



Does Reading During the Summer Build Reading Skills? Evidence from a Randomized Experiment in 463 Classrooms

Jonathan Guryan

Associate Professor of Human Development and Social Policy
Faculty Fellow, Institute for Policy Research
Northwestern University

James Kim

Associate Professor of Education
Harvard University

David Quinn

Doctoral Student, Graduate School of Education
Harvard University

Version: December 22, 2014

DRAFT

Please do not quote or distribute without permission.

Abstract

There are large gaps in reading skills by family income among school-aged children in the United States. Correlational evidence suggests that reading skills are strongly related to the amount of reading students do outside of school. Experimental evidence testing whether this relationship is causal is lacking. The researchers report the results from a randomized evaluation of a summer reading program called Project READS, which induces students to read more during the summer by mailing 10 books to them, one per week. Simple intent-to-treat estimates show that the program increased reading during the summer, and show significant effects on reading comprehension test scores in the fall for third grade girls but not for third grade boys or second graders of either gender. Analyses that take advantage of within-classroom random assignment and cross-classroom variation in treatment effects show evidence that reading more books generates increases in reading comprehension skills, particularly when students read carefully enough to be able to answer basic questions about the books they read, and particularly for girls.

1. Introduction

According to the National Assessment of Educational Progress (NAEP), there are large gaps in reading performance by family income among school-aged children in the United States. In 2013, the gap in average NAEP reading scores between 4th-grade students who were and were not eligible for free or reduced-price lunch was about 0.8 of a standard deviation.¹ This gap has not narrowed significantly in the last decade, and income gaps in reading skills measured from other sources have grown substantially over the last half-century (Reardon, 2011).

Research suggests that the gap in reading skills by family income exists even before school entry, and appears to grow during the summer. Using data from the Early Childhood Longitudinal Study: Kindergarten Cohort (ECLS-K), a nationally representative sample of over 20,000 children entering kindergarten in 1998, Downey et al. (2004) found that gaps associated with socioeconomic status (a composite index of family income, parental education and parental occupational status) exist at kindergarten entry, and grow more quickly during the summer months than while school is in session. In addition, girls enjoyed a significant advantage in reading at school entry and also enjoyed larger gains during Kindergarten compared to boys.

Some have hypothesized that socioeconomic reading skills gaps grow more during the summer because family investments and resources matter more for skill development when school is not in session than when it is (Alexander, Entwisle and Olson, 2001). If school is a substitute for parental investments and low-income families invest less time and other resources in their children's development (Guryan et al., 2008), one might expect that children from these families would fall behind their higher-income peers during the summer. This would be particularly true for reading as compared with math if, as many experts believe, reading skill

¹ See *Digest of Education Statistics*, 2013, Table 221.75.

development is more dependent on activities outside of school, such as exposure to language (Hart and Risley, 1995) and time spent reading at home (Heyns, 1978; Fryer and Levitt, 2006), and if parents tend to view literacy development as the joint responsibility of parents and schools, but view math skill development as the responsibility primarily of schools (Evans et al., 2004).

Research suggests that a strong correlate of reading skills is the time a child spends reading outside of school. Heyns (1978) found that for a sample of 6th and 7th grade students, the number of books read outside of school and the number of hours they spent per day reading were the best predictors of reading scores. Strikingly, book reading explained a larger portion of the variance in reading skills than socioeconomic status. Hofferth and Sandberg (2001) evaluated time-use data for school-aged children and found that the activity that most strongly predicted achievement on the Woodcock-Johnson Achievement Test was time spent reading. Fryer and Levitt (2006) found that a one standard deviation increase in the number of children's books in the home was associated with a 0.115 standard deviation increase in reading scores, controlling for socioeconomic status and other background characteristics.²

As is surely clear, the strength and volume of the correlational research does not fully answer the causal question: Does additional time spent reading outside of school cause reading

² In a meta-analysis of 99 studies, Mol and Bus (2011) found positive associations between print exposure measures and verbal ability across the life span. To conduct the analysis, they found studies that measured reading volume with print exposure tasks, which are viewed as more objective and valid proxy measures of reading volume than self-reported surveys. Measures such as the Title Recognition Test typically include a list of titles of best-selling authors and foils (false or non-existing titles). In the elementary grades (1-6), Mol and Bus found stronger correlations between print exposure measures like the TRT and reading skills for children with lower ability levels ($r = .39$) than children with age-appropriate abilities ($r = .20$). Echols, West, Stanovich, and Zehr (1996) found that Grade 4 TRT scores predicted reading comprehension 2 years later after controlling for baseline measures of cognitive ability. These correlations, however, suffer from endogeneity bias since the volume of reading is not randomly assigned across children's environments.

skills to improve, and if so by how much? Families that have more books at home are certainly different in unobservable ways from families that have fewer books, and children who read more at home, whether for pleasure or because their parents strongly encourage them, are surely different from children who do not read much outside of school. To isolate the causal effect of out of school reading on reading skill development from the effect of other unobserved correlates of reading, we would ideally randomly assign some children to read more than others outside of school.

We implemented just such an experiment. We implemented a voluntary scaffolded summer reading program called Project READS (Reading Enhances Achievement During Summer) for 2nd and 3rd graders in 463 classrooms in 59 public schools in 7 North Carolina school districts in the spring and summer of 2013. Students randomly selected to be in the treatment group were given six reading comprehension lessons in the spring that focused on reading activities designed to foster children's engagement with books at home during the summer. Parents were also invited to an afterschool family literacy event where they learned about the READS activities. Treatment group students were then mailed 10 books, one per week, during the summer. The books were matched to students based on their baseline reading skill level and their interests. Students were encouraged to read the books, and were asked to mail a tri-fold after they read each book; the tri-fold included three comprehension questions about each book and a few questions designed to prompt the students to use the reading strategies taught during the spring lessons. Students assigned to the control group received no books during the summer, and participated in six mathematics lessons during the spring while the treatment students participated in the reading lessons.

The research design, which includes within classroom student- and teacher-level random assignment, presents an opportunity to evaluate the effect of randomly induced out of school reading on reading skill development. Both treatment and control group students were tested at baseline in the spring, but only treatment group students were mailed 10 books during the summer before post-testing in the fall. Because control children were given math lessons and books were mailed to treatment group children's homes, there was no diffusion of the treatment components across conditions. In addition, children's homeroom teachers, who provided instruction during the school year, were also randomly assigned to new classrooms for the intervention-related lessons at the end of the school. Thus, the design eliminates student and teacher confounds and enhances the internal validity of the intention-to-treat effects.

Our findings suggest the answer to the question of whether reading more books makes kids better readers is not a simple 'yes' or 'no.' On average, treatment group students read more books than control group students but ITT analyses show that the two groups had similar reading comprehension post-test scores, suggesting that simply reading more books does not in general cause an increase in reading comprehension skills. Breaking this result up by grade and gender, ITT estimates show no difference in reading comprehension scores for 2nd graders, consistent with the idea that 2nd graders do not in general have strong enough basic reading skills to benefit from an out of school reading program. The small and insignificant test-score effects among 2nd graders replicate previous non-significant results of READS with younger children in first- and second-grade (Kim, 2007) and with low-income language minority children with weak basic reading skills (Kim and Guryan, 2010). Based on the consistent finding that READS has not shown positive test score effects for 2nd graders, we conclude that READS is not developmentally appropriate for children of that age.

Among 3rd graders, however, we found significant reading comprehension test score effects for girls but not for boys, and the difference in effects by gender was statistically significant. These ITT test score estimates suggest that on average the intervention did not cause large increases in reading comprehension over the summer for third grade boys, but did for third grade girls. However, the pooled ITT estimates mask interesting variation in treatment effects across classrooms that may be informative about the underlying mechanism by which reading might build reading skills. Focusing on the 3rd graders, where there are treatment effects to learn from, we take advantage of the multi-site nature of the experimental design – there were essentially 218 different experiments in the 218 different classrooms – to investigate whether different types of reading might be differentially effective at building reading skills. Does reading more books translate into greater reading comprehension skill gains? Does the quantity of reading matter regardless of the child’s focus during the activity, or does reading only build skills when it is done in an engaged and focused way?

When we compare the experimental results across classrooms, we find that in third grade classrooms where the treatment induced students to read more books over the summer there were larger reading comprehension gains for girls. However, this result is stronger when summer reading is measured based on the number of tri-folds students returned indicating they had read the book, or when a student is only credited with reading a book if he or she answered basic questions about the book correctly, than when book reading is based on self-reports. This suggests that either behavioral measures of book reading are more accurate, or focused engaged reading builds transferrable reading comprehension skills while disengaged reading does not.

We also explore whether mediators we are able to measure explain the gender difference in treatment effects.³ We find that the difference in treatment effects by gender is not explained by the difference in the number of books Project READS induced girls to read relative to boys. Rather, we find that the effects of book reading were stronger for girls than for boys, which could have several different interpretations. It may be that reading during the summer builds reading skills for girls but not for boys. It is also possible that on average when 3rd grade girls read during the summer, they do so in a productive way whereas 3rd grade boys tend not to. This would be consistent with research showing that girls tend to have more positive attitudes toward recreational reading than boys (Logan and Johnston 2009) and that parents are more likely to read to girls than boys when they are younger (Bertrand and Pan 2013). For boys, but not for girls, there was some weak evidence that the amount they liked the books they read during the summer may have contributed to improved reading skills. It may be the case that when boys find books that they like, they read in a more engaged and focused way and build reading skills as a result. Taken together, these results suggest that, the amount of engaged, focus and careful reading, and not simply the number of books a child reads, may be what builds reading skills among beginning readers.

2. Project READS

Project READS (Reading Enhances Achievement During Summer) is a voluntary scaffolded summer reading program with two primary components.⁴ In the spring, just prior to

³ The results also add to the growing literature showing larger positive effects of social policy interventions for girls than boys (see e.g. Anderson 2008, Angrist, Lang and Oreopolous 2009, Angrist and Lavy 2009, Dynarski 2008, Kling, Leibman and Katz, 2007)

⁴ Scaffolding is a term used in the reading and literacy literature that refers to supports given to students as they learn to read that are meant to be substitutes for direct help from teachers,

the end of the school year, students in the program are taught six lessons (each lasting approximately one hour) during the school day. These lessons are focused around reading strategies that are designed to help beginning readers to read outside of school, with limited or no adult supports. Parents are also invited with their children to attend an afterschool family literacy event focused on the READS comprehension activities. Then during the summer, each student is mailed 10 books, one per week. Students are tested for reading comprehension in the spring and fall. The spring test serves as a baseline measure of reading skills, and the fall test serves as a post-test.

The books are chosen to match each student's reading skill level and reading interests as best as possible. Reading skill levels are measured using the spring baseline reading comprehension test, which are translated into Lexiles, a proprietary system designed to align reading skills with the difficulty of children's books. Students are also asked questions about the types of books they would like to read. Using an algorithm, books are then chosen that best match each student's interest among those that are the appropriate difficulty given the student's baseline reading skill level.

Previous studies of Project READS have shown evidence of its effectiveness at improving reading comprehension and highlighted for whom and under what conditions READS works best. Kim (2006) describes the results of a randomized controlled trial in which students assigned to Project READS had higher reading comprehension test scores than control students, particularly for African-American and Hispanic students. Kim and White (2008) found that students assigned to a version of Project READS where students were prompted during the summer to use the reading strategies taught during the spring lessons had larger test score gains

parents or peers. In the case of READS, scaffolding refers to specific reading strategies and tactics that the child was taught to use when reading on his or her own.

than students who were assigned to receive books mailed to them over the summer with no prompting to use the reading strategies. In addition, prior experimental studies of READS experiments found no positive effects on reading comprehension for children in first- and second-grade or for low-income Latino children. None of the prior studies of READS included measures of book reading that were correlated with reading skills. An important contribution of this study is the creation of a book-specific tri-fold, which was mailed with each READS book, and which asked children comprehension questions about each of their 10 mailed books. Children were asked to return the tri-fold when they had completed reading a READS book. The tri-folds provide a measure of summer reading that is based on behavior rather than self-report, and the comprehension questions on the tri-folds also provide an indication of how carefully the student read the books.

The current study is on a much larger scale than the previous studies, and addresses the mechanism by which reading might build reading skills. We use data from the tri-folds to shed light on the mechanisms that may underlie improvement in reading comprehension. Finally, whereas the previous studies took place in a single district with 400-550 students, the current study took place in 59 North Carolina elementary schools in 7 districts and included 6,383 students (Grade 2 and 3 at baseline).

3. Estimation

We present two types of analyses. The first are standard intent to treat (ITT) estimates of the effect of Project READS. We estimated the following model

$$(1) \quad y_i^{Fall} = \alpha_0 + \alpha_1 R_i + \alpha_y y_i^{Spring} + \gamma_c + \varepsilon_i$$

where y_i^{Fall} is a post-random assignment outcome measured in the fall, R_i is an indicator for being randomly assigned to Project READS, y_i^{Spring} is a baseline reading comprehension test score, γ_c is a set of classroom fixed effects, α_0 , α_1 and α_y are parameters to be estimated, and ε_i is a random error term. The parameter of interest, α_1 , measures the difference in average outcomes between students randomly assigned to Project READS and students in the control group. We refer to it as the intent to treat (ITT) estimate because R_i is the actual assignment rather than a measure of participation. However, since all students in the study were consented prior to random assignment, virtually every student who was randomly assigned to receive the Project READS program did so. Thus, in practice our estimates of α_1 are very close to measures of the effect of participation, if participation is defined as receiving the spring lessons in the classroom, an invitation to an afterschool family literacy event and being mailed 10 books over the summer.

Assignment to Project READS was selected randomly within classroom. Random assignment ensures that in large samples R_i is uncorrelated with ε_i (or equivalently with potential outcomes for y_i^{Fall}), and as a result we can interpret estimates of α_1 as the causal effect of assignment to Project READS on y_i^{Fall} . Furthermore, within grade and school, classroom teachers were randomly assigned to teach either six READS lessons or six math lessons at the end of the school year. Students who had been randomly assigned to treatment were then randomly assigned to one of the teachers teaching READS lessons for their end-of-year lessons, while control students were randomly assigned to one of the teachers teaching math lessons. The quality of classroom teachers treatment and control students were exposed to for the prior school year is balanced mechanically because random assignment was within classroom; random assignment of teachers for the end-of-year lessons promoted balance in the quality of teachers

student were exposed to for those six lessons. We ran balance tests which show that teachers assigned to teach the math and READS lessons were balanced on race, gender, and the selectivity of the college they attended as an undergraduate; teachers assigned to teach the math classes were older on average (2.4 years, $p=0.04$).

Because random assignment was conditional on classroom, we include classroom fixed effects in all models.⁵ Inclusion of the baseline reading test as a control is not necessary for internal validity, but it improves precision because it explains a large portion of the residual variance.

The second type of model we estimate takes advantage of the fact that there were essentially 218 different experiments, one in each third grade spring classroom. It would be possible to estimate 218 treatment effects, each based on random assignment within a specific classroom. Presumably, students responded to the treatment in different ways in different classrooms. Students may have been induced to read more by Project READS in some classrooms than others, for example. It would be interesting to know if in the classrooms where students were induced to read more by the treatment, there were also larger treatment effects on test scores. The logic of this analysis follows directly from Kling, Liebman and Katz (2007). We

⁵ In practice, the probability of being assigned to the treatment group was close to 0.5 in each classroom. Since the treatment assignment probability was not correlated with classroom, excluding the classroom fixed effects should not cause any omitted variables bias, and are therefore not necessary. Within each classroom, randomization was also stratified by whether students were limited English proficient (LEP) and whether they were eligible for free or reduced lunch (FRL). Within each classroom, the probability of assignment to the treatment was the same in the FRL/LEP and non-FRL/LEP strata, so in practice this level of stratification is ignorable. We report results from models that include classroom fixed effects because the subsequent analysis is at the classroom level, and models that include classroom, classroom-by-LEP/FRL status or no fixed effects are all unbiased. The results from all three sets of models yield the same qualitative findings.

use classroom-by-treatment fixed effects as instruments for potential mediators of the relationship between assignment to Project READS and test scores.

The first stage model is

$$(2) \quad M_i = \pi_c R_i + \pi_y y_i^{Spring} + \delta_c + u_i$$

where M_i is the potential mediator (e.g. number of books read during the summer), π_c is a full set of classroom-level treatment fixed effects, and u_i is an error term. The second stage model is

$$(3) \quad y_i^{Fall} = \beta_0 + \beta_1 M_i + \beta_y y_i^{Spring} + \lambda_c + e_i$$

Classroom fixed effects are included in both the first and second stage models so that the variation identifying the effect of the mediator on fall test scores is treatment-control comparisons within classrooms. The identification of these mediator parameters are vulnerable to the possibility that treatment effects on the mediators are correlated with an unobservable cause of the test score treatment effects, what might be referred to as “omitted mediator bias.”⁶

However, relative to the common alternative of adding controls for potential mediators to the ITT test score model, the assumptions of the multi-site IV method are much weaker. Whereas the commonly used approach is invalid if the level of unobservable determinants of test scores are correlated with treatment effects across sites, the multi-site IV only requires that there are no treatment-control differences in unobservables that are correlated with test score treatment effects across sites. Since treatment is randomly assigned at each site (in our case classrooms), baseline unobservables should be balanced across treatment and control observations within sites. Thus, the identifying assumption is that variation in treatment effects on unobserved *mediators* are not correlated with treatment effects on the included mediator.

⁶ We thank Jens Ludwig for suggesting this term.

4. Data and descriptive statistics

Table 1 presents descriptive statistics and baseline equivalence tests. The left panel shows means and equivalence tests for the full study sample, while the right panel shows means for the analysis sample, defined as observations that have non-missing baseline and post-test scores. The majority of missing test score data is from the post-test, but the similarity of spring standardized ITBS scores in the analysis sample and the full sample among the control group provides comfort that restricting the analysis to observations with non-missing pre- and post-test data is unlikely to bias the results. For observations with non-missing fall ITBS scores, we imputed missing values for other variables using the mean value within the same classroom and experimental group (in cases where all students in a classroom/treatment cell were missing on the variable in question, we imputed the grade level/treatment group-specific mean). All regressions include as controls indicators for observations with imputed values. Imputation allows us to analyze a consistent sample across the different analyses we perform. Analyses not reported here show qualitatively similar results when we do not impute missing values and allow the samples to vary across analyses.

4.1. Demographics of the sample

Fifty-three percent of the students in the study were female, approximately 77 percent of the sample were eligible for free or reduced-price lunch, a common measure of poverty in education data, and seventeen percent of students in the study were classified as limited English

proficient. Twenty-three percent of the sample was white, 38 percent were black, 22 percent were Hispanic and 17 percent were a race/ethnicity other than white, black or Hispanic.⁷

4.2. Measures of reading comprehension

A key outcome in the study was the reading comprehension section of the Iowa Test of Basic Skills, which was administered in both the spring (baseline) and fall (post-test) specifically for this study. ITBS scaled scores were normalized to have mean zero and standard deviation one in each grade in the control group for the full sample. The table also reports means of three measures of summer book reading. Students in the treatment group were mailed a tri-fold (a piece of paper folded into thirds) with each book. The tri-folds asked the student to respond to a series of questions aimed at reminding the students to use the reading strategies they had been taught during the spring lessons. Students were also instructed to mail the tri-fold back, addressed and postage paid, when they had read the book. The number of tri-folds returned is one measure of the number of Project READS books the students read during the summer. On average, treatment students returned slightly fewer than 4 of the 10 possible tri-folds. There were a very small number of control students who returned tri-folds. These were students who shared living arrangements with a treatment student, either because they were siblings or because they lived in the same household.

4.3. Measures of summer reading

Students were also asked to report the number of books they read over the summer in a survey administered in the fall after they returned from summer vacation.⁸ This question was

⁷ In the data students are placed into one of the following race/ethnicity categories: White, Black, Hispanic, Asian, Native Hawaiian/Pacific Islander, American Indian and Multi-Racial.

asked of all students, regardless of whether they were in the treatment or control group, and asked about all books read, not just books that were mailed as a part of Project READS. Students in the control group reported reading 9.8 books during the summer on average, while students in the treatment group reported reading 10.9 books. The difference was statistically significant, suggesting that assignment to participate in Project READS induced students to read an additional 1.1 books over the summer.

On the tri-folds, students in the treatment group were asked three comprehension questions about the book they had been mailed. The number of questions answered correctly on returned tri-folds is a measure of book reading that requires students to have read the book carefully enough to have been able to answer basic questions about the book. The average treatment student answered 8 total questions correctly. Since students on average returned 3.75 tri-folds, this represents approximately 70 percent of the total possible questions on the tri-folds students returned.

4.4. Test of baseline equivalence between treatment and control groups

Random assignment should make the treatment and control groups comparable on average. If random assignment was carried out correctly, both unobservables and observables should be balanced across the treatment and control groups. This is of course what allows us to draw strong causal inferences; any average difference in post random assignment outcomes must be driven by assignment to the treatment since other determinants of the outcome should be the

⁸ The number of books read variable is derived from the fall student survey, on which students from both conditions were asked to respond to the question “During summer vacation, about how many books did you read?” with answer choices “0-1,” “2-3,” up to “20 or more.” We assigned students the mid-point value for their answer choice (e.g. students who reported having read 2-3 books were given a value of 2.5), and students who selected “20 or more” were given a value of 20.

same on average in the treatment and control group. It is impossible to test for baseline equivalence of unobservables, but it is possible to test for baseline equivalence of observables. Table 1 presents means of several variables measured at baseline, separately for the treatment and control groups, along with p-values from t-tests of the difference in means. None of the differences in averages at baseline is large, and the p-values of the pairwise t-tests are all larger than 0.3. Random assignment appears to have generated balance in observables across the treatment and control groups.

5. Intent to treat estimates

Tables 2-4 present treatment-control differences in book reading during the summer, how much students liked the books they read over the summer, and ITBS reading comprehension test scores in the fall after the intervention. Students were surveyed in the fall after returning from summer vacation and asked how many books they read during the summer. Table 2 presents estimates from a regression of self-reported summer book reading on an indicator for being assigned to the treatment condition. Table 3 shows results from similar regressions with a measure of how much the students reported liking the books they read during the summer, and table 4 shows results from similar models with fall ITBS reading comprehension test scores as the dependent variable. Each cell in the tables shows the coefficient on a treatment dummy from a different regression. The first row shows estimates for grades 2 & 3 combined, while the next two rows show estimates for grades 2 and 3 separately. The first column shows estimates for boys and girls combined, while the second and third columns show estimates for boy and girls separately. Because random assignment was conditional on spring classroom, all regressions control for spring classroom fixed effects.

The consent process occurred prior to random assignment, so only consented students were eligible to be randomly assigned. All students randomly assigned to the treatment condition participated in the spring reading lessons and were in fact mailed 10 books during the summer. Thus, the estimates shown in tables 2-4 can be thought of as intent to treat (ITT) estimates of the effect of being assigned to Project READS. However, since the take-up rate was 100 percent among those assigned to treatment and there was very little control crossover, the estimates in tables 2 and 4 are very close to estimates of the effect of participation in Project READS on summer reading and reading comprehension.

On average, control group students reported reading 9.8 books during the summer. Control group third graders reported reading fewer books than second graders. Third graders reported reading 9.2 books; second graders reported reading 10.3 books.⁹ From that baseline, mailing books to students during the summer caused them to increase their summer reading activity by about 1.1 books on average. The increase was statistically significant, and treatment effects tended to be slightly larger for boys than girls. Second and third grade boys in Project READS read about 1.3 more books than their control group counterparts. Second and third grade girls read approximately 0.8 more books than their control group counterparts.

The treatment effect of Project READS on book reading was slightly larger for third graders than for second graders, though this difference was not statistically significant. The treatment effect for second graders was an additional 1.0 books, while the treatment effect for third graders was an additional 1.2 books. There was not a statistically significant effect of Project READS on book reading for second grade girls (the point estimate was 0.560 books, with a standard error of 0.380), whereas the treatment effect for second grade boys was an additional

⁹ Second graders and third graders typically read different types of books, so the amount of text read by third graders was likely greater than the amount of text read by second graders.

1.4 books (standard error of 0.411). The gender difference in the treatment effect on book reading among second graders was marginally significant. There was no corresponding gender difference in the effect of READS on summer reading for third graders. READS increased summer reading by 1.2 books for both boys and girls in third grade.

Project READS did not just induce children to read more books during the summer. It also may have caused children to read different books than they otherwise would have. Books are matched to students based both on their baseline reading skills and on their interests. Students responded to survey questions in the spring designed to elicit their reading tastes. To match students to books of the appropriate reading difficulty, baseline ITBS scores were translated into Lexiles, a proprietary system in which student reading skills and books are rated on a comparable scale. An algorithm was then used to choose 10 books for each student that most closely matched the student's reading interests among the books that were within the recommended range on the Lexile scale. By mailing particular books to students, Project READS may have induced some students to read those books instead of books they would have read in the absence of the program. The books may be better matched based on the Lexile measure of difficulty, but they may be matched less well in other dimensions. Table 3 shows how Project READS affected how much students reported liking the books they read during the summer. The estimates show that treatment students reported liking the books they read significantly less than control students. This effect was fairly stable across grade and gender, with estimates ranging from approximately 0.15 to 0.22 standard deviation declines.

Table 4 shows the estimated ITT effects of Project READS on ITBS reading comprehension scores. The pooled second and third grade sample showed no treatment control difference in reading comprehension scores. The point estimate on the standardized test score

(standardized to have mean zero and standard deviation 1 in the control group) was 0.014 with a standard error of 0.017. The estimate for second graders was also small and statistically insignificant. The point estimate for second graders was 0.010 with a standard error of 0.023. The estimated effect for second grade boys was 0.027 with a standard error of 0.034, and for second grade girls was -0.004 with a standard error of 0.034. The lack of an effect on reading comprehension skills for second grader was perhaps not surprising since Project READS was developed for 3rd-5th graders. As a voluntary scaffolded summer reading program, Project READS requires students to read at home with limited or no adult support. Second graders are perhaps not skilled enough readers yet to benefit from reading on their own. There also was not a statistically significant effect on summer reading for second grade girls, though the confidence interval for that estimate is too wide to rule out reasonably large increases in summer reading.

The third grade sample showed a small and insignificant reading comprehension effect when boys and girls are pooled, but there was a notable gender difference. For the pooled sample of boys and girls in the third grade, the treatment effect on test scores was 0.020 with a standard error of 0.024. For third grade boys, there was a negative and insignificant reading score effect of -0.024 (standard error of 0.036). For third grade girls, there was a positive treatment effect of 0.073 with a standard error of 0.034. The reading comprehension effect for third grade girls was statistically significantly different from zero, and also statistically significantly different from the third grade boy effect. The reading comprehension effect for third grade girls was fairly large in magnitude. For that group, the treatment caused an increase in reading of about 1.2 books during the summer and an increase of 7.3 percent of a standard deviation in reading comprehension skills relative to the control group. One way to think about the magnitude of the effect is in months of learning. Measured this way, the treatment effect for third grade girls is equal to the

test score gain that the average student experiences in 1.4 months between 2nd and 4th grade.¹⁰

Put a different way, in 2013 the difference in NAEP reading scores between free lunch eligible and free lunch ineligible students was 0.79 standard deviations; the ITT effect for third grade girls was 9.2 percent as large as the poverty gap in reading achievement.¹¹

Based on these results, there is not a clear answer to the question that motivated this study: whether reading books outside of school improves reading skills. Second graders may not have the necessary skills to benefit from the type of reading activity that Project READS encourages, a finding that is consistent with past studies of READS (see e.g. Kim 2007). Among third graders, boys appeared not to benefit on average from additional reading while girls benefited significantly from a modest increase in reading activity. Why is there this gender difference in the degree to which reading builds reading skills? Are there particular types of reading that are more effective at building reading skills for beginning readers? Perhaps exploring these differences can help us to understand how reading skills are developed.

6. Are there more and less effective ways to read?

To delve deeper into the mechanisms by which reading builds reading skills, we focus on third graders where there were treatment effects on test scores to try to understand, and we take advantage of the fact that random assignment occurred within classrooms. In a sense, this means that there were really 218 different third grade experiments, one in each third grade classroom.

This allows us to ask whether the treatment effects on reading comprehension scores were larger

¹⁰ Measured in scaled test score points, the treatment effect for 3rd grade girls was 1.9 points. The average annual growth in scaled scores in the ITBS for grade 2 to 3 and grade 3 to 4 is 16.0 points, or approximately 1.33 points per month.

¹¹ See *Digest of Education Statistics*, 2013, Table 221.75.
http://nces.ed.gov/programs/digest/d13/tables/dt13_221.75.asp

in the classrooms where treatment students reacted to the treatment differently. Were test score effects larger in classrooms where students read more books? Did students build more general reading comprehension skills in classrooms where students exhibited better comprehension of the specific books they read?

These are essentially questions about what the mediators are for the treatment effect on test scores. To identify the effect of potential mediators on test scores, following Kling, Liebman and Katz (2007), we estimated a series of instrumental variables (IV) models using classroom-by-treatment assignment dummies as instruments for the potential mediators. This IV model takes advantage of the random assignment – the instruments are the randomly assigned treatment dummy interacted with pre-random-assignment classroom – and essentially asks whether test score effects were significantly larger in classrooms where treatment assignment caused the potential mediator to increase more relative to the control group.

For the mediator analysis, we focus on third graders because we believe the intervention was not properly designed for second graders and because there were no apparent test score effects for boys or girls in second grade. As described above, Project READS has been implemented for a number of years in multiple locations, and prior to this study it has only shown test score effects when it focused on third through fifth graders. Originally, the current study was intended for third and fourth graders who were to receive the intervention for two consecutive summers (grade three to grade four, and then grade four to grade five), consistent with the sampling design of grade three through five students in earlier studies (Kim and White, 2008). Fourth graders were replaced with second graders just prior to implementation because of a change in testing that would have made it impossible to obtain posttest data for fourth graders in this cohort.

6.1. Does reading more books during the summer improve reading comprehension skills?

Table 5 presents the estimates of mediator effects from these IV models. Each row of the table shows results for a different potential mediator. While it is clearly not possible to test for every potential mediator, we focused on measures of summer reading activity, the potential mediator around which READS was designed. We present estimates for several mediators, each of which measures summer reading behavior in a slightly different way. For each mediator, we show results for the pooled sample of boy and girl third graders in the first column, followed by estimates for third grade boys and third grade girls. The first two rows show estimates in which self-reported summer book reading is the potential mediator. Students were surveyed in the fall after returning from summer vacation and asked how many books they read over the summer. A positive effect would indicate that classrooms where mailing books caused a larger increase in summer book reading also had larger treatment effects on reading comprehension skills. The first row presents results from a specification in which the self-reported number of books is the potential mediator; the second row shows results for a median split of self-reported book reading, or a binary indicator for having read more than 8.5 books. In both summer-book-reading specifications, the point estimates are positive but not significantly different from zero. The point estimates are moderate in magnitude. The estimate for the pooled third grade sample is 0.007, which would mean that in a classroom where mailing 10 books home caused an increase in reading of 5 books relative to the control group, the treatment effect would have been 3.5 percent of a standard deviation. Given the size of the standard errors, based on the estimates in rows 1 and 2 of table 5, we cannot rule out fairly large effects or no effect at all of inducing students to read during the summer.

Self-reported reading measures may not accurately measure actual reading activities. One nice feature of the research design is that we have alternative measures of reading behavior. Every student in the treatment group was asked to return a tri-fold every time they read one of the books they had received in the mail. While this not a direct measure of reading, it has the feature that it is based on behavior (the student has to mail the tri-fold) rather than a report of behavior. Rows 3 and 4 show the results where the number of tri-folds returned is treated as the potential mediator. A linear count of tri-fold returns is used in row 3, and median split indicator for having returned more than 2 tri-folds is used in row 4. Virtually all control students returned zero tri-folds, so the first stage estimates are essentially the average number of tri-folds returned by the treatment students in the classroom.¹² The analysis therefore asks whether treatment effects on test scores were higher in classrooms where the treatment students read more of the books they were mailed during the summer.

The results suggest that returning more tri-folds, and presumably reading more of the Project READS books, was associated with a larger test score effect for girls and not for boys. The point estimates for boys were small and insignificant, while the point estimates for girls were reasonably large and statistically significant. In the specification with the linear measure of tri-fold returns, the point estimate for girls was 0.014, which indicates that in a classroom where the students read all 10 books that were mailed home, the girls in the program scored 0.14 standard deviations higher than the control group on the reading comprehension post-test. In other words, reading the Project READS books improved reading comprehension skills for girls

¹² There were a very small number of control students who lived in the same household as treatment students and who returned tri-folds that were mailed to the treatment student with whom they were living. This control group crossover would presumably tend to reduce any measured treatment effects since some control students received the treatment. Since this represents a very small portion of the sample, we suspect any understatement of treatment effects is small.

in classrooms where the students participated in the program as intended. It is possible that there is something else about the classrooms where the treatment students read a lot of the mailed books that explains the high treatment effects on test scores. However, it is interesting that the treatment effects are not large for boys in those classrooms. This gender pattern suggests there may be something about the ways that boys and girls read that makes summer reading more effective at building reading skills for girls than boys. It is possible, for example, that on average third grade girls are more diligent and engaged than boys when they read in unsupervised settings.

6.2. Does reading carefully over the summer improve reading comprehension skills?

Does the extent to which students are engaged when they read matter for how effectively reading builds reading skills? Do students who are focused enough when they read books over the summer that they can answer questions on basic facts about the book gain more general reading comprehension skills than students who did not retain an understanding of the specific books they read over the summer? We address these questions with the analyses presented in the fifth row of table 5. On each tri-fold, students were asked to answer reading comprehension questions about the book they had been mailed. In the IV models shown in row 5, the potential mediator is the total number of items the student answered correctly on those tri-fold reading comprehension questions. The analysis asks whether in classrooms where the students read carefully enough to be able to correctly answer basic reading comprehension questions about the books they read during the summer, the treatment effect on general reading comprehension skills was larger. In other words, does reading carefully enough to understand a particular book generate reading comprehension skills that carry over to reading other materials?

The estimates are shown in row 5 of table 5. For boys, the point estimate was zero, while for girls there was a marginally significant positive effect of 0.006 with a standard error of 0.003 ($p=0.052$). The estimate suggests that in classrooms where students were focused and engaged enough when they read over the summer that they were able to answer comprehension questions about the books, girls gained reading comprehension skills that carried over to other reading activities. This suggests that encouraging children to read during the summer may not be enough; it may be important for them to read in an active and engaged way to build reading skills.

6.3. Does liking the books you read matter for building reading skills?

Why does reading in this way not appear to improve reading skills for boys in the way it does for girls? One clue might be found in the results shown in row 6 of table 4. In those models, the potential mediator is the degree to which students reported liking the books they read over the summer.¹³ For the pooled sample of boys and girls, the estimate of the effect of reading books students like was not statistically distinguishable from zero. The estimate for girls was close to zero and insignificant as well. For boys, the point estimate was positive but not statistically different from zero. Though it was not significant, the point estimate for boys was reasonably large in magnitude, and the contrast with the estimate for girls is interesting. Care should be taken when interpreting coefficients that are not statistically significant, but the results are weakly suggestive that liking the books they read may be important for boys to build reading skills. Again, we emphasize that this estimate is not statistically significant; this result may be worthy of further inquiry.

¹³ The survey question was, “Overall, how much did you like or dislike the books you read this summer?” The choices were: “I loved them,” “I liked them,” “They were okay,” and “I didn’t like them.” Scores of 4-1 were assigned to these answers, and that variable was normalized to have mean zero and standard deviation one.

7. Conclusion

Gaps in reading skills by family income are present early in childhood and appear to grow over the summer when children are not in school. Numerous studies document a strong correlation between reading skills and the amount of reading students do outside of school. It is possible that differences in out of school reading explain the income-based gaps in reading skills, and that policies aimed at increasing the amount of reading that children do outside of school could improve reading skills generally and narrow gaps. On its own, however, the correlational evidence does not establish a causal relationship, and may not be informative about policy.

In this paper, we report the results from a randomized evaluation of a summer reading program called Project READS, which induces students to read more during the summer by mailing ten books to them, one per week. Comparisons of students randomly assigned to Project READS with control students showed that the program increased reading during the summer by slightly more than one book, relative to baseline reading levels of approximately 10 books. Similar comparisons showed significant effects on reading comprehension test scores in the fall for third grade girls, but not for third grade boys or second graders of either gender. The effects on reading comprehension achievement test scores for third grade girls were large in magnitude relative to the intensiveness of the intervention; assuming the standard deviation of ITBS scores in our control group is comparable to the standard deviation of NAEP reading scores, the treatment effect of 0.073 standard deviations is approximately 9.2 percent of the free-lunch eligible/ineligible gap in NAEP reading scores, and is equal to the gain in test scores a typical third grader experiences in 1.4 months.

We also used the multi-site nature of the experiment – randomization took place in 218 separate classrooms – to explore the causal relationship between reading and the development of reading skills. Analyses that used classroom-by-treatment status indicators as instruments for the amount and kind of reading students did showed that reading more books generated increases in reading comprehension skills, particularly when students read carefully enough to be able to answer basic questions about the books they read, and particularly for girls. We find that for third grade girls reading an additional book during the summer increases reading comprehension by 1 to 1.4 percent of a standard deviation, depending on our measure of book reading. To put these estimates into context, this would imply that reading an additional five books over the summer would generate reading score gains equal to 6 to 9 percent of the poverty gap in reading achievement scores in the NAEP.

Another way to measure the magnitude of the treatment effects is relative to costs. Compared with many interventions in education, READS is moderate in per-student costs. We estimate that that cost to provide the Project READS intervention described in this study was between \$250 and \$400 per student. These costs include the cost of the books, the mailings, stipends to teachers, salaries of staff who managed the implementation, and testing. Relative to costs of this magnitude, the treatment effect for third grade girls of 0.073 SD compares favorably with effect sizes from the Tennessee STAR class size experiment (Krueger 1999, Schanzenbach 2006) and Perry Preschool (Schweinhart et al. 2005). For the third grade girls subgroup, the test score effect per \$1,000 of per-student cost ranges from 0.18 to 0.29. For comparison, the test score effect per \$1,000 of per-student cost was approximately 0.01 for class size reduction in the Tennessee STAR study, and approximately 0.03 for Perry Preschool.

Of course, cost benefit ratios for the treatment effects for other subgroups that we estimate to be statistically insignificant and close to zero clearly cannot be distinguished from zero. The size of our study does not allow for sufficiently powerful tests of the hypothesis that the benefit-cost ratio is greater than 1 for other subgroups, or for the full sample of third graders, based on the intent to treat analyses.

The instrumental variables estimates provide an alternative way to estimate cost-benefit ratios for the average student in the study. On average, third grade students randomly assigned to Project READS read 1.3 more books over the summer than control students. Using the per-student costs described above, the intervention increased book reading by 3.1 to 5.0 books per \$1,000 spent. As just mentioned, the instrumental variables estimates in table 5 imply that reading an additional book during the summer generates an increase in reading scores between 0.010 and 0.014 standard deviations. This would imply a benefit-cost ratio of approximately 0.03 to 0.07 for the average third grader in the study, which is substantially smaller than the benefit-cost ratios we estimate for the third grade girl subgroup, but which compares favorably with class-size reduction and Perry Preschool.

Taken together, the results of this study suggest that the answer to the question posed by the title is not a simple “yes” or “no.” Going through the motions of reading without being focused enough to remember basic facts about the book may not be effective at building lasting reading skills. Reading carefully and with enough focus to be able to answer reading comprehension questions about the book appears to build reading skills that improve comprehension of other texts weeks or months later. How best to get children learning to read in an engaged and focused way remains an open question, and would be a promising area for future research.

References

- Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson. 2001. "Schools, Achievement, and Inequality: A Seasonal Perspective," *Educational Evaluation and Policy Analysis* 23(2):171-191.
- Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103(484):1481-1495.
- Angrist, Joshua, Daniel Lang and Victor Lavy. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics* 1(1):136-163.
- Angrist, Joshua D. and Victor Lavy. 2009. "The Effect of High School Matriculation Awards: Evidence from Randomized Trials," *American Economic Review* 99(4):1384-1414.
- Downey, Douglas B., Paul T. von Hippel, and Beckett A. Broh. 2004. "Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year," *American Sociological Review* 69:613-635.
- Dynarski, Susan. 2008. "Building the Stock of College-Educated Labor," *Journal of Human Resources* 63(3):576-610.
- Echols, Laura D., Richard F. West, Keith E. Stanovich and Kathleen S. Zehr. 1996. "Using Children's Literacy Activities to Predict Growth in Verbal Cognitive Skills: A Longitudinal Investigation," *Journal of Educational Psychology* 88(2):296-304.
- Fryer, Roland G. Jr., and Steven D. Levitt. 2006. "The Black-White Test Score Gap Through Third Grade," *American Law and Economic Review* 8(2):249-281.
- Guryan, Jonathan, Erik Hurst and Melissa Kearney. 2008. "Parental Education and Parental Time with Children," *Journal of Economic Perspectives* 22(3):23-46.
- Hart, Betty and Todd R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H. Brookes Publishing Co.
- Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling*. Academic Press.
- Hofferth, Sandra L. and John F. Sandberg. 2001. "How American Children Spend Their Time," *Journal of Marriage and Family* 63:295-308.
- Kim, James S. 2006. "Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results from a Randomized Field Trial," *Educational Evaluation and Policy Analysis* 28(4): 335-355.

- Kim, James S. and Thomas G. White. 2008. "Scaffolding Voluntary Summer Reading for Children in Grades 3 to 5: An Experimental Study," *Scientific Studies of Reading* 12(1): 1-23.
- Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects," *Econometrica* 75(1): 83-119.
- Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics* 114(2):497-532.
- Mol, Suzanne E., and Adrianna G. Bus. 2011. "To Read or Not to Read: A Meta-Analysis of Print Exposure From Infancy to Early Adulthood," *Psychological Bulletin* 137(2):267-296.
- Reardon, Sean. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations," (pp. 91-116) in Greg J. Duncan and Richard J. Murnane, Eds. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage Foundation.
- Schanzenbach, Diane W. 2006. "What Have Researchers Learned from Project STAR?" *Brookings Papers on Education Policy*. 9:205-228.
- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, Michigan: High/Scope Press.

**Table 1:
Descriptive Statistics for Treatment and Control Students by Grade**

	Full Sample				Analysis Sample			
	Grades 2 & 3							
	Control	READS	<i>p</i> -value	<i>n</i>	Control	READS	<i>p</i> -value	<i>n</i>
% Female	51.79	50.54	0.34	6383	52.97	51.18	0.97	5319
% FRL	77.63	77.39	0.68	6383	77.04	76.58	0.34	5319
% LEP	16.62	16.31	0.89	6383	17.14	16.74	0.37	5319
% Hispanic	21.37	21.25	1.00	6383	21.81	21.86	0.75	5319
% Black	40.05	39.31	0.47	6383	38.24	37.87	0.61	5319
% White	22.60	23.10	0.63	6383	23.37	23.68	0.42	5319
% Other Race	15.98	16.35	0.66	6383	16.57	16.59	0.51	5319
Spr. Comp. Z-Score	0.00	0.01	0.67	6088	0.03	0.03	0.94	5319
Spr. Comp Std. Score	174.31	174.77	0.65	6088	175.05	175.23	0.94	5319
Fall Comp. Z-Score	0.00	0.00	0.64	5518	0.01	0.02	0.95	5319
Fall Comp Std. Score	176.92	177.18	0.58	5518	177.13	177.58	0.95	5319
# Trifolds Returned	0.04	3.62	0.00	6383	0.04	3.85	0.00	5319
# Trifold Qs Correct	0.07	7.50	0.00	6383	0.08	7.99	0.00	5319
# Books Read	9.82	10.82	0.00	6383	9.81	10.88	0.00	5319
Enjoyment of Books (std.)	0.11	-0.10	0.00	6383	0.11	-0.10	0.00	5319
	Grade 2							
% Female	50.43	50.18	0.97	3433	51.83	51.11	0.97	2819
% FRL	77.96	76.78	0.34	3433	77.36	75.62	0.34	2819
% LEP	17.48	16.20	0.37	3433	17.74	16.68	0.37	2819
% Hispanic	22.56	22.90	0.75	3433	22.90	23.75	0.75	2819
% Black	39.67	38.81	0.61	3433	37.88	37.23	0.61	2819
% White	22.02	23.15	0.42	3433	22.90	23.60	0.42	2819
% Other Race	15.75	15.14	0.51	3433	16.32	15.41	0.51	2819
Spr. Comp. Z-Score	0.00	-0.01	0.94	3255	0.04	0.01	0.94	2819
Spr. Comp Std. Score	167.60	167.46	0.94	3255	168.38	167.77	0.94	2819
Fall Comp. Z-Score	0.00	-0.02	0.95	2938	0.01	-0.01	0.95	2819
Fall Comp	169.61	169.18	0.95	2938	169.84	169.39	0.95	2819

Std. Score									
# Trifolds Returned	0.03	3.74	0.00	3433	0.04	3.94	0.00	2819	
# Trifold Qs Correct	0.05	7.40	0.00	3433	0.06	7.80	0.00	2819	
# Books Read	10.31	11.22	0.00	3433	10.32	11.26	0.00	2819	
Enjoyment of Books (std.)	0.15	-0.02	0.00	3433	0.15	-0.03	0.00	2819	

Grade 3

% Female	53.38	50.95	0.97	2950	54.29	51.27	0.97	2500	
% FRL	77.24	78.09	0.34	2950	76.67	77.64	0.34	2500	
% LEP	15.60	16.45	0.37	2950	16.45	16.81	0.37	2500	
% Hispanic	19.97	19.34	0.75	2950	20.56	19.78	0.75	2500	
% Black	40.50	39.88	0.61	2950	38.66	38.58	0.61	2500	
% White	23.27	23.03	0.42	2950	23.92	23.76	0.42	2500	
% Other Race	16.26	17.75	0.51	2950	16.86	17.88	0.51	2500	
Spr. Comp. Z-Score	0.00	0.04	0.94	2833	0.03	0.06	0.94	2500	
Spr. Comp Std. Score	182.16	183.00	0.94	2833	182.73	183.47	0.94	2500	
Fall Comp. Z-Score	0.00	0.03	0.95	2580	0.01	0.05	0.95	2500	
Fall Comp Std. Score	185.37	186.14	0.95	2580	185.54	186.61	0.95	2500	
# Trifolds Returned	0.04	3.48	0.00	2950	0.05	3.75	0.00	2500	
# Trifold Qs Correct	0.09	7.62	0.00	2950	0.10	8.19	0.00	2500	
# Books Read	9.24	10.36	0.00	2950	9.22	10.46	0.00	2500	
Enjoyment of Books (std.)	0.06	-0.19	0.00	2950	0.06	-0.18	0.00	2500	

Note. The analysis sample is based on students with non-missing data for pre- and posttest. *P*-values for mean T/C differences are derived from regressions of variable on indicator for treatment assignment and fixed effects for homeroom. *Comp Std. Score* is students' standard score on the ITBS reading comprehension subtest. FRL=student is eligible for free or reduced-price lunch. LEP=student is classified as limited English proficiency. "# Books Read" is based on students self-report on fall survey (see footnote 8). "Enjoyment of Books (std.)" is students' answers to fall survey question, "Overall, how much did you like or dislike the books you read this summer? The choices were: "I loved them," "I liked them," "They were okay," and "I didn't like them." Scores of 4-1 were assigned to these answers, and that variable was normalized to have mean zero and standard deviation one. Missing values for demographic variables, "# Books Read," and "Enjoyment of Books (std)" are imputed with homeroom and treatment group-specific sample means, or grade-level and treatment-group specific means when all students in homeroom are missing on the variable ("# Books Read," and "Enjoyment of Books (std)" only).

Table 2:
ITT Effect of Project READS on Self-Reported Number of Books Read

	(1) Pooled	(2) Boys	(3) Girls
<i>Grades 2 & 3:</i>			
READS	1.099*** (0.172)	1.313*** (0.263)	0.844*** (0.245)
<i>N</i>	5319	2549	2770
<i>R</i> ²	0.166	0.252	0.221
<i>Grade 2:</i>			
READS	0.993*** (0.239)	1.377*** (0.359)	0.573~ (0.343)
<i>N</i>	2819	1368	1451
<i>R</i> ²	0.176	0.268	0.228
<i>Grade 3:</i>			
READS	1.255*** (0.248)	1.321*** (0.387)	1.176*** (0.351)
<i>N</i>	2500	1181	1319
<i>R</i> ²	0.150	0.226	0.211

~p<0.10 *p<0.05 **p<0.01 ***p<0.001 Note. Standard errors are in parentheses.
 All models control for fixed effects of spring classroom and baseline test scores.
 Outcome is censored at “20 or more books.”

Table 3. Effect of Project READS on Students' Enjoyment of Books Read over the Summer (Standardized)

	(1) Pooled	(2) Boys	(3) Girls
<i>Grades 2 & 3:</i>			
READS	-0.203*** (0.026)	-0.157*** (0.040)	-0.220*** (0.036)
<i>N</i>	5319	2549	2770
<i>R</i> ²	0.157	0.228	0.215
<i>Grade 2:</i>			
READS	-0.181*** (0.036)	-0.168** (0.055)	-0.218*** (0.050)
<i>N</i>	2819	1368	1451
<i>R</i> ²	0.147	0.222	0.199
<i>Grade 3:</i>			
READS	-0.229*** (0.037)	-0.150* (0.059)	-0.222*** (0.051)
<i>N</i>	2500	1181	1319
<i>R</i> ²	0.166	0.224	0.234

~p<0.10 *p<0.05 **p<0.01 ***p<0.001

Note. Standard errors are in parentheses. All models control for fixed effects of spring classroom and baseline test scores.

The outcome is students' answers to the fall survey question, "Overall, how much did you like or dislike the books you read this summer?" The choices were: "I loved them," "I liked them," "They were okay," and "I didn't like them." Scores of 4-1 were assigned to these answers, and that variable was normalized to have mean zero and standard deviation one.

Table 4:
ITT Effect of Project READS on Fall ITBS Reading Comprehension Scores

	(1) Pooled	(2) Boys	(3) Girls
<i>Grades 2 & 3:</i>			
READS	0.014 (0.017)	0.004 (0.024)	0.033 (0.024)
<i>N</i>	5319	2549	2770
<i>R</i> ²	0.675	0.718	0.688
<i>Grade 2:</i>			
READS	0.010 (0.023)	0.027 (0.034)	-0.004 (0.034)
<i>N</i>	2819	1368	1451
<i>R</i> ²	0.662	0.703	0.677
<i>Grade 3:</i>			
READS	0.020 (0.024)	-0.024 (0.036)	0.073* (0.034)
<i>N</i>	2500	1181	1319
<i>R</i> ²	0.690	0.734	0.701

~p<0.10 *p<0.05 **p<0.01 ***p<0.001 Note. Standard errors are in parentheses. All models control for fixed effects of spring classroom and baseline test scores.

Table 5:
IV Estimates of Effect of Various Measures of Book Reading on Fall ITBS Reading Comprehension Scores, Third Grade

	(1) Pooled	(2) Boys	(3) Girls
Self-reported books read	0.008 (0.006)	0.005 (0.006)	0.010 (0.006)
<i>N</i>	2500	1181	1319
High self-reported books read	0.149* (0.076)	0.097 (0.085)	0.141~ (0.082)
<i>N</i>	2500	1181	1319
Number of tri-folds returned	0.004 (0.006)	0.001 (0.009)	0.014* (0.007)
<i>N</i>	2500	1181	1319
High number of tri-folds returned	0.035 (0.041)	0.009 (0.062)	0.115* (0.053)
<i>N</i>	2500	1181	1319
Tri-fold questions answered correctly	0.002 (0.003)	0.000 (0.004)	0.006~ (0.003)
<i>N</i>	2500	1181	1319
Liked books read over the summer	0.032 (0.037)	0.065~ (0.039)	0.002 (0.042)
<i>N</i>	2500	1181	1319

~p<0.10 *p<0.05 **p<0.01 ***p<0.001 Note. Standard errors are in parentheses. All models control for fixed effects of spring classroom and baseline test scores.