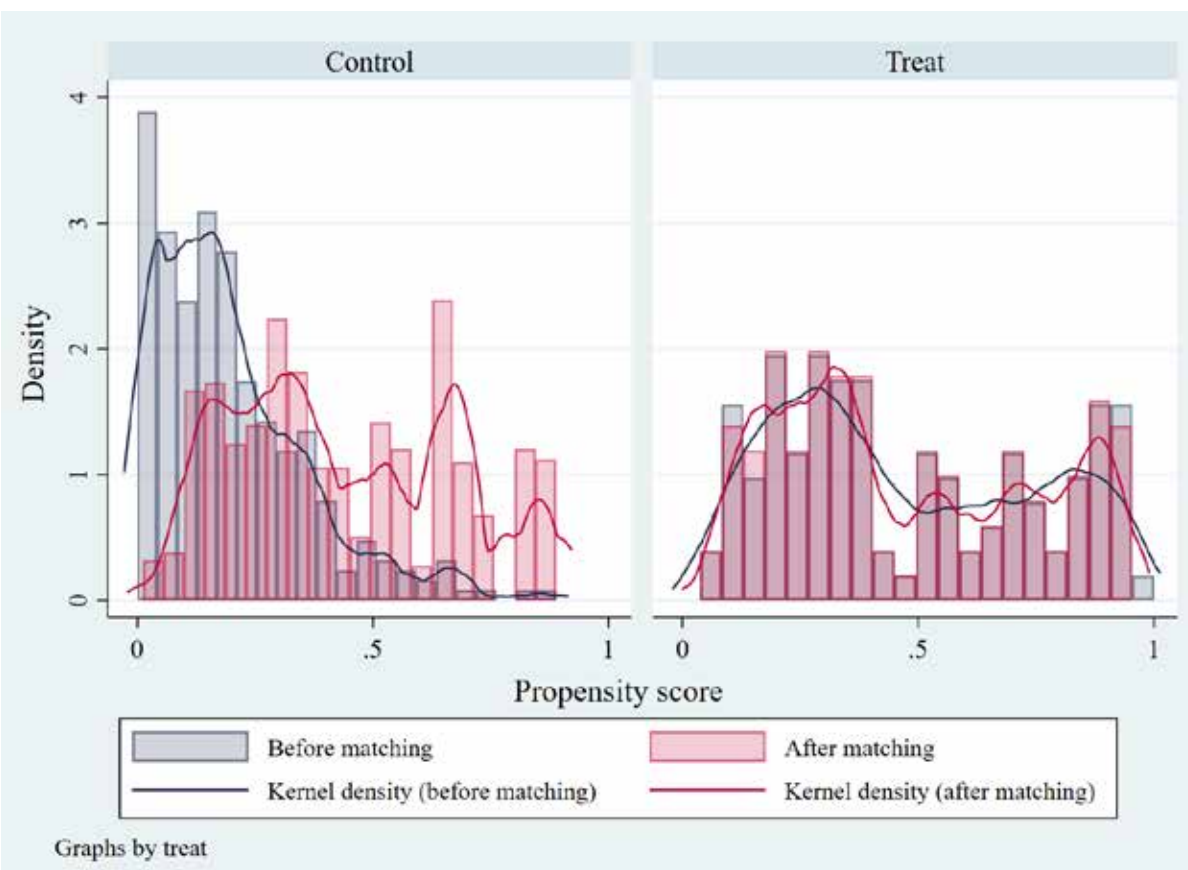
	Before matching			After matching		
	Non-KRN student	KRN student		Non-KRN student	KRN student	
Variables	Mean (SD)	Mean (SD)	Mean Difference	Mean (SD)	Mean (SD)	Mean Difference
2018 Fall	0.00 (1.03)	0.09 (1.02)	-0.09	0.30 (1.04)	0.20 (1.11)	0.11
2018 Winter	0.02 (1.00)	0.12 (0.97)	-0.10	0.38 (0.91)	0.22 (1.00)	0.16
2019 Spring	-0.03 (1.02)	0.11 (0.97)	-0.14	0.34 (0.91)	0.21 (1.01)	0.13
Female	0.41	0.50	-0.09*	0.44	0.44	0.01
Economic disadvantage	0.79	0.50	0.29***	0.60	0.58	0.02
Black	0.45	0.11	0.34***	0.18	0.15	0.03
Asian	0.07	0.02	0.04	0.02	0.02	0.00
White	0.39	0.70	-0.30***	0.68	0.70	-0.02
Hispanic	0.05	0.04	0.01	0.07	0.05	0.02
Multiracial	0.05	0.14	-0.09***	0.06	0.08	-0.03
Minority	0.55	0.25	0.31***	0.25	0.24	0.01
Female×Economic disadvantage	0.35	0.28	0.07	0.28	0.27	0.01
Female×Black	0.17	0.06	0.11***	0.08	0.07	0.01
Female×Asian	0.03	0.02	0.00	0.02	0.02	0.00
Female×White	0.16	0.35	-0.19***	0.31	0.31	0.00
Female×Hispanic	0.03	0.02	0.01	0.03	0.04	0.00
Female×Multiracial	0.02	0.04	-0.02	0.00	0.00	0.00
Economic disadvantage×Black	0.36	0.09	0.26**	0.16	0.13	0.03
Economic disadvantage×Asian	0.06	0.01	0.05**	0.02	0.02	0.00
Economic disadvantage×White	0.30	0.26	0.04	0.30	0.31	-0.01
Economic disadvantage×Hispanic	0.04	0.04	0.00	0.07	0.05	0.02
Economic disadvantage×Multiracial	0.04	0.10	-0.06***	0.05	0.07	-0.03
N	440	161		133	133	

Note: Statistical tests for mean pretest differences employ a *t*-test; and a Chi-Square test for all other binary covariates; * $p < .05$. **; $p < .01$. ***; $p < .001$.

excluded because they did not have viable propensity score matches within the comparison pool. Of the 298 comparison students, 156 students were matched on either one, two, or three occasions to 110 treatment students. Specifically, a single treatment case was matched to three comparison cases in 102 instances, in 5 instances there were five treatment cases matched to two comparison cases, and in three instances 3 treatment students were matched to a single comparison student. All analyses applied analytic weights to weight the 156 comparison cases and 110 treatment cases to an effective sample size of 133 for both groups.

The left panel of Figure 1 displays the distribution of propensity scores for comparison students and the right panel shows the distribution for treatment students. The density of comparison cases from the propensity score distribution that were not matched to treatment students are identified in blue and labeled as "Before Matching." Those 133 weighted comparison cases that were matched are identified in red and labeled "After Matching." Finally, the kernel density plots reveal very similar distributions of propensity scores for the matched control and treatment groups.

Figure 1. Propensity Score Distributions Before and After Matching for Treatment and Control.



In Table 3, we summarize the results of various balancing tests, including calculations of standardized mean differences, variance ratios, eta-squared effect sizes, and hypothesis tests of mean differences to check that treatment and control group students were statistically equivalent on covariates after matching (Lee, 2013; Richardson, 2011; Zhang, et al., 2019). The results reveal additional information to assess the outcomes of PSM, including the extent to which the matching produced more comparable treatment and control groups having smaller standardized mean differences, eta-squared effect sizes for treatment, and variance ratios closer to 1. As seen

Table 3. Treatment-Control Balance Checks Before and After Matching.

	Before Matching				After Matching			
	Mean difference	Standardized mean difference	Eta-squared effect size	Variance Ratio	Mean difference	Standardized mean difference	Eta-squared effect size	Variance Ratio
2018 Fall	0.09	0.092	0.002	0.983	-0.11	-0.102	0.002	1.143
2018 Winter	0.10	0.102	0.002	0.925	-0.16	-0.174	0.007	1.215
2019 Spring	0.14	0.137	0.004	0.913	-0.13	-0.148	0.005	1.239
Female	0.09*	0.186	0.007	1.036	-0.01	-0.012	0.000	1.000
Economic disadvantage	-0.29***	-0.715	0.081	1.53	-0.02	-0.034	0.000	1.015
Black	-0.34***	-0.683	0.099	0.384	-0.03	-0.079	0.002	0.86
Asian	-0.04	-0.165	0.006	0.395	0.00	0.000	0.000	1.003
White	0.30***	0.619	0.072	0.891	0.02	0.042	0.000	0.968
Hispanic	-0.01	-0.049	0.001	0.793	-0.02	-0.059	0.001	0.798
Multiracial	0.09***	0.417	0.023	2.606	0.03	0.112	0.003	1.423
Minority	-0.26***	-0.524	0.054	0.814	-0.02	-0.043	0.000	0.964
Female×Economic disadvantage	-0.07	-0.148	0.004	0.889	-0.01	-0.02	0.000	0.983
Female×Black	-0.12***	-0.304	0.021	0.375	-0.01	-0.043	0.001	0.871
Female×Asian	-0.00	-0.015	0.000	0.917	0.00	0.000	0.000	1.003
Female×White	0.20***	0.532	0.045	1.716	0.00	0.007	0.000	1.008
Female×Hispanic	-0.01	-0.04	0.000	0.79	0.00	0.017	0.000	1.09
Female×Multiracial	0.02	0.139	0.003	1.88	.	.	.	.
Economic disadvantage×Black	-0.26***	-0.55	0.067	0.37	-0.03	-0.090	0.002	0.826
Economic disadvantage×Asian	-0.05**	-0.198	0.01	0.222	0.00	0.000	0.000	1.003
Economic disadvantage×White	-0.04	-0.081	0.001	0.926	0.01	0.013	0.000	1.014
Economic disadvantage×Hispanic	-0.00	-0.018	0.000	0.918	-0.02	-0.059	0.001	0.798
Economic disadvantage×Multiracial	0.06**	0.315	0.014	2.419	0.03	0.121	0.003	1.511

Note: Statistical tests for the mean differences are conducted with a *t*-test for pretest scores and a Chi-Square test for all other binary covariates; * $p < .05$. **; $p < .01$. ***; $p < .001$.

in Table 3, PSM provided notable improvements for treatment-control balance, such that no baseline covariate or interaction term differences exceeded 0.25 standard deviations.

Quasi-experimental Estimates of Treatment Effects

The main analyses compare the fall posttest scores of the KRN students to those of the comparison students. The left panel of Table 4 presents the results of the doubly-robust regression model estimating the quasi-experimental impacts associated with student participation in KRN on the 2019 fall reading outcomes, controlling for the covariates and school-by-grade fixed effects. The 2019 fall test scores for the treatment group students were statistically significantly higher than those for the comparison group students. The calculated effect size derived by dividing the coefficient by the pooled standard deviation of the outcome was 0.15.

Table 4. Doubly-Robust Regression Outcomes for Intent-to-Treat Estimate and Two-Stage-Least-Squares Outcomes for Treatment-on-the-Treated Estimate.

	Intent-to-Treat			Treatment-on-the-Treated		
	Coefficient	*SE*	Effect size (*d*)	Coefficient	*SE*	Effect size (*d*)
Treatment	0.149*	0.071	0.145			
Number of books				0.023*	0.011	0.023
2018 Fall	0.200**	0.059		0.197**	0.060	
2018 Winter	0.367**	0.078		0.366**	0.078	
2019 Spring	0.338**	0.061		0.339**	0.061	
Black	0.048	0.288		0.058	0.288	
Asian	0.312	0.282		0.348	0.282	
White (Ref.)						
Hispanic	1.035**	0.241		1.043**	0.241	
Multiracial	-0.127	0.385		-0.078	0.386	
Female	0.095	0.124		0.094	0.124	
Female×Black	-0.300	0.238		-0.280	0.238	
Female×White (Ref.)						
Female×Hispanic	-1.102**	0.323		-1.084**	0.323	
Economic disadvantage	-0.148	0.119		-0.138	0.119	
Economic disadvantage×Black	0.163	0.339		0.154	0.339	
Economic disadvantage×White (Ref.)						
Economic disadvantage×Multiracial	-0.068	0.429		-0.123	0.430	

	Intent-to-Treat			Treatment-on-the-Treated		
	Coefficient	*SE*	Effect size (*d*)	Coefficient	*SE*	Effect size (*d*)
Female×Economic disadvantage	0.264	0.185		0.252	0.185	
Constant	-0.222	0.208		-0.239	0.209	
Observations	266			266		

*Note:*Interactions of Female×Asian, Female×Multiracial, Economic disadvantage×Asian,Economic disadvantage×Hispanic are omitted because of collinearity; * $p < .05$. **; $p < .01$. ***; $p < .001$. Both models include school/grade fixed effects (not shown).

The right panel of Table 4 shows the results of the 2SLS model. Because receipt of the KRN books can occur only through participation in the program and because the receipt of the books is the hypothesized mechanism through which KRN impacts reading and reading achievement, this makes KRN participation a strong instrument and book receipt is the key mediator of KRN impacts. After controlling for the covariates and school-by-grade fixed effects, the results suggest that the 2019 fall test score increased statistically significantly by over 0.02 standard deviation units for each additional book a student received from KRN. Because KRN delivers up to 6 books beyond the initial three provided, the model predicts that a KRN student who received the maximum number of 9 books would realize standard deviation increase on the fall reading achievement outcome of 0.21.

In addition to these overall impacts across grade levels and schools, Table 5 shows subgroup analyses by grade level, 1-4. Although no treatment effect estimates were statistically significant, due to limited grade-by-grade sample sizes, the point estimates for both the ITT and TOT outcomes suggest that the treatment effect may be most powerful for grade 1 students. This suggests that the youngest group of students, from grade 1, appeared to benefit the most both overall and with corresponding increases in the number of books that they requested and received from KRN.

Table 5. Kids Read Now Estimated Effects by Grade Level, 1-4.

Intent-to-Treat Estimates					
Grade	Coefficient	SE	Effect size (*d*)	p-value	N
1	0.280	0.255	0.274	0.279	63
2	0.107	0.149	0.104	0.481	32
3	0.091	0.070	0.089	0.202	84
4	0.081	0.081	0.079	0.321	87
Treatment-on-the-Treated Estimates					
Grade	Coefficient	SE	Effect size (*d*)	p-value	N
1	0.045	0.041	0.044	0.279	63
2	0.020	0.028	0.020	0.481	32
3	0.013	0.010	0.012	0.202	84
4	0.014	0.014	0.014	0.321	87

Supplemental Analyses

Impact Estimate Differences by District. Because the selection of students into the KRN program differed by school district, with the two participating Battle Creek Schools having an open enrollment process while the one Troy City school attempted to target recruitment of the lowest-performing students, we investigated whether the treatment effects varied by district. We divided the sample into each district then estimated a doubly-robust regression model similar to our main model of the quasi-experimental ITT effect.

DISCUSSION

Our results suggest that KRN participants outperformed comparison group students, with a mean effect size of nearly $d = .15$. Additional model estimates of the impacts for those students who read more of the books provided by KRN revealed that those who received all 9 books realized an effect size of $d = .21$ relative to the outcomes for matched comparison students. Supplemental analyses revealed larger impacts in Battle Creek $d = .20$ relative to Troy City, $d = .06$. Finally, the estimated effects seemed particularly strong for first grade students, $d = .27$.

In addition to the statistical significance of these results, their practical significance is considerable. Applying the widely-used criteria from Bloom et al. (2008), the average effect of KRN equaled over 2 months of learning, or greater than 23% of the learning that takes place over a typical 9-month school year. Considering the full impact of KRN, we find that those students who received all 9 books attained the equivalent of over 3 months of learning, or 33% of the learning taking place over the school year. Similar to the 2018 results for KRN reported by Borman, Yang, & Xie (in press), when students and parents commit to the full KRN program, these results suggest that the impact of KRN can more than compensate for the 2 months of summer learning loss typically experienced by low-income students.

REFERENCES

aimswebPlus. (n.d.). aimswebPlus overview. Retrieved from https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/aimswebPlus-overview.pdf

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association, 90*(430), 431-442.

Bloom, H. S., Hill, C. J., Black, A. B., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289-328.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.

Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 15*(1), 56–68.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology, 173*(7), 761-767.

Glazerman, S., Levy, D.M., & Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review.* Princeton, NJ: Mathematica Policy Research, Inc

Ichimura, H., & Taber, C. (2001). Propensity-score matching with instrumental variables. *American Economic Review, 91*(2), 119–124.