A Quasi-Experimental Study of the Impacts of the Kids Read Now Summer Reading Program

Geoffrey D. Borman
Hyunwoo Yang
Xin Xie
University of Wisconsin—Madison

Significant Findings Summary

"…we find that the impact of [participating in] KRN was equivalent to approximately 1.7 months of learning, or nearly 20% of the learning that takes place over a typical school year." Link To Conclusions

"….when students and parents take advantage of the full complement of 9 books delivered by KRN, the results are more profound.… equivalent to approximately 2.5 months of learning, or nearly 28% of the learning that takes place over a typical school year."

"…our results indicate that the impact of KRN can more than eradicate the entire 2 months of summer learning loss experienced by low-income students."

"…Kim and Quinn found that the effective home-based summer reading programs that they reviewed produced average impacts on total reading outcomes of 0.13 standard deviation units. The KRN outcomes, ranging from 0.12 to 0.18 standard deviations compare favorably. "

"…Similarly, the average impact of 0.19 standard deviation units found for the typically more intensive and expensive school-based programs are essentially the same as those we found for full implementation of KRN."

"...these results meet rigorous evidence standards, such as those established by the What Works Clearinghouse…"

Geoffrey D. Borman, Ph.D.
Vilas Distinguished Achievement Professor
Faculty in Departments of Educational Leadership and Policy Analysis, Educational Policy Studies, Educational Psychology, and Sociology
Director, Interdisciplinary Training Program in the Education Sciences, Wisconsin Center for Education Research (WCER) University of Wisconsin--Madison

For over 100 years, researchers have documented the *summer slide*, or the achievement losses by students from the end of the school year, during the spring, to the beginning of the next school year during the fall (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). A meta-analysis of this research by Cooper and his colleagues indicated that the typical student loses approximately 1 month of grade-equivalent skill or knowledge in combined math and reading achievement over the summer. In the reading domain, the summer slide has a particularly harmful impact on the achievement of students from low-income backgrounds (Alexander, Pitcock, & Boulay, 2016; Borman & Boulay, 2004; Cooper et al. 1996; Entwisle, Alexander, & Olson, 1997; Heyns, 1978). While middle-class children's spring-fall reading scores reveal small gains during the summer months, low-income children's scores show declines of over two months of grade-level equivalency (Cooper et al., 1996). As a result, during a typical year low-income children's reading skill levels fall approximately 3 months behind those of their middle-class peers—a difference equivalent to about a third of the typical amount of learning that takes place during a regular nine-month school year.

More recent analyses of seasonal learning outcomes based on the national Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set have revealed similar findings. Downey, Von Hippel, and Broh (2004) concluded that nearly every minority-white and income-based achievement gap grew faster during the summer after kindergarten than during the kindergarten and first-grade school years. Similar to the result from the previous synthesis of the summer learning literature by Cooper et al. (1996), Downey and his colleagues estimated that the reading achievement level of a child with a household income of $40,000 fell 2.5 months behind the achievement level of a child with a household income of $100,000. Burkam, Ready, Lee, and LoGerfo (2004) also found that there was stratification in summer learning for young children between kindergarten and first grade. However, their results suggested that the relationship between family income and children's summer learning was not linear. Instead, the most important summer learning differences were those concentrated mainly in the highest and lowest quintiles of the income distribution, among very advantaged and very disadvantaged children. These summer learning differences may be explained by the "faucet theory," whereby children from all socioeconomic backgrounds receive approximately equal benefits when school is in session, but once the school resource "faucet" is turned off in the summer, students from less-advantaged backgrounds have less access within their homes and communities to books and activities that promote reading growth (Entwisle, Alexander, & Olsen, 2001).

Even more disconcerting than these achievement disparities found over the course of one summer is the finding from a long-term Baltimore-based study showing that the summer learning deficits of low-income children *accumulate* over the elementary-school years, and that their achievement scores fall farther and farther behind the scores of their more economically advantaged peers as they progress through school (Alexander & Entwisle, 1996). By the end of the sixth grade, these summer reading losses produced a cumulative lag of two years in reading achievement despite the fact that lower- and higher-income children learned at essentially the same rate while in school. By the beginning of high school, two-thirds of the socioeconomic-based achievement gap was explained by summer learning differences. These summer learning differences, in turn, substantially accounted for achievement-related differences by family income in high school track placements (college preparatory or not), high school non-completion, and four-year college attendance (Alexander, Entwisle, & Olsen, 2007). As a result of this and other research on summer learning, it has become clear that these seasonal learning differences have considerable implications for understanding and addressing the persistent achievement and attainment gaps that separate economically disadvantaged students from their middle-class peers.

Combatting the Summer Slide: School- and Home-based Interventions. Though these summer-based reading deficits are profound, more optimistically two meta-analyses suggest that interventions that target summer learning can help. Policymakers and educators have adopted two primary strategies for improving children's reading achievement during the summer months: school- and home-based summer reading programs (McCombs et al., 2011). School-based summer reading interventions, which were the primary focus of a meta-analysis conducted by Cooper, Charlton, Valentine, and Muhlenbruck (2000), are designed to support children's academic growth through instructional activities led by school teachers, college and graduate students, and researchers. Cooper and colleagues identified 93 studies of summer school programs and achievement that were amenable to quantitative synthesis and indicated that the average effect size for remedial summer programs was equal to nearly one fifth of a standard deviation (d =.19).

More recently, home-based summer reading interventions have been implemented as a potentially cost-effective strategy for mitigating reading loss among low-income children (McCombs et al., 2011). Rather than teacher-directed literacy instruction, which is the prior mechanism through which classroom-based interventions aim to improve achievement, home-based programs depend on the quantity and quality of child-initiated book reading activities to promote literacy growth. A meta-analysis by Kim and Quinn (2013) concluded that both types of programs have statistically equivalent effects on students' total reading achievement (school-based programs: d =.09; home-based programs: d =.13) and reading comprehension outcomes (school-based programs: d =.25; home-based programs: d =.22). Though there are key differences between school-based and home-based models, it seems that both approaches may hold promise for stemming the summer achievement slide.

The Kids Read Now Program. Many of the summer programs identified by the Cooper et al., (2000) and Kim and Quinn (2013) research reviews, and the vast majority of those currently operating across the country are local, district-based interventions that are not guided by evidence or explicit criteria for implementation (Borman, Schmidt, & Hosp, 2016). In contrast, the Kids Read Now program is a nationally disseminated model, which is capable of widespread replication. Since 2010, the non-profit Kids Read Now organization has served over 65,000 students, delivering over 290,000 new books to nearly 40,000 K-3 students in 2018 alone (National Summer Learning Association, 2018). The Kids Read Now program is guided by the summer learning evidence base and is available through direct purchase or through select matching grants available to schools.

The Kids Read Now program consists of several key components, which we describe below, that include both school-based and home-based features. As Kim and Quinn (2013) noted, very few interventions combine school- and home-based components, yet these authors argued that more effectively forging such home-school connections is a highly desirable goal for summer learning programs. The central home-based feature of Kids Read Now is the collection of 9 books that students receive directly from Kids Read Now. Providing the free books, though, also involves several school-based components designed to address the bifurcations that researchers and policymakers have largely reinforced: that classrooms and homes are separate spheres for children's development and distinct settings where summer programs are usually implemented (Cooper et al., 2000; McCombs et al., 2011; Kim & Quinn, 2013).

Specifically, at the end of each school year, teachers initially help students select 9 books from an educator-curated "Wish List" of 150 titles that incorporate fiction, non-fiction, bilingual and multicultural choices, listed by AR, Lexile, and Fountas and Pinnell levels to help children self-select the books they would like to receive and read over the summer. At an end-of-year family reading night, students receive their initial three books. All parents, guardians, and caregivers are also invited to attend the event and receive direct information from teachers and written guidance, in a bilingual Parent Guide, from Kids Read Now to help them support their children's summer reading. Among other things, the guidance provides clear, common-sense ideas to encourage parent-child discussions focusing on dialogic reading activities, extended discourse about text, and elaborative reminiscing, which prior research within the context of preschool and emergent literacy has suggested, promotes improved oral language, comprehension, and vocabulary outcomes (Hart & Risley, 1995; Reese, Sparks, & Leyva, 2010). During the school-based event, parents commit to encouraging their students to read over the summer and helping them stay on track. In these ways, teachers are involved in helping students select appropriate books to match each student's reading ability and interests and the school staff connect with families to encourage summer reading.

The primary home-based component of Kids Read Now is delivery via mail of up to an additional 6 free books to each participating child. The program model encourages students to read each book and discuss it with their parent or guardian. In addition, within each book a "Discovery Sheet" is affixed with questions that facilitate the discussion with a parent, challenge children to think about what they read, and foster improved comprehension. Each "Discovery Sheet" incorporates text-to-self, text-to-text, text-to-world, and creativity questions or activities, each written at the reading level of the book, and specific to each book. Each week throughout the summer, Kids Read Now sends reminder calls, texts, or emails, in the parent's preferred language and mode of contact, asking if the participating child has read a book that week. After completing each book, parents reply to the weekly call, text, or email to inform Kids Read Now of the book(s) that the child has read. Upon receiving this information, Kids Read Now sends another book to the home of the participating student. Additionally, a start-of-summer postcard is mailed home with program information and to verify postal deliverability. Bi-weekly motivational messages are sent via text, email, or voice call with tips on reading and literacy. If students read all nine books they chose, they receive a personalized certificate and a prize at the end of the summer program. Since some schools prefer literacy-oriented prizes, the school can opt to receive more books for dissemination instead of other prize options (stuffed animals, keychains, water bottles, etc.) This book distribution model has many similarities to the annual book fair approach of Allington et al. (2010), which achieved a "near top tier" evidence rating by the Coalition for Evidence-Based Policy.[1] However, the Kids Read Now model, arguably, includes a greater emphasis on establishing the home-school connection in order to complement and support improved reading behaviors and attitudes and increased literacy achievement growth during both the school year and summer.

The Current Study. Drawing on administrative data and reading achievement data provided by two Midwestern school districts for three participating Kids Read Now schools, the current study provides the first opportunity to study the reading outcomes of Kids Read Now

---

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

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.

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. Located in central Miami County, Ohio, the Troy City School District is located in a suburban setting comprised of six elementary schools, one 6th grade building, one junior high, and one high school, and a population of 4,241 students. According to state test scores, 69% of students are at least proficient in math and 58% in reading (Niche, n.d.-d). The Battle Creek Public School system is an urban school district located in Calhoun County, Michigan. Serving a major portion of the City of Battle Creek, portions of Emmett Township, Pennfield Township, Bedford Township, and the City of Springfield, the district includes one high school, two middle schools, six elementary schools, and a new Alternative High School housed at the W.K. Kellogg Middle School (Battle Creek Public Schools, n.d.). It has 4,118 students and, according to state test scores, 13% of students are at least proficient in math and 23% in reading (Niche, n.d.-a).

Table 1 presents the specific information on the schools implementing KRN in the two districts, including the number of KRN students and the number of students not enrolled in KRN in each school. Verona Elementary School serves 308 students in grades Prekindergarten-6, 59% of whom are minority students (16% Hispanic, 30% Black, and 13% two or more races) and 84% of whom are eligible for free or reduced-price lunch (Public School Review, n.d.). 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 (Niche, n.d.-b). 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 (Niche, n.d.-c).

Table 1. Information on KRN Schools in the Battle Creek District and the Troy City District

District	School Name	# Non-KRN students	# KRN students
Battle Creek	Verona Elementary	117	88
Troy	Hook Elementary	83	64
	Kyle Elementary	70	62

KRN Implementation in Battle Creek and Troy City

Participating students from the two school districts self-selected into the Kids Read Now Program. Participating schools from the two school districts received a book wish list of over sixty books for each participating student. Students selected nine books, received parent approval to participate, and won a prize for signing up. The school hosted a Family Reading Night where students received their first three books and parents received reading tips. Students read the books and discussed reading comprehension questions printed inside each book. Weekly calls, text messages, and emails asked parents to respond after each book was completed. After parents reported that their child had read a book, the KRN program mailed a new book to the student's home. Students who read all nine books got a prize and a certificate of recognition in the fall.

The literature on propensity score matching suggests that estimation of the average treatment effect of the treated (ATT) is most efficient and effective in situations with many more control than treated subjects (Stuart, 2010; Pirracchio et al., 2016). We identified a total of 214 students as KRN participants in 1st grade through 4th grade, who had complete data and were deemed eligible for the quasi-experimental study in the two districts. A total of 270 students, who were enrolled at the three schools but did not participate in the KRN program were identified as the comparison group pool. As a result, 270 1st through 4th graders who had complete data were deemed eligible for the comparison group sample. More details of the student samples across the three schools can be seen in Table 2.

Table 2. Information on the Sampled Schools

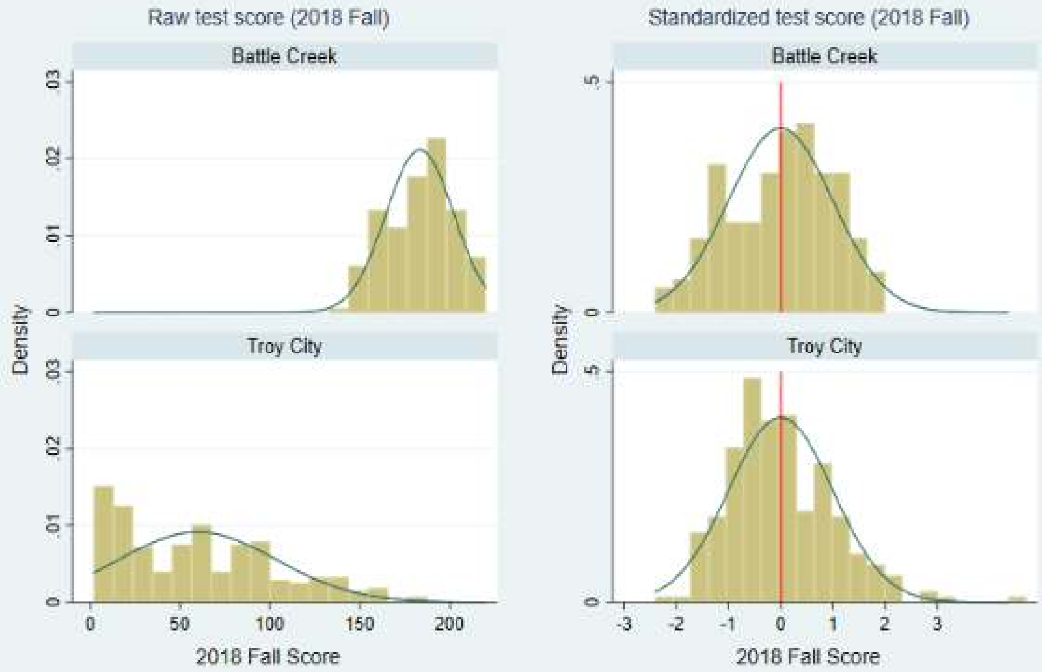
	Troy City District				Battle Creek District	
	Hook Elementary		Kyle Elementary		Verona Elementary	
Grade	Non-KRN student	KRN student	Non-KRN student	KRN student	Non-KRN student	KRN student
1	28	19	21	25		
2	34	25	21	15		
3	21	20	28	22		
4					117	88
Total	83	64	70	62	117	88

Measures

Dependent variable. We used students' test scores in fall 2018 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, 2019). For the students in Michigan, we used the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) Reading Fluency scores (NWEA, 2019), 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, we standardized the test scores within the district and grade level to have the same mean and standard deviations across the districts.

Figure 1 shows the standardized results.

Figure 1. 2018 Fall test scores before and after standardization.



Note: Raw scores are standardized by district and grade level.

***Independent variables*.** For students' academic background information, we used three aimswebPlus or NWEA scores as pretest scores from fall 2017, winter 2018, and spring 2018. 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 2018. 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), which was determined by whether a student was eligible for free or reduced-price lunch, was coded 1 = EDS, 0 = non-EDS. Finally, we included a dummy indicator of each student's grade and school for the propensity score matching as the literature on quasi-experimental studies is clear in suggesting that bias is lower when the comparison group is locally matched to treatment (Glazerman et al., 2002). Prior to matching, Table 3 shows a comparison of the posttest scores for treatment and control students.

Table 3. Comparison of Posttest Scores for the Control Group and the Treatment Group

School	Grade	Control Group			Treatment Group		
		N	Mean	SD	N	Mean	SD
Hook Elementary	1	28	-0.08	0.87	18	0.20	1.01
	2	34	0.09	0.82	25	0.43	0.95
	3	21	-0.04	1.09	19	-0.01	1.12
Kyle Elementary	1	21	0.23	1.39	25	-0.21	0.72
	2	21	-0.14	1.17	15	-0.61	0.97
	3	28	-0.02	0.95	22	0.01	0.92
Verona Elementary	4	115	0.28	0.90	87	-0.44	1.01
Total		268	0.12	0.98	211	-0.19	1.00

Analytical Strategy

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.

By using logistic regression, we calculated each student's conditional probability, namely a propensity score, to receive the treatment based on the predetermined covariates. The logistic regression model included the covariates mentioned above and the interaction term between economic status and race/ethnicity. Then we matched each treatment student with a control student who had the nearest propensity score to that of each treatment student in the terms of a one-to-one within 0.01 caliper (Lunt, 2014). Finally, we allowed a control student to be matched with multiple treatment students if their propensity scores were nearest with replacement (Caliendo & Kopeinig 2008). Various balancing tests, such as standardized mean differences, variance ratios, eta-squared effect sizes, and hypothesis test of mean differences, were computed to check that treatment and control group students were balanced on covariates resulting from PSM (Lee, 2013; Richardson, 2011; Zhang, et al., 2019).

After constructing comparable groups through matching, we utilized two main models to estimate the treatment effect of the KRN program on academic achievement. The first was a doubly-robust regression model to estimate the intent-to-treat (ITT) effect of the KRN program. A doubly-robust estimation method is a form of regression combining a model for the exposure (the propensity scores, in this case), which enables us to control for the remaining bias (Funk, et al., 2011; Linden, 2014; Robins et al., 2007). The second model was a two-stage least squares (2SLS) regression analysis to estimate the treatment-on-the-treated (TOT) effect (Angrist &

Imbens, 1995; Ichimura & Taber, 2001). The 2SLS regression is particularly useful in two aspects. First, this approach can deal with 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 receive all 9 of the 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 actually received from KRN. Both the doubly robust regression and 2SLS models included all the covariates used for PSM and school-by-grade fixed effects.

Results

Descriptive Statistics and Balance Checks

Table 4 provides background information for the samples and descriptive statistics for the baseline covariates. The tabulated information includes the three pretest scores and 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 pretests and a chi-squared test for the other dichotomous covariates. Before PSM, all three pretests and economically disadvantaged status significantly differed between the control and treatment groups. Specifically, treatment students had 0.35 standard deviation lower pretest scores relative to control students, and the treatment students were 20 percent more likely to be economically disadvantaged than their counterparts.[2] Additionally, there were seven percent more multiracial students in the control group. The right panel shows the same descriptive information after PSM. Although some differences remained, none of these mean differences were statistically significant. The final matched sample of 111 treatment and 111 control students was smaller than the original non-matched samples. This difference is due to two factors: (1) students who had at least one missing value for any covariate could not be matched and were excluded from the final sample; and (2) students whose propensity scores were not within the 0.10 caliper matching criterion were also excluded. Figure 2 shows the unmatched samples, which are represented as "off-support." Racial/ethnic indicator is driven, in part, by the very small number of Asian students in the sample. Figure 3 also visually corroborates the imbalance before PSM and balance after PSM.

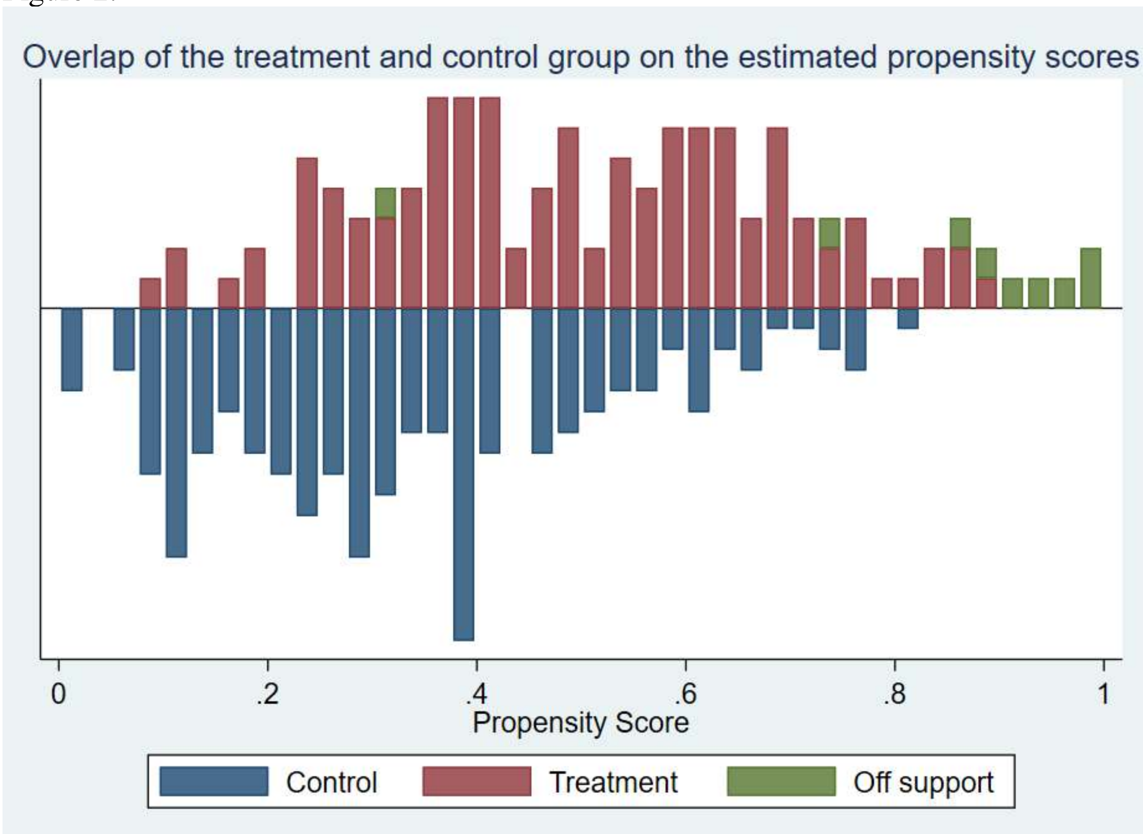
[2] Different strategies were used to target students in Troy and Battle Creek, which contributed to these baseline differences. Troy administrators directed school principals to focus program enrollment on students with low reading scores, encouraging students with **the** greatest needs to participate. **On the other hand,** Battle Creek sent enrollment materials to all students, without regard to pre-summer reading scores.

Table 4. Comparison of Baseline Characteristics for KRN before and after PS Matching

		Before PS matching				After PS matching			
Variables	Condition	N	Mean	SD	Mean Difference	N	Mean	SD	Mean Difference
2017 Fall	Control	231	0.15	0.99	0.35***	111	0.00	1.05	0.10
	Treatment	166	-0.21	0.98		111	-0.09	1.05	
2018 Winter	Control	244	0.15	0.97	0.37***	111	-0.07	1.10	0.04
	Treatment	176	-0.22	1.00		111	-0.11	1.05	
2018 Spring	Control	247	0.15	0.95	0.38***	111	-0.02	0.99	0.08
	Treatment	180	-0.23	1.03		111	-0.10	1.06	
Female	Control	268	0.51	0.50	-0.01	111	0.47	0.50	-0.02
	Treatment	174	0.52	0.50		111	0.49	0.50	
Economic Disadvantage	Control	268	0.63	0.48	0.19***	111	0.59	0.49	0.03
	Treatment	206	0.44	0.50		111	0.56	0.50	
Asian	Control	267	0.01	0.11	-0.03	111	0.00	0.00	-0.04
	Treatment	164	0.04	0.19		111	0.04	0.19	
White	Control	267	0.61	0.49	-0.04	111	0.74	0.44	0.08
	Treatment	164	0.64	0.48		111	0.66	0.48	
Hispanic	Control	267	0.09	0.29	0.04	111	0.04	0.19	-0.02
	Treatment	164	0.05	0.22		111	0.05	0.23	
Black	Control	267	0.18	0.38	-0.06	111	0.14	0.34	-0.08
	Treatment	164	0.23	0.42		111	0.22	0.41	
Two or more	Control	267	0.12	0.32	0.07***	111	0.09	0.29	0.05
	Treatment	164	0.04	0.20		111	0.04	0.19	
Minority	Control	267	0.38	0.03	0.06	111	0.26	0.44	0.04
	Treatment	164	0.32	0.04		111	0.30	0.46	

Note: Statistical test for the mean differences are conducted with a T-test for pre-test scores and a chi-squared test for all other binary covariates; * p < .05. **; p < .01. ***; p < .001.

Figure 2.



In addition to the mean difference test, Table 5 reveals additional information to assess the outcomes of PSM, and to what extent 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 in Table 5, PSM provided notable improvements to the treatment-control balance on all baseline measures. Most importantly, no baseline covariates, with the exception of the variable indicating students' race/ethnicity as Asian, revealed standardized mean differences greater than 0.25 standard deviation units. Because treatment and control students were within 0.10 standard deviation units on all three key pretest measures, the sample meets conventional criteria for baseline equivalence for quasi-experimental studies (What Works Clearinghouse, 2018). The seeming imbalance for the Asian

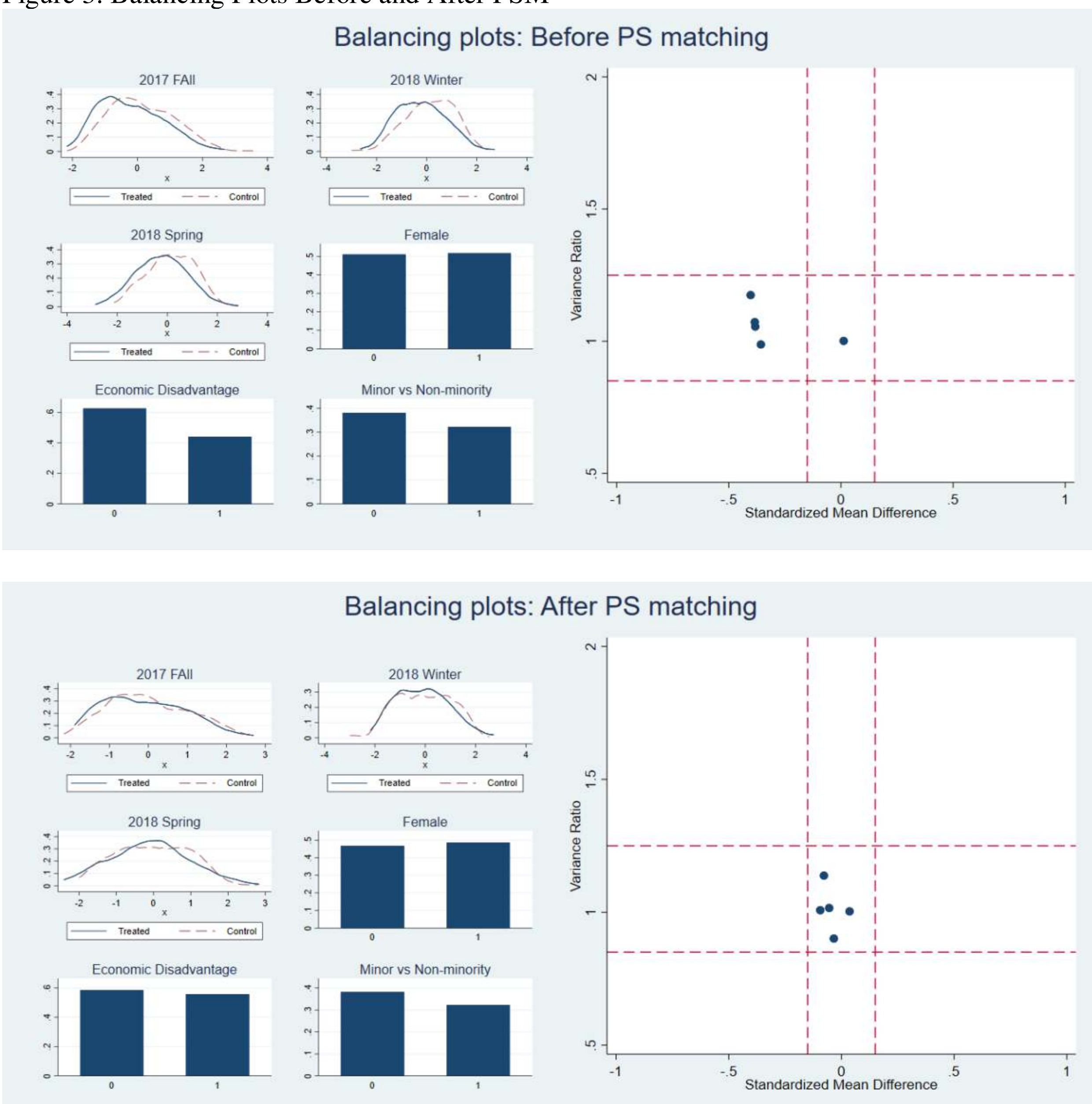
Table 5. Comparison of Balance Checks

	Before PS matching				After PS matching			
	Mean difference	Standardized mean difference	Eta-squared effect size	Variance Ratio	Mean difference	Standardized mean difference	Eta-squared effect size	Variance Ratio
2017 Fall	-0.35***	-0.36	0.03	0.99	-0.10	-0.09	0.00	1.01
2018 Winter	-0.37***	-0.38	0.03	1.07	-0.04	-0.03	0.00	0.90
2018 Spring	-0.38***	-0.40	0.04	1.18	-0.08	-0.08	0.00	1.14
Economic Disadvantage	-0.19***	-0.38	0.03	1.06	-0.03	-0.06	0.00	1.02
Female	0.01	0.01	0.00	1.00	0.02	0.04	0.00	1.00
Asian	0.03	0.24	0.01	3.18	0.04	.	0.02	.
White	0.03	0.07	0.00	0.97	-0.08	-0.18	0.01	1.17
Hispanic	-0.04	-0.14	0.01	0.57	0.02	0.10	0.00	1.47
Black	0.06	0.15	0.01	1.23	0.08	0.24	0.01	1.45
Two or more	-0.07**	-0.23	0.02	0.40	-0.05	-0.19	0.01	0.42
Minority	-0.06	-0.121	0.00	0.93	0.04	0.10	0.00	1.10

Note: * $p < .05$. **; $p < .01$. ***; $p < .001$.

and control groups having smaller standardized mean differences, eta-squared effect sizes for treatment, and variance ratios closer to 1. As seen in Table 5, PSM provided notable improvements to the treatment-control balance on all baseline measures. Most importantly, no baseline covariates, with the exception of the variable indicating students' race/ethnicity as Asian, revealed standardized mean differences greater than 0.25 standard deviation units. Because treatment and control students were within 0.10 standard deviation units on all three key pretest measures, the sample meets conventional criteria for baseline equivalence for quasi-experimental studies (What Works Clearinghouse, 2018). The seeming imbalance for the Asian racial/ethnic indicator is driven, in part, by the very small number of Asian students in the sample. Figure 3 also visually corroborates the imbalance before PSM and balance after PSM.

Figure 3. Balancing Plots Before and After PSM



Treatment effects

The main analyses compared the fall posttest scores of the KRN students to those of the comparison students. The left panel of Table 6 presents the results of the doubly-robust regression model estimating the effect of participating in KRN on the 2018 fall reading outcomes, controlling for the covariates and school-by-grade fixed effects. The 2018 fall test scores for the treatment group students were statistically significantly higher than those for the control group by 0.12 standard deviation, on average. The calculated effect size derived by dividing the coefficient by the pooled standard deviation of the outcome was also 0.12.

The right panel of Table 6 shows the results of the 2SLS model. The 2018 fall test score increased statistically significantly by 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 a 0.18 standard deviation increase on the fall reading achievement outcome.

Table 6. Doubly-Robust Regression for ITT and 2SLS Model for TOT Estimate of Treatment Effect

	Intent-to-Treat			Treatment-on-the-Treated		
	Coefficients	SE	Effect size (SD)	Coefficients	SE	Effect size (SD)
Treatment	0.12*	0.06	0.12			
Number of books				0.02*	0.01	0.02
2017 Fall	0.27***	0.06		0.26***	0.06	
2018 Winter	0.26***	0.06		0.26***	0.06	
2018 Spring	0.38***	0.07		0.38***	0.07	
Female	-0.15*	0.06		-0.14*	0.06	
Economic Disadvantage	-0.04	0.09		-0.02	0.09	
Asian	-0.03	0.22		-0.01	0.22	
White						
Hispanic	-0.13	0.15		-0.11	0.15	
Black	-0.19*	0.09		-0.18*	0.09	
Two or More	-0.18	0.13		-0.19	0.13	
Grade 1 in Hook						
Grade 2 in Hook	0.30*	0.11		0.31**	0.11	
Grade 3 in Hook	0.37**	0.11		0.38***	0.11	
Grade 1 in Kyle	0.00	.		0.00	.	

Grade 2 in Kyle	0.66**	0.23	0.71**	0.23
Grade 3 in Kyle	0.40**	0.12	0.42***	0.12
Grade 4 in Verona	0.49***	0.13	0.54***	0.13
Constant	-0.36***	0.01	-0.39***	0.10
Observations	222		222	

Note: * p < .05. **; p < .01. ***; p < .001.

In addition to these overall impacts across grade levels and schools, Table 7 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 the ITT outcomes suggest that the treatment effect may be most powerful grade 1 students, and the estimates for the TOT analyses suggest that the oldest group of students, from grade 4, benefited the most with corresponding increases in the number of books received.

Table 7. Treatment Effects by Grade Level, 1-4.

Intent-to-treat effects by grade level

Grade	Coefficients	SE	Effect size (SD)	p-value	N
1	0.20	0.27	0.21	0.48	26
2	0.08	0.12	0.09	0.50	36
3	0.06	0.09	0.06	0.48	60
4	0.13	0.09	0.14	0.13	92

Treatment-on-the-treated effects by grade level

Grade	Coefficients	SE	Effect size (SD)	p-value	N
1	0.02	0.02	0.02	0.29	26
2	0.01	0.01	0.01	0.44	36
3	0.01	0.01	0.01	0.43	60
4	0.04	0.02	0.04	0.14	92

Conclusion

The results of this quasi-experimental suggest that KRN can have statistically significant impacts on the reading achievement of students across grade 1 through 4. Further, when students and parents take full advantage of the program's offerings, the results for student achievement are amplified. Summer is often a time of stagnant learning gains or even important losses in the academic skills and knowledge that students accumulate over the school year. How should we consider and interpret the results of this study and what are the implications?

Though somewhat dated, the meta-analysis by Cooper and colleagues (1996) suggests that economically disadvantaged students, who represented the majority of our sample, lose, on average, over 2 months of learning during the summer. Using the well-established estimates of Bloom, Hill, Black, and Lipsey (2008), we can estimate the average amount of nine-month, school-year achievement growth in the reading domain across grades 1 through 4 to be approximately 0.64 standard deviation units. Given that the average impact of KRN over a single summer was 0.12 standard deviations, and applying the estimates of Bloom et al (2008), we find that the impact of KRN was equivalent to approximately 1.7 months of learning, or nearly 20% of the learning that takes place over a typical school year. This suggests that the impact of KRN may effectively compensate for nearly the total summer learning loss of 2 months that is typically experienced by low-income students.

Considering the full impacts of the program, when students and parents take advantage of the full complement of 9 books delivered by KRN, the results are more profound. Specifically, we find that the impact estimate for receiving all 9 books is equivalent to an effect size, or standardized gain, of 0.18. Again, applying the criteria from Bloom et al. (2008), this suggests that the full impact of KRN was equivalent to approximately 2.5 months of learning, or nearly 28% of the learning that takes place over a typical school year. When students and parents commit to the full KRN program, our results indicate that the impact of KRN can more than eradicate the entire 2 months of summer learning loss experienced by low-income students.

These results for KRN are highly promising and are quite consistent with the impacts noted by both Cooper, Charlton, Valentine, Muhlenbruck, and Borman (2000) and by Kim and Quinn (2013). Notably, Kim and Quinn found that the effective home-based summer reading programs that they reviewed produced average impacts on total reading outcomes of 0.13 standard deviation units. The KRN outcomes, ranging from 0.12 to 0.18 standard deviations compare favorably. Similarly, the average impact of 0.19 standard deviation units found for the typically more intensive and expensive school-based programs are essentially the same as those we found for full implementation of KRN.

These results are promising and suggest two key future directions for KRN. First, given the reliable impacts found across three independent school sites and given the replicable nature of the KRN program, this suggests a strong set of preconditions for continued scale up. Second, though these results meet rigorous evidence standards, such as those established by the What Works Clearinghouse, they also point toward the potential to conduct an even more powerful, larger demonstration of KRN's impact through a widely deployed randomized controlled trial.

References

aimswebPlus. (n.d.). aimswebPlus overview. Retrieved from Conclusion

Alexander, K. L., & Entwisle, D. R. (1996). Alexander, K.L., & Entwisle, D.R. (1996). Schools and children at risk. In A. Booth, & J.F. Dunn (Eds.). *Family-school links: How do they affect educational outcomes?* (pp. 67-89). Mahwah, NJ: Erlbaum.

Alexander, K. L., Entwisle, D.R., & Olson, L.S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*(2), 167–180.

Alexander, K., Pitcock, S., & Boulay, M. (Eds.) (2016). *The summer slide: What we know and can do about summer learning loss.* New York: Teachers College Press.

Allington, R. L., McGill-Franzen, A., Camilli, G., Willimas, L., Graff, J., Zeig, J., Zmach, C., Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*, 411-427.

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.

Battle Creek Public Schools. (n.d.). District information. Retrieved from https://www.battlecreekpublicschools.org/

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.

Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004) Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education, 77*, 1-31.

Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys, 22*(1), 31–72. https://doi-org.ezproxy.library.wisc.edu/10.1111/j.1467-6419.2007.00527.x

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. https://doi-org.ezproxy.library.wisc.edu/10.1037/a0018536

Cooper, H., Charlton, K., Valentine, J.C., Muhlenbruck, L. & Borman, G.D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*, 1-127.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research 66,* 227-268.

Downey, D. B., Von Hippel, P. T., & Broh, B. (2004). *Are schools the great equalizer? Cognitive inequality during the summer months and the school year*. American *Sociological Review, 69*, 613-635.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.

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

Heyns, B. (1978). *Summer learning and the effects of schooling.* New York, NY: Academic Press.

Ichimura, H., & Taber, C. (2001). Propensity-Score Matching with Instrumental Variables. *American Economic Review, 91*(2), 119–124. https://doi-org.ezproxy.library.wisc.edu/10.1257/aer.91.2.119

Kids Read Now. (n.d.) The Kids Read Now solution. Retrieved from https://kidsreadnow.org/about/

Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology, 99*, 505-515.

Kim, J. S. & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from Kindergarten to Grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*, 386-431.

Lee, W. S. (2013). Propensity score matching and variations on the balancing test. *Empirical Economics, 44*(1), 47. https://doi-org.ezproxy.library.wisc.edu/10.1007/s00181-011-0481-0

Linden, A. (2017). Improving causal inference with a doubly robust estimator that combines propensity score stratification and weighting. *Journal of Evaluation in Clinical Practice, 23*(4), 697–702. https://doi-org.ezproxy.library.wisc.edu/10.1111/jep.12714

Lunt, M. (2014). Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *American Journal Of Epidemiology, 179*(2), 226–235. https://doi-org.ezproxy.library.wisc.edu/10.1093/aje/kwt212

McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B., Lichter, D. S., & Cross, A. B. (2011). *Making summer count: How summer programs can boost children's learning.* Santa Monica, CA: RAND Corporation.

National Summer Learning Association. (2018). Spotlighting Kids Read Now. Retrieved from https://www.summerlearning.org/spotlighting-kids-read-now/

Niche. (n.d.-a). Battle Creek School District. Retrieved from https://www.niche.com/k12/d/battle-creek-public-schools-mi/

Niche. (n.d.-b). Hook Elementary School. Retrieved from https://www.niche.com/k12/hook-elementary-school-troy-oh/

Niche. (n.d.-c). Kyle Elementary School. Retrieved from https://www.niche.com/k12/kyle-elementary-school-troy-oh/

Niche. (n.d.-d). Troy City School District. Retrieved from https://www.niche.com/k12/d/troy-city-school-district-oh/

NWEA. (n.d.). The MAP Suite. Retrieved from https://www.nwea.org/the-map-suite/

Pirracchio, R., Carone, M., Rigon, M. R., Caruana, E., Mebazaa, A., & Chevret, S. (2016). Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. *Statistical Methods in Medical Research, 25(5)*, 1938. https://doi-org.ezproxy.library.wisc.edu/10.1177/0962280213507034

Public School Review. (n.d.). Verona Elementary School. Retrieved from https://www.publicschoolreview.com/verona-elementary-school-profile

Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool children's language and emergent literacy. *Journal of Early Childhood Literacy*, *10*, 97-117.

Richardson, J. T. E. (2011). Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research. *Educational Research Review, 6*(2), 135–147. Retrieved from http://search.ebscohost.com.ezproxy.library.wisc.edu/login.aspx?direct=true&AuthType=ip,uid&db=eric&AN=EJ927266&site=ehost-live&scope=site

Robins, J., Sued, M., Lei-Gomez, Q., & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when" inverse probability" weights are highly variable. *Statistical Science, 22*(4), 544-559.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3-4), 169-188.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1)*, 1.

Troy City Schools. (n.d.) About our district. Retrieved from http://www.troy.k12.oh.us/Content2/240

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals Of Translational Medicine, 7*(1), 16. https://doi-org.ezproxy.library.wisc.edu/10.21037/atm.2018.12.10