

Effects of a Voluntary Summer Reading Intervention on Reading Achievement:
Results From a Randomized Field Trial

James S. Kim
University of California, Irvine

Citation: Kim, J. S. (2006). Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results From a Randomized Field Trial. *Educational Evaluation and Policy Analysis*, vol. 28, no. 4, pp. 335-355.

ABSTRACT

The effects of a voluntary summer reading intervention were assessed in a randomized field trial involving 552 students in 10 schools. In this study, fourth-grade children received 8 books to read during summer vacation, and were encouraged by their teachers to practice oral reading at home with a family member and to use comprehension strategies during independent, silent reading. Reading lessons occurred during the last month of school in June, and 8 books were mailed to students on a biweekly basis during July and August. The estimated treatment effects on a standardized test of reading achievement (Iowa Test of Basic Skills) were largest for Black students ($ES = .22$), Latino students ($ES = .14$), less fluent readers ($ES = .17$), and students who reported owning fewer than 50 children's books ($ES = .13$). The main findings suggest that a voluntary summer reading intervention may represent a scaleable policy for improving reading achievement among lower-performing students.

Numerous empirical studies indicate that the achievement gap in reading forms and widens during summer vacation rather than during the school year. In a study of summer learning in Atlanta, Heyns (1978) found that “the gap between black and white children, and between low- and high-income children widens disproportionately during the months when schools are not in session” (p. 187). A synthesis of studies on summer learning loss (Cooper *et al.*, 1996) showed that middle-income students enjoyed reading gains during the summer whereas low-income students lost ground. Longitudinal studies have continued to show that gaps in reading achievement based on children’s socioeconomic status grow larger during summer vacation than during the school year (Alexander *et al.*, 2001; Downey *et al.*, 2004). In addition, there is some evidence that summer reading loss is greater for minority students than for White students (Heyns, 1987; Klibanoff & Haggart, 1981; Murnane, 1975; Phillips *et al.*, 1998).¹

Although there are many potential causes of summer reading loss, access to books and voluntary reading are likely to play a critical role in promoting reading achievement outside school (Cunningham & Stanovich, 1998; Entwisle *et al.*, 2000; Heyns, 1978). Some scholars have suggested that policies designed to increase access to books may keep the learning faucet open when schools are closed during summer vacation (Entwisle *et al.*, 2000). Voluntary reading interventions, in which children receive free books and are encouraged to read at home, may represent a scaleable policy strategy for promoting reading achievement during summer vacation. However, there is little experimental evidence supporting the use of voluntary reading interventions as a large-scale instructional policy. Given the limitations of previous research, the National Reading Panel (2000) underscored “a need for rigorous evaluations of the effectiveness of encouraging wide reading on reading achievement” (p. 3-27).

This experimental study examined the effects of a voluntary summer reading intervention on the reading skills of fourth-grade students in one of the nation's largest public school districts. The intervention attempts to improve reading skills by increasing children's access to books, matching books to children's reading levels and preferences, and encouraging children to read orally with a parent/family member and to practice comprehension strategies learned in school. Finding a cost-effective reading intervention is important for policymakers and practitioners given the current goals of federal education policy. Under the No Child Left Behind Act, public schools must adopt scientifically based reading interventions based on experimental evidence and accelerate the achievement of underperforming subgroups, including students from low-achieving, low-income, and minority backgrounds. Moreover, a voluntary reading intervention that improved reading outcomes for students in Title I schools in particular would complement more intensive reading interventions and compensatory education policies that occur during the regular school year (Borman *et al.*, 2005; D'Agostino & Murphy, 2004).

The paper is organized into five sections. Section I summarizes the research literature on correlational studies that examine the effects of access to books and voluntary reading on reading achievement. Section II discusses how findings from the National Reading Panel's review of experimental studies on voluntary reading motivated the design of the present study. Section III describes the methods. Section IV provides estimated treatment effects for all students and subgroups of students, and section V concludes with a discussion of the implications for research and policy.

I. Correlational Studies on Access to Books and Voluntary Reading

Several studies on summer learning have underscored the potential impact of voluntary reading on children's learning. In pioneering work on summer learning in the Atlanta public

schools, Heyns (1978) studied correlates of reading growth on a standardized test of word knowledge among a sample of nearly 3,000 students in grade 6 and 7. Heyns (1978) found that the “number of books read during the summer is consistently related to achievement gains; the strength of this relationship often exceeds that of socioeconomic status when prior achievement is controlled” (p. 119). Additional results from multivariate analyses indicated that measures of voluntary reading (e.g., number of books read, time spent on daily leisure reading) explained a larger proportion of the variation in reading than other recreational and enrichment activities. Based on these findings, Heyns suggested that “[w]hatever the reasons, the unique contribution of reading to summer learning suggests that increasing access to books and encouraging reading may well have a substantial impact on achievement” (p. 172). Similarly, Phillips and Chin’s (2004) analysis of the Prospects study of Title I from the 1990s revealed that “reading with children, encouraging them to read on their own, and providing access to a wide range of new books, improve[d] children’s performance on reading comprehension and vocabulary tests” (p. 278). In addition to studies focusing on summer learning, numerous studies have revealed strong positive correlations between reading books outside school and improvements in general reading ability (Anderson *et al.*, 1988; Donahue *et al.*, 2001).

Analyses of national survey data suggest that policies designed to increase minority children’s access to books at home could potentially ameliorate ethnic disparities in reading achievement. For example, Fryer and Levitt (2004) analyzed data from the Early Childhood Longitudinal Survey (ECLS-K), a nationally representative sample of approximately 20,000 children entering Kindergarten in fall 1998, and found that White families ($M = 93, SD = 65$) reported owning more books than Black families ($M = 39, SD = 42$), Latino families ($M = 41, SD = 48$), and Asian families ($M = 49, SD = 56$). In multivariate analyses exploring ethnic

disparities in test score performance during the first two years of school, Fryer and Levitt found that the inclusion of a composite measure of socioeconomic status and the number of books in children's homes accounted for the entire gap in reading scores between Black students and White students and most of the gap between Latino/a students and White students. Based on these results, Fryer and Levitt argued that the number of books in a child's home was a "useful proxy for capturing the conduciveness of the home environment to academic success" (p. 452). In addition to the results from the ECLS-K, similar ethnic disparities in book ownership between White and minority students have also been noted on the fourth-grade NAEP reading assessment (Donahue et al., 2001) and the Minority Student Achievement Network's survey of over 40,000 high school students in ten multiethnic suburban school districts (Ferguson, 2002). Ultimately, it is difficult to establish a causal link between access to books and reading achievement based on multivariate analyses of survey data. Nonetheless, previous research suggests that giving children more books to read during the summer could encourage more reading practice and better reading outcomes for Black and Latino/a students.

II. The National Reading Panel's Review of Voluntary Reading Interventions

The key question, then, is whether voluntary reading interventions are an effective and scaleable instructional policy for improving reading fluency, vocabulary, and comprehension. To address this question, the National Reading Panel's report, *Teaching Children to Read* (2000), reviewed 14 studies that focused on the effects of a "widely recommended approach to developing fluent readers—encouraging children to read a lot" (p. 3-21). Policies designed to encourage voluntary reading have different names (e.g., sustained silent reading), but they share three characteristics: students choose their own books, read silently on their own, and receive little or no feedback on the selection of books or the reading activity from teachers or parents

(Filgreen, 2000). As noted by Chall (2000), voluntary reading can be viewed as a “student-centered” instructional policy because it assumes that “learning is accomplished on one’s own, based on one’s interests” (p. 34). In general, the National Reading Panel’s review found little experimental support for the use of voluntary reading as an effective instructional policy (p.3-28). In particular, the National Reading Panel noted that the inconsistent and inconclusive findings on the effects of voluntary reading could stem from either “deficiencies in the instructional procedures themselves or to weaknesses and limitations evident in the study designs” (p. 3-27).

Four findings from the National Reading Panel’s review shaped the design of the current study. First, research design appeared to mediate the treatment effects obtained from the 14 studies. None of the 6 studies using a quasi-experimental design reported positive effects (Carver & Leibert, 1995; Cline & Kretke, 1980; Morrow & Weinstein, 1986; Peak & Dewalt, 1994; Summers & McClelland, 1982; Vollands *et al.*, 1999). In contrast, 5 of the 8 experimental studies showed positive and significant treatment effects (Burley, 1980; Davis, 1988; Holt & O’ Tuel, 1989; Langford & Allen, 1983; Manning & Manning, 1984). Thus, all of the positive effects of voluntary reading were restricted to randomized experiments, which yield more valid causal estimates of treatment effects than quasi-experimental studies. Since most of the experimental studies involved a small number of students, a large multi-site randomized field trial would address design limitations in previous research.

Second, voluntary reading interventions usually involved students in grade 5 and above and the benefits were strongest among older students. Among the 14 studies reviewed by the National Reading Panel, none involved first-grade students, and the two studies involving second-grade students showed no significant effects (Collins, 1980; Morrow & Weinstein, 1986).

Indeed, 12 of the 14 studies involved students in grade 5 to high school. In most studies of voluntary reading, researchers usually focused on older students who had sufficiently strong decoding skills to read words independently. Put another way, there is no reason to believe that a voluntary reading intervention would work if students were given free books but were unable to decode the words. Developmental theories of reading assume that decoding ability must be sufficiently strong for students, especially by the beginning of fourth-grade when children must use reading as a tool for “learning the new” (Chall, 1983). As noted by the National Reading Council (1998), in “fourth grade and up, it is taken for granted that they are capable— independently and productively—of reading to learn” (p. 207). Thus, prior research suggests that voluntary reading would benefit students in the upper elementary grades who can decode words but need additional reading practice in order to improve their reading fluency, vocabulary, and comprehension.

Third, the failure to match texts to readers was cited as a major reason why voluntary reading interventions were ineffective. For example, among the 14 studies, the National Reading Panel cited a study by Carver and Liebert (1995) as providing “one of the clearest tests of the effect of reading by studying students during the summer” (p. 3-26). In this study, 43 students in grades 3, 4, and 5 read easy fiction books for six weeks (Monday through Friday), 2 hours each day. The 60 hours of leisure reading time did not translate into higher reading levels, increased reading rate and efficiency, or vocabulary gains. Since students were allowed to self-select texts, most students selected easy reading materials, which were several grade levels below their independent reading level. As a result, Carver and Liebert concluded that their study “was seriously flawed in that the apparent difficulty levels used to assign books were not the real levels as measured objectively by a readability formula” (p. 43). They added that the results

“may be explained by a failure to manipulate the difficulty level of the material read” (p.44).

This study suggests that improving the quality of the match between reader level and text difficulty could enhance the effects of voluntary reading.

Fourth, the National Reading Panel’s review suggested that guided oral reading and comprehension strategies enhanced the effectiveness of reading practice. Moreover, the National Reading Panel noted that teaching oral guided reading and comprehension strategies did not require a large investment of instructional time. The Panel found strong evidence that guided oral reading strategies, in which children receive feedback during oral reading of text, improved fluency, and to a lesser extent, comprehension. This finding is important because fluent readers can read connected text accurately, quickly, and with proper expression, thereby freeing the mind to focus on the meaning of the text (LaBerge & Samuels, 1974; Stahl, 2004). Less fluent readers, however, often lack the skills to understand more challenging texts encountered in fourth-grade reading assessments (Daane *et al.*, 2005). In addition to oral reading of text, the National Reading Panel found that comprehension strategies used by good readers—question generation, question answering, summarizing, and re-reading—improved understanding of text. Research on the use of multiple strategies also indicated that achievement gains were similar regardless of whether teachers spent 6 or 25 classes teaching these strategies (Rosenshine & Meister, 1994). Moreover, the National Reading Panel found that most studies involved older children, which implies that “researchers taught readers who had achieved decoding and other basic reading skills before they were taught strategies” (p. 4-51). The Panel’s findings indicated that encouraging students to practice oral reading of text and comprehension strategies during silent reading could potentially enhance the effects of reading practice.

Overview of Study Design and Research Questions

The current study was intended to address both the methodological and instructional limitations of previous research on voluntary reading. With respect to methodology, we conducted an experimental study to examine the causal effects of the intervention on the reading achievement of fourth-graders. In Chall's (1983) stage theory of reading, fourth-grade presents a key transitional point in schooling because it is often assumed that students have mastered the foundational skills needed to decode individual words (Chall, 1983). For example, the California framework for English language arts (California Department of Education, 1999) states that "third grade is often considered the last period of formal instruction in decoding...[and] at the end of this pivotal year, instruction in phonics is phased out from the formal curriculum" (p. 80). Therefore, we hypothesized that most fourth-graders could decode words independently and that poor readers in particular needed opportunities to read books and practice their reading skills during the extended summer recess.² Another advantage of targeting fourth-grade students is the availability of data from the main NAEP reading assessment (Grade 4), which can be used to compare the achievement level of students in the study sample to a nationally representative sample of U.S. school children (Daane et al., 2005; Perie *et al.*, 2005).

With respect to the instructional design, the current intervention addressed multiple factors—the characteristics of the students, the books, and the home context—that are likely to shape opportunities to read in the summer and affect reading outcomes (RAND Reading Study Group, 2002). Thus, one goal of the instructional design was to address the variability in access to books, reading levels, and reading preferences between White and minority students. To increase access to books, each student in the treatment group received 8 free books to read during summer vacation. To facilitate a better match between text difficulty and reader ability, a text leveling system was used to provide books that were within each student's independent

reading level. One widely used method for matching books to readers is the Lexile Framework (U. S. Department of Education, 2001). This framework bases the readability of individual texts on measures of semantic difficulty (word frequency) and syntactic difficulty (sentence length) as in traditional measures of readability (Chall & Dale, 1995). Like other readability measures, the Lexile Framework has shown that measures of semantic and syntactic difficulty are the best predictors of text difficulty. For this study, the reading level of individual students was based on a standardized test of reading (Iowa Test of Basic Skills), and reading preferences were obtained through a spring survey. Based on information from the reading test and survey, students received books that were within their independent reading level and matched to their reading preferences.

Three research questions motivated this study: (1) Did the intervention increase children's access to books at home and literacy related activities during summer vacation? (2) What is the magnitude of the treatment effect for all students and by student ethnicity and income status? (3) What is the magnitude of treatment effect based on pretest measures of silent reading ability, oral reading fluency, and self-reported measures of children's ownership of books?

III. Methods

District Context for the Randomized Field Trial

The randomized field trial took place in the Lake County Public School District, a large, multi-ethnic school district located in a mid-Atlantic state.³ Earlier studies on summer learning in Lake County indicated that Black and Latino students from both low-income and middle-income families lost ground in reading during summer vacation in the upper elementary grades (3 to 5). In an attempt to address the problem of summer reading loss, we designed the current

intervention in order to expand summer learning opportunities and to improve the reading skills of minority students and less skilled readers. Therefore, the practical goal of addressing racial and ethnic disparities in reading achievement motivated the sampling plan.

With over 100 elementary schools in Lake County, the district is organized into small sub-districts, each with its own assistant superintendent.⁴ We selected one sub-district based on two criteria: (1) it included high-poverty schools that administered Title I schoolwide programs; and, (2) it included multiracial schools in which reading scores for Black and Latino students contributed to the federal adequate yearly progress ratings. To select the school sample, we rank ordered the schools by the percentage of Black and Latino students and included enough schools to have approximately 500 students in the study.⁵ The final sample included four Title I schools and six non-Title I schools with the largest percentage of minority students.⁶ Our study was designed to have sufficient power to detect small to modest sized effects between .10 and .20 standard deviation units, which was the magnitude of the treatment effects from a pilot study of the current intervention and remedial summer programs (Cooper *et al.*, 2000). We also wanted to gather evidence on the effects of the intervention for less-skilled readers, minority students, and low-income students before scaling up in the lowest performing schools. In short, we planned to use evidence from this study to inform the next phase of research and policy implementation.

School principals agreed to implement the experimental design, and students were required to obtain active consent from their parents. To obtain high participation, we used incentives (e.g., ice cream, pizza, extra recess) to encourage students to return consent forms and translated the permission forms into five different languages in order to increase participation

among parents whose primary language was not English. Overall, 85% of the students returned their consent forms and participated in the study.

Table 1 provides descriptive statistics for the demographic and achievement variables. Each minority subgroup includes at least 10% of the sample, facilitating analyses of separate treatment effects for White and minority students.⁷ In addition, the national percentile rank (NPR) of 52 on the ITBS was near the national norm. The mean reading fluency, which is a combination of reading accuracy and rate, was 120 words correctly read per minute. This figure is similar to the national norm of 123 reported in a 2005 norming study of nearly 16,000 students in grade 4 (Behavioral Research & Teaching, 2005) and the 119 average from the 2002 NAEP study of oral reading in grade 4 (Daane et al., 2005). Finally, the reading attitude score ($M = 58.45$) translates into a NPR of 58 based on norms from a national study of elementary school children's attitudes toward reading (McKenna & Kear, 1990).

[INSERT TABLE 1]

To provide a context for understanding ethnic disparities in reading achievement and children's access to books during the summer, Table 2 displays cross tabulations between column variables denoting student ethnicity and row variables denoting reading ability and ownership of books. The large chi-square statistics reveal significant associations between student ethnicity and each of the subgroup categories. More precisely, a larger percentage of minority students than White students scored below the median on the ITBS and the measure of reading fluency, and reported owning fewer than 100 books in general or 50 kids books in particular.

[INSERT TABLE 2]

Measures

Five sets of measures were collected for this study, and descriptions of the measures are provided below.

1. *Student Demographic Measures.* Student demographic variables were obtained from the Lake County Public Schools testing and evaluation office. The demographic variables were dummy-coded (1 = yes, 0 = no) and included data on gender, free lunch status, English language proficiency, and whether or not the student was White, Black, Latino, Asian, or other ethnicity. The “other ethnicity” category included 2 Native American students and 21 multiethnic students. These demographic variables are used to determine the performance of subgroups as required by the No Child Left Behind Act’s accountability rules.⁸

2. *Reading Outcomes.* All students were administered a pre and posttest measure of silent reading and oral reading.

The Iowa Test of Basic Skills (ITBS), Form A, was the primary outcome used to measure reading achievement. The ITBS was nationally normed in 2000 and is a widely used standardized test of general reading ability. The complete battery of the ITBS, Form A (Total Reading), Level 10, was administered to grade 4 students in June 2005 and Form B was administered to grade 5 students in September 2005. The KR-20 reliability coefficients for the ITBS reading test are above .93 in the spring and fall (Hoover *et al.*, 2003). Spring ITBS scores were also used to obtain a reading level for each student that was converted into Lexiles. The ITBS also provides a Lexile range for each student. A 100-point Lexile range on the ITBS corresponds to the independent reading level used to match books to students (U. S. Department of Education, 2001). The ITBS provides a 100-point Lexile range (+50 Lexiles above student’s observed score and -50 Lexiles below), which represents each student’s independent reading

level. Each student's Lexile range was used to select books within the student's independent reading level.

To assess oral reading fluency, retired elementary school teachers administered an oral fluency assessment from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Reliability coefficients (alternate forms) on the DIBELS oral fluency assessment have ranged from .89 to .94 (Good & Kaminski, 2003; Tindal *et al.*, 1983), and the test-retest reliability from this study was .89. The oral fluency assessment (Good & Kaminski, 2003) is individually administered to students and provides a measure of the number of "words correctly read per minute" (WCPM), which is a widely used indicator of oral reading fluency (Fuchs *et al.*, 2001). Each student read a grade level passage in the spring and fall for one minute. To compute the WCPM, each tester subtracted the number of decoding errors from the total number of words read.⁹

3. *Spring Reading Survey.* Students were administered a spring reading survey, which included a 20-item Elementary Reading Attitude Survey (ERAS) and a 25-item reading preferences survey.

The Elementary Reading Attitude Survey (ERAS) is a 20-item scale used to measure attitudes toward academic reading (10 items) and recreational reading (10 items). The ERAS was normed in 1989 in a national sample of 18,138 students in grades 1 to 6. For grade 4, Cronbach alpha reliabilities were .89 for the full scale, .83 for the recreational subscale, and .83 for the academic subscale (McKenna & Kear, 1990).

The reading preferences survey asked students how much they enjoyed reading books from one of 25 categories of children's books. The prompt asked students how they would feel reading books from a particular genre/category of children's literature. Each response option

included a picture (smiley faces) representing 1 of 4 options: (a) I don't like it, (b) It's okay, (c) I like it, and (d) I really like it! The reading categories were developed using an article on reading preferences by the American Library Association (2003), and recent editions of the National Council of Teachers of English's (NCTE) *Adventuring with Books* series (McClure & Kristo, 2002). The reading categories were validated using other published surveys on children's reading preferences (Galda *et al.*, 2000; Ivey & Broaddus, 2001; Monson & Sebesta, 1991; Summers & Lukasevich, 1983).¹⁰ In addition, four elementary school English language arts teachers checked the face validity of the categories and helped to revise and simplify the readings categories used in the spring reading preference survey. There were 25 reading categories.¹¹ A total of 240 books were used in the study and were placed into the reading categories.¹² Each book was available through two major children's books sellers (Scholastic, Borders), and contained information on text difficulty (Lexile level). Survey data was used to match 8 books to children's preferences. In particular, matched books were selected for each child by a computer algorithm that merged data from two files. One file contained a Lexile level and preference categories for each of 240 available book titles. The second file contained each student's Lexile range from the spring ITBS and reading preferences from the spring survey. The algorithm generated a list of the eight best matches for each child.¹³

4. *Fall Reading Survey.* The fall survey included measures of reading activity during the summer and access to books at home.

The literacy habits survey (Paris *et al.*, 2004) has been used in previous studies involving beginning readers (grade 1 to 3). The measure was adapted for this study, which involves students in the upper elementary grades and focuses specifically on reading during summer vacation. Each of the items begins with the statement, "During summer vacation" and is

followed by 1 of 5 activities: (1) how often did you read at home for fun? (2) how often did you read books or stories at bedtime? (3) how often did you read books? (4) how often did your parents help you read at home? (5) how often did you read out loud to someone at home? The four response options were (a) never or hardly ever, (b) once or twice a month, (c) once or twice a week, and (d) almost every day. The internal reliability coefficient for the total score on the 5-item literacy habits measure was .71. Since the literacy habits scale included items related to both silent and oral reading, we conducted a principal components analysis with varimax rotation to examine the dimensionality of the data (Comrey & Lee, 1992). The analysis revealed two dimensions of literacy related activities, including a first component describing independent reading activity (eigenvalue = 2.4) and a second component (eigenvalue = 1.1) describing social interaction around books with family members.¹⁴

Two measures of book ownership were also included in the fall survey. The first pertained to children's books and asked students "Some homes have 0 books for kids while others have more than 50 books for kids. About how many books for kids do you have in your home?" Response options included (a) 0-10 books, (b) 11 to 20 books for kids, (c) 21 to 30 books for kids, (d) 31 to 50 books for kids, and (e) more than 50 books for kids. The second measure was taken from the National Assessment of Educational Progress (NAEP) grade 4 reading assessment. It asked students "about how many books are in your home?" which is followed by 1 of 4 options: (a) few (0-10), (b) enough to fill one shelf (11-25), (c) enough to fill one bookcase, (d) enough to fill several bookcases (more than 100). To facilitate interpretation of analyses, we converted the scale to two new measures: (1) the percentage of students who reported owning more than 100 books, and (2) the percentage of students who reported owning more than 50 books.

6. *Summer Reading Postcards*. Students in the treatment group received a postcard with each book. The postcard encouraged students to practice the oral reading and silent reading skills taught by their teachers. The first set of questions asked students to write down the title of the book, and whether they read the book (yes, no). The next question asked students: what did you do to better understand this book? (check all that apply). The options included, (a) I re-read parts of this book, (b) I asked questions about this book, (c) I made predictions about this book, (d) I summarized parts of this book, and (e) I made connections (text to text, text to self). The final question prompted students to read orally with a parent or family member, and to ask whether they read the passage more smoothly, with more expression, and knew more words after a second reading. A line was provided for an adult signature indicating that the student read aloud from a 100-word passage.

Procedures for Teacher Training, Reading Lessons, and Delivery of Treatment

In June 2005, all participating teachers in fourth-grade attended a 2 hour training session after school to learn how to administer the reading tests and surveys and how to teach the reading lessons. A veteran English language arts teacher who had participated in a 2004 pilot study led the training session. In particular, she reviewed two strategies for encouraging both silent reading and oral reading. First, teachers were asked to review five comprehension strategies that help children understand text during silent reading as noted by the National Reading Panel (2000). During the training, the lead teacher used the children's book, *The Wreck of the Zephyr*, to model five comprehension strategies: re-reading, asking questions, making predictions, summarizing, and making connections to self and to other text. Second, teachers were asked to instruct their students about paired reading (Topping, 1987), a widely used oral fluency strategy in which a student chooses a favorite part of a book (100 words) to read out loud to a parent or

family member. During the second reading, the student asks if he/she read with more prosody—that is, did they read smoothly, with expressiveness, and knowledge of more words. These lessons were designed to help students practice oral reading with their parents or another family member. A letter (English and Spanish) accompanied each postcard, encouraging a family member to read aloud with students.¹⁵

During the last two weeks of school, teachers modeled the comprehension strategies and paired reading in their classrooms and assigned homework assignments asking students to practice filling out the postcards at home.¹⁶ The postcard asked students to check comprehension strategies used during independent reading and to obtain a signature from a parent or family member after reading aloud from the text. Since both treatment and control students were in the classrooms during the reading lessons, this study examines the value-added of giving children free books and encouraging reading practice at home during summer vacation. Children in the treatment group received their books and postcards by airmail during July and August. Thus, the only difference between the two groups is that the treatment group received additional resources (i.e., books, reading postcards, and letters to a parent/family member) during summer vacation whereas control students received these materials after fall posttests.

IV. Results

Analytic Strategy

A total of 552 students received consent to participate in the study and took pretests in June 2005. Students were randomly assigned to treatment and control groups within their English language arts classroom (i.e., the blocking variable). Randomization of students within each of the 34 classroom blocks was intended to create probabilistically equivalent treatment and control groups with respect to teacher effectiveness in reading and language arts classroom. All

analyses include fixed effects for the classroom in which students were randomly assigned to experimental conditions.

As would be expected from the random assignment of students to the treatment and control group, there were no statistically significant differences between the two groups at the beginning of the experiment on numerous demographic and achievement characteristics. Results from Table 3 suggest that the percentage of minority students, low-income students, students with limited English proficiency, and mean scores on the reading measures were statistically equivalent in both groups. When these analyses were conducted for each of the four major ethnic groups, there were no statistically significant differences between treatment and control groups on most observed characteristics. However, among Asian students, the control group ($M = 45\%$) had a significantly larger percentage of females than the treatment group ($M = 28\%$), suggesting baseline differences in the gender composition of the two groups.

[INSERT TABLE 3]

A total of 552 students completed baseline ITBS reading tests, including 282 treatment group students and 270 control group students. Due to attrition during the summer, the final analytic sample included 486 students, including 252 students in the treatment group and 234 students in the control group. A total of 66 students moved during the summer. Since these students had missing posttest scores, they were not part of the final analytic sample. Although missing data represents a potential threat to internal validity, there was no systematic relationship between missing ITBS test scores and experimental condition, $\chi^2(1, N = 552) = .95, p = .33$. Among students who moved during the summer, there was also no statistically significant difference in mean ITBS scores for the 30 students in the treatment group ($M = 199.43, SD = 26.75$) and 36 students in the control group ($M = 194.44, SD = 22.35$), $t(64) = .83, p = .41$. This

suggests that the initial reading skills of treatment and control group students excluded from the analyses were statistically equivalent.

Using data for the analytic sample of 486 students, we conducted analyses to address three research questions. First, we conducted a two-way analysis of variance (ANOVA) to examine whether the intervention increased children's access to books and literacy activities during summer vacation. Second, we conducted a series of ordinary least squares (OLS) regression analyses to estimate treatment effects for all students and separately for each of the four ethnic groups. We also conducted an analysis of covariance (ANCOVA) to examine whether the treatment had differential effects based on student ethnicity and income status. Third, we estimated the magnitude of the treatment effect for subgroups of students based on pretest measures of oral reading fluency, silent reading ability, and access to books.

Research Question #1: Did the intervention increase children's access to books at home and literacy related activities during summer vacation?

A two-way analysis of variance on both self-reported measures of book ownership revealed two key findings. First, the two-way ANOVA revealed no significant difference in the percentage of students in the treatment group ($M = 35\%$) and the control group ($M = 35\%$) who reported owning more than 100 books, $F(1, 460) = .00, p = .99$, nor was there an interaction between the treatment and student ethnicity, $F(3, 460) = .67, p = .573$. Similarly, the two-way ANOVA revealed no significant difference in the percentage of students in the treatment group ($M = 39\%$) and the control group ($M = 40\%$) who reported owning more than 50 kids books, $F(1, 459) = .02, p = .89$, nor was there an interaction between the treatment and student ethnicity, $F(3, 459) = 1.203, p = .308$. Second, there were significant differences on both measures of book ownership between White and minority students. A two-way ANOVA on the first measure of book ownership (100 or more books) revealed a significant main effect of ethnicity, $F(3, 460)$

= 16.492, $p < .001$. A two-way ANOVA on the second measure of books ownership (50 or more kids books) also revealed a significant main effect of ethnicity, $F(3, 467) = 13.113, p < .001$. We conducted pairwise comparisons with Bonferonni adjustments to examine differences in books ownership between White students and each of the three minority subgroups. These comparisons showed that a significantly larger percentage of White students ($M = 58\%$) reported owning 100 books than Black students ($M = 24\%$), Latino students ($M = 24\%$), and Asian students ($M = 32\%$). Furthermore, a significantly larger percentage of White students ($M = 61\%$) reported owning 50 or more kids books than Black students ($M = 30\%$), Latino students ($M = 33\%$), and Asian students ($M = 35\%$).

A series of two-way analyses of variance on the literacy habits score revealed three findings. First, a two-way ANOVA on the total literacy habits score revealed no significant difference between the treatment and control group, $F(1, 454) = .49, p = .48$, nor an interaction between the treatment and student ethnicity, $F(3, 454) = .04, p = .99$.¹⁷ Second, a two-way ANOVA on the silent reading factor score revealed no significant difference between the treatment and control group, $F(1, 454) = .04, p = .85$, nor the interaction between the treatment and student ethnicity, $F(3, 454) = .51, p = .674$. Third, a two-way ANOVA on the oral reading factor score revealed a significant difference between the treatment and control group, $F(1, 454) = 4.31, p = .045$, but no significant interaction between the treatment and student ethnicity, $F(3, 454) = .51, p = .68$.

To summarize, results from these analyses suggest that the intervention did not increase children's access to books nor the amount of independent, silent reading. However, children in the treatment group engaged in significantly more oral reading activities at home with family members than control group children.

Research Question #2: What is the magnitude of the treatment effect for all students and by student ethnicity and income status?

Since the purpose of this study was to examine the effects of a voluntary summer reading on student reading outcomes, individual students were assigned to treatment and control groups and analyses involved the estimation of treatment effects at the student-level. In other words, this study did not involve random assignment of schools to experimental conditions and there was no attempt to use hierarchical linear models (Raudenbush & Bryk, 2002) as is common in studies of schoolwide interventions such as Success for All (Borman et al., 2005). Ordinary least squares (OLS) regression is the most appropriate method for analyzing the achievement of individual students during summer vacation (Burkham et al., 2004).¹⁸ Formally, the following OLS model was used to estimate the treatment effects:

$$(1) Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 RB_i + \varepsilon_i$$

where Y_i represents the fall reading scores for the i^{th} student, T_i is a binary variable representing the experimental condition (i.e., treatment or control group), X_i is the baseline covariate (i.e., spring ITBS score), RB_i is the randomization block (i.e., fourth-grade English language arts classroom) in which a student was randomly assigned to treatment or control groups, and ε_i is the residual for the i^{th} student. Using equation (1), the goal is to obtain a parameter estimate for β_1 , which represents the treatment effect. This intention-to-treat estimate provides an unbiased estimate of the intervention's impact on posttest reading scores and the baseline covariate is designed to improve the precision of the estimated treatment effect. To facilitate interpretation of the results, standardized scores (i.e., z-scores) were used in the OLS models. Thus, the coefficient for the treatment effect represents the effect size, or standardized mean difference, between the treatment and control group.

The regression results for the ITBS reading score for all students and by ethnicity are displayed in Table 4. Among all students, the estimated treatment effect ($B = .08$, $SE = .04$, $t = 1.90$, $p = .059$) was marginally significant. The treatment effect for all students, however, was smaller than the treatment effect for Black and Latino students. For example, the estimated treatment effect was significant for Black students ($B = .22$, $SE = .09$, $t = 2.59$, $p = .011$), marginally significant for Latino students ($B = .14$, $SE = .08$, $t = 1.76$, $p = .081$), and non-significant for White students ($B = .11$, $SE = .09$, $t = 1.23$, $p = .22$). Although the treatment effect for each of these subgroups was positive, there was a negative and non-significant effect for Asian students ($B = -.17$, $SE = .11$, $t = -1.55$, $p = .125$). This idiosyncratic finding may stem from baseline differences in the gender composition of the treatment and control group.¹⁹

Since the magnitude of the treatment effects varied across the four ethnic groups, we replicated the OLS results using an analysis of covariance with interaction terms for the treatment group and student background characteristics.²⁰ The ANCOVA revealed a significant interaction between the treatment and student ethnicity, $F(3, 446) = 2.91$, $p = .035$, but no significant interaction based on student income status, $F(1, 446) = 2.27$, $p = .133$. There was also no significant three way interaction involving the treatment, student ethnicity, and income status, $F(3, 446) = 1.14$, $p = .33$.

[INSERT TABLE 4]

Table 5 displays the regression results for oral reading fluency outcomes, which was defined as the number of words read correctly in one minute. The estimated treatment effects were not statistically significant in the OLS models for all students and for each of the four ethnic groups. In short, the treatment students did no better than control students on a grade level measure of oral reading fluency in the fall.

[INSERT TABLE 5]

Research Question #3: What is the magnitude of treatment effect based on pretest measures of silent reading ability, oral reading fluency, and self-reported measures of children's ownership of books?

Since the positive treatment effects were largely concentrated among Black and Latino students, we used both OLS regression and ANCOVA to further examine whether subgroups of students with a large percentage of minority students benefited the most from the intervention. First, since there was a significant association between student ethnicity and pretest reading scores, we conducted OLS regression to estimate treatment effects for subgroups of students based on a median split of spring reading scores and ownership of books. According to the first two rows of Table 6, the largest positive effects were concentrated among students who scored below the median on the ITBS reading assessment ($ES = .10$) and the measure of oral reading fluency ($ES = .17$). Second, since there was an association between student ethnicity and measures of access to books, we estimated treatment effects based on self-reported data on ownership of books. As shown in the last rows of Table 6, the positive effects were concentrated among children who reported owning fewer than 100 books ($ES = .10$) or fewer than 50 kids books ($ES = .13$). Third, we included interaction terms to examine whether treatment effects were significantly different for subgroups based on prior reading and ownership of books. None of the interaction terms was statistically significant. In sum, the magnitude of the treatment effects was largest among lower-performing students, and there were no significant interactions between the treatment and measures of reading ability or ownership of books.²¹

[INSERT TABLE 6]

Discussion

This study reports findings from a randomized field trial of a voluntary summer reading intervention involving Grade 4 students in a multiracial public school district. In our primary analyses, we compared standardized test scores for students who were randomly assigned to a treatment or control group. The estimated treatment effects reveal the impact of giving children 8 books to read during summer vacation and encouraging them to engage in independent reading of books and oral reading of text with family members. Formally, the evaluation of the present intervention is a randomized encouragement design (Hirano *et al.*, 2000), which provides the causal effect of *assignment* to the treatment rather than *receipt* of the treatment. In other words, although teachers encouraged children to read books during summer vacation, children were not required to comply with the actual treatment protocol, which was formalized through end of year reading lessons.

The main study findings are tentative but promising. The magnitude of the average treatment effect for all students was .08 standard deviations, which is “small” by social science standards (Cohen, 1988; Lipsey & Wilson, 1993). Moreover, there was no significant treatment effect for White students, students who reported owning more than 100 books in general or 50 kids books, and students scoring above the median on pretest measures of reading ability. These findings indicate that wide scale adoption of the current intervention would not be an effective policy for improving the reading achievement of all students. The primary findings contribute to the research literature on the effects of summer programs in particular and voluntary reading in general.

To begin, it is useful to compare our findings to results from a review of remedial summer school programs. This comparison is relevant because the benefits of the current intervention were restricted largely to children with weaker reading skills, and such students are

likely candidates for a remedial summer school program. In a meta-analysis of summer school programs, Cooper et al. (2000) found that random assignment studies of remedial programs yielded an average effect size of .14, which is similar to the magnitude of the positive effects observed in the current study. Cooper et al. also found that summer programs had larger effects on the achievement of students from middle-class families ($ES = .46$ to $.56$) than students from disadvantaged backgrounds ($ES = .20$ to $.24$).

Unlike results from Cooper's meta-analysis, the current intervention did not produce larger benefits for advantaged students. Instead, the magnitude of the effect sizes was larger among disadvantaged subgroups of students, and the interaction between the treatment and student ethnicity revealed larger effects for minority students than for White students. First, when we analyzed the data for different subgroups of students, the largest effect sizes were concentrated among Black students ($ES = .22$) and Latino students ($ES = .14$); students who scored below the median on a pretest measure of oral reading fluency ($ES = .17$) and silent reading ability ($ES = .10$); and students who reported owning fewer than 50 kids books ($ES = .13$) or 100 books in general ($ES = .10$). Second, when we examined statistical interactions between the treatment and student income status and measures of book ownership, none of the interaction terms was statistically significant. Thus, the major findings suggest that the current intervention would not expand disparities in reading achievement between White and minority students. In addition, the positive effects were largely concentrated among underperforming students, whether defined by minority status, skill level, or the availability of books at home.

This study provides experimental evidence that confirms the robust relationship between reading achievement and voluntary reading of books outside school. Numerous studies have also shown that cumulative differences in children's exposure to print and reading practice contribute

to the widening gap in reading achievement between skilled and unskilled readers (Cunningham & Stanovich, 1998; Stanovich, 1986, 2000). Since good readers usually find reading to be an enjoyable and rewarding experience, they engage in more self-initiated reading practice than poor readers and frequent reading practice further strengthens children's reading skills (Juel, 1988). As noted by the National Reading Panel (2000), there are "literally hundreds of studies that find that the best readers read the most and that poor readers read the least" (p. 3-21). As a result, researchers have suggested the importance of encouraging more leisure reading among minority students (Ferguson, 2001) and providing struggling readers with enjoyable reading experiences with parents and family members (Baker, 2003). Some research finds that additional leisure reading outside school would be especially beneficial for lower-performing students. In an analysis of reading growth from second- to fifth-grade, Anderson, Wilson, and Fielding's (1988) found that the relationship between time spent reading books and reading achievement was curvilinear since "reading comprehension rises sharply between 0 and about 10 minutes a day of book reading and then levels off" (p. 297). In other words, there were diminishing returns to reading practice and the largest gains were concentrated among children who devoted little time to leisure reading. Thus, a voluntary intervention that provided children with books and encouraged reading practice at home during summer vacation would potentially have the greatest benefit for poor readers. In general, the results from this study highlight the cognitive benefits of wide reading outside school.

Implications for Policy and Practice

There is a growing recognition that schools alone cannot accomplish the goal of reducing academic achievement disparities, and that summer programs can supplement the educational services provided to underperforming students (Lauer *et al.*, 2006; U. S. Department of

Education, 2003b). Therefore, voluntary summer policies represent a pragmatic effort to improve the reading achievement of minority and low-income youth. This study highlights several conditions under which such policies might be most effective.

Voluntary reading of books may produce greater benefits when children engage in both independent reading practice and oral reading of text with family members. Traditionally, voluntary reading programs have focused primarily on providing children with opportunities to read and allotting time to activities such as “uninterrupted sustained silent reading” (USSR) and “super quiet independent reading time” (SQUIRT) (National Reading Panel, 2000; Shanahan, 2004). Rather than viewing voluntary reading as a purely independent activity, our goal was to improve children’s engagement with books by encouraging both independent, silent reading of books and oral reading of text with family members. Indeed, the results from this study suggest that treatment group children read more frequently with their parents and family members than the control group children. Previous research also suggests that reading aloud from children’s books produces cognitive and affective benefits for elementary school children (Baker, 2003). Thus, voluntary reading can be viewed as both a solitary and social activity that is designed to foster learning and recreation.

Recent research highlights the importance of providing children with some minimal reading instruction and opportunities to practice reading regardless of whether the context for these activities is a school classroom or a child’s home. To date, policymakers have sought to implement (1) cheap programs that encourage children to read books at home or (2) intensive programs that provide children with instruction in summer school classrooms. On one hand, some policies such as the No Child Left Behind’s Summer Reading Achievers programs (2003a) and Connecticut’s Summer Reading Challenge (2001) encourage wide reading of books at home

and provide children with little formal instruction. On the other hand, non-profit organizations such as BELL (Building Educated Leaders for Life) (Chaplin & Capizzano, 2006) and Teach Baltimore (Borman & Dowling, 2006) have recently implemented voluntary summer programs that provide instruction in classroom settings, including phonics instruction, read alouds, and wide reading of children's books and multicultural literature. With respect to the duration and intensity of the instruction offered to children, the present intervention stands in the middle of two dominant approaches to summer programs. At the end of the school year, regular classroom teachers instructed children how to use comprehension strategies during independent reading of books and to read aloud from their favorite passages with family members. Thus, teachers provided instruction in a classroom setting during the last month of school year in June, and the reading activities occurred in children's homes during summer vacation. Nonetheless, the magnitude of the treatment effects from the current intervention falls in line with one-year impact estimates based on rigorous evaluations of the BELL and Teach Baltimore summer programs.²² Thus, the current intervention may represent a scaleable and effective policy for improving the reading achievement of lower-performing students during summer vacation.

Limitations and Future Research

Additional research is needed to address several limitations of the current study. First, the results showed no difference in the estimated treatment effects based on student income status. This result conflicts with the faucet theory of learning (Cooper et al., 1996; Entwisle et al., 2000) which suggests that providing additional books and instruction would benefit poor children more than middle-class children. One limitation of the present study was the use of a coarse binary measure of socioeconomic status (SES)—that is, whether or not a student received free lunch. Usually, studies have employed granular measures of socioeconomic status (SES)

such as family income (Heyns, 1978) or a composite based on parental income, occupational status, and education (Downey et al., 2004; Entwisle et al., 1997). A more sensitive measure of SES would increase the power to detect significant treatment effects among children from different social class backgrounds (Aguinis, 1995).

Second, future work should examine whether the treatment effects are generalizable across different school districts and are cumulative across multiple school years and summers. A larger experiment involving multiple schools district would allow us to assess the external validity of the results by exploring treatment effects in sites with diverse social and economic conditions and local education policies. The present study was undertaken in a school district with strong support from the assistant superintendent and school principals. It is unclear whether the intervention could be effectively implemented in multiple districts where administrative support may vary across sites. Moreover, since this study examined changes in reading achievement over a single summer, a longitudinal study is needed to examine long-term impacts.

Third, information on the costs and effects of different types of summer programs is needed given the range of options available to policymakers and practitioners. As noted earlier, voluntary summer programs differ along several dimensions, including costs, duration, and the setting where reading instruction and activity occurs. Although our intervention is inexpensive relative to other summer programs, we cannot conclude that the intervention is cost-effective until there is more information on the costs and effects of different policy options.²³ Thus, a future evaluation on the cost-effectiveness of different summer programs would be useful for policymakers (Levin & McEwan, 2001).

Fourth, randomized experiments have been criticized for providing little information on the causal mechanisms that produce outcome differences between treatment and control groups

(Maxwell, 2004). Survey data from this study were too limited to provide detailed information on the processes that ultimately led to better reading outcomes among less-skilled readers, children with fewer books at home, and children from minority backgrounds. Unfortunately, we did not have sufficient resources to collect information from multiple sources of data including self-reported surveys, interviews with parents and students, and observations of family literacy activities. Better measures of social settings such as children's home are needed to reveal the quality of the social interactions within families when children are given free books to read at home and encouraged to read with their family members. Ethnographies of children's home environment (Chin & Phillips, 2004; Lareau, 1989) have shown that minority and low-income parents often lack the knowledge to help their children select appropriately challenging and interesting books. Such insights, gained through ethnographic research, could shed light on the etiology of summer reading loss and how it varies across families from different social class and ethnic backgrounds. Future research should include an experimental design and ethnographic data on the summer reading experiences of students from various ethnic and social backgrounds.

Conclusion

The main findings motivate cautious optimism regarding current efforts to address academic achievement disparities. Caution is needed until the current findings are replicated in a study that includes a larger sample of districts and schools and employs a richer set of reading measures. Optimism, however, may also be justified given the promising findings for lower-performing students. Encouraging voluntary reading during summer vacation may be one useful strategy for helping struggling readers acquire the skills needed to succeed in school.

TABLE 1

Characteristics of Students at the Beginning of the Study (N = 552)

Variable	%	Min	Max	<i>M</i>	<i>SD</i>
Female	0.47				
White	0.33				
Black	0.19				
Latino	0.26				
Asian	0.17				
Other	0.05				
Free-Reduced Lunch	0.39				
Limited English Proficiency	0.38				
Title I School	0.26				
Age (Months)		108	140	123.45	4.74
Iowa Test of Basic Skills (DSS)		142	263	202.78	24.08
Iowa Test of Basic Skills (NPR)		1	99	51.97	28.08
Oral Reading Fluency (WCPM)		6	242	120.27	37.83
Elementary Reading Attitude Survey (Total)		23.00	80	58.45	11.12

Note: DSS = Developmental Standard Score, NPR = National Percentile Rank,

WCPM = Words Correctly Read Per Minute.

TABLE 2

Relationship Between Student Ethnicity and Pretest Measures of Silent Reading Ability, Oral Reading Fluency, and Ownership of Books

Characteristic of Individual Students	White	Black	Latino	Asian
1. Spring Iowa Test of Basic Skills (total reading) scores above median?				
Yes	73%	38%	29%	42%
No	27%	62%	71%	58%
$\chi^2 = 73.182$ ($df = 3$), $p < .001$				
2. Spring reading fluency (Words Read Correctly Per Minute) above median?				
Yes	65%	39%	33%	46%
No	35%	61%	67%	54%
$\chi^2 = 36.856$ ($df = 3$), $p < .001$				
3. About how many books are in your home? (more than 100)				
Yes	58%	25%	25%	32%
No	42%	75%	75%	68%
$\chi^2 = 44.761$ ($df = 3$), $p < .001$				
4. About how many books for kids do you have in your home? (more than 50 books for kids)				
Yes	61%	30%	33%	35%
No	39%	70%	67%	65%
$\chi^2 = 36.458$ ($df = 3$), $p < .001$				

TABLE 3

Comparison of Baseline Achievement and Demographic Characteristics by Condition

Variable	Treatment		Control		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
White	0.35	0.48	0.32	0.47	0.72	550	0.471
Black	0.19	0.39	0.19	0.40	-0.14	550	0.890
Latino	0.26	0.44	0.26	0.44	-0.01	550	0.995
Asian	0.17	0.37	0.18	0.39	-0.46	550	0.647
Other	0.04	0.20	0.05	0.22	-0.51	550	0.607
Title I School	0.26	0.44	0.26	0.44	0.09	550	0.929
Age (Months)	123.57	4.89	123.32	4.58	0.62	550	0.538
Female	0.45	0.50	0.49	0.50	-0.90	550	0.367
Limited English Proficiency	0.56	0.50	0.55	0.50	0.20	550	0.840
Free-Reduced Lunch	0.37	0.48	0.43	0.50	-1.46	550	0.145
Iowa Test of Basic Skills (DSS)	203.24	24.10	202.30	24.09	0.46	550	0.647
Oral Reading Fluency (WCPM)	120.29	39.89	120.25	35.60	0.01	527	0.990
Elementary Reading Attitude Survey	58.43	11.39	58.46	10.86	-0.03	542	0.974

Note: Scores on the fluency assessment and total attitude score do not add up to 552 due to missing data.

TABLE 4

Ordinary Least Squares Models Predicting Treatment Effect on ITBS (Total Reading Scores)

Variables	All	White	Black	Latino	Asian
	B (SE)				
Treatment	.08~ (.04)	.11 (.09)	.22* (.09)	.14~ (.08)	-0.17 (.11)
Spring ITBS	.87** (.02)	.84** (.04)	.83** (.05)	.77** (.05)	.88** (.07)
(Constant)	-0.07 (.05)	-0.03 (.10)	-.17~ (.09)	-0.12 (.09)	0.07 -0.13
R ²	0.76	0.71	0.76	0.69	0.71
N	486	160	93	125	85

Note: All models include fixed effects for the randomization block. Standard errors in parentheses.

The model for "other ethnic students" (21 multiethnic, 2 Native American) revealed non-significant treatment effects.

~p<.10, *p<.05, **p<.01

TABLE 5

Ordinary Least Squares Models Predicting Treatment Effect on Oral Reading Fluency (WCPM)

Variables	All	White	Black	Latino	Asian
	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
Treatment	-2.09 (1.50)	-2.83 (2.73)	-1.79 (3.31)	-2.21 (2.81)	-.41 (3.95)
Spring-WCPM	.83*** (0.02)	0.86*** (.04)	0.83*** (.05)	0.77*** (.05)	0.77*** (0.05)
(Constant)	11.08*** -3.07	7.24 -5.51	12.21~ -6.72	17.00** -6.37	20.14* (8.54)
R ²	0.80	0.80	0.81	0.73	0.75
N	450	150	85	116	80

Note: All models include fixed effects for the randomization block. Standard errors in parentheses.

Sample sizes for OLS models predicting fluency are not equal to the ITBS analysis due to missing data on the fall fluency assessment.

~p<.10, *p<.05, **p<.01

TABLE 6

Ordinary Least Squares Models Predicting Treatment Effects for Subgroups of Students Based on Pretest Measures of Silent Reading Ability, Oral Reading Fluency, and Ownership of Books

Characteristic of Students	<i>B</i>	<i>SE</i>	<i>Lower Bound</i>	<i>Upper Bound</i>	<i>t</i>	<i>p</i>
1. Spring Iowa Test of Basic Skills (total reading) scores above median?						
Yes (n= 246)	0.064	0.066	-0.065	0.193	0.982	0.327
No (n = 240)	0.104	0.058	-0.010	0.218	1.805	0.072
2. Spring reading fluency (Words Read Correctly Per Minute) above median)?						
Yes (n = 236)	0.023	0.063	-1.000	0.146	0.367	0.714
No (n = 232)	0.167	0.063	0.044	0.290	2.669	0.008
3. About how many books are in your home? (more than 100)						
Yes (n = 186)	0.057	0.078	-0.097	0.206	0.728	0.467
No (n = 294)	0.097	0.051	-0.004	0.198	1.887	0.060
4. About how many books for kids do you have in your home? (more than 50 books for kids)						
Yes (n = 206)	0.024	0.072	-0.117	0.166	0.339	0.735
No (n = 272)	0.131	0.056	0.021	0.242	2.348	0.020

Note: Samples are unequal due to missing data from the fall survey.

References

- Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, 21, 1141-1158.
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23, 171-191.
- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23(3), 285-303.
- Baker, L. (2003). The role of parents in motivating struggling readers. *Reading and Writing Quarterly*, 19, 87-106.
- Behavioral Research & Teaching. (2005). *Oral reading fluency: 90 years of measurement*. Eugene, OR: Author.
- Blackburn, M. (2003). Boys and literacies: What difference does gender make? *Reading Research Quarterly*, 38, 276-287.
- Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the teach baltimore randomized field trial. *Educational Evaluation and Policy Analysis*, 28, 25-48.
- Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005). Success for all: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, 27, 1-22.
- Burkham, 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.
- Burley, J. E. (1980). Short-term, high intensity reading practice methods for upward bound students: An appraisal. *Negro Educational Review*, 31, 156-161.
- California Department of Education. (1999). *Reading/language arts framework for California public schools, kindergarten through grade twelve*. Sacramento, CA: California Department of Education.
- Carver, R. P., & Leibert, R. E. (1995). The effect of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly*, 30, 26-48.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Chall, J. S. (2000). *The academic achievement challenge: What really works in the classroom?* New York: Guilford Press.
- Chaplin, D., & Capizzano, J. (2006). *Impacts of a summer learning program: A random assignment study of Building Educated Leaders for Life (BELL)*. Washington, DC: The Urban Institute.
- Chin, T., & Phillips, M. (2004). Social reproductive and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of Education*, 77, 185-210.
- Cline, R. K. J., & Kretke, G. L. (1980). An evaluation of a long-term SSR in the junior high school. *Journal of Reading*, 23, 503-506.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Collins, C. (1980). Sustained silent reading periods: Effects on teachers' behaviors and students' achievement. *Elementary School Journal*, 81, 108-114.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Connecticut State Department of Education. (2001). *Connecticut reads 2001, governor's summer reading challenge*. Hartford, CT.
- Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). *Making the most of summer school: A meta-analytic and narrative review*. Washington, DC: Society for Research in Child Development.
- 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.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, Spring/Summer*, 8-15.
- D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of reading recovery in the united states. *Educational Evaluation and Policy Analysis, 26*, 23-39.
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006-469)*. Washington, DC: U. S. Department of Education. Institute of Education Sciences, National Center for Education Statistics.
- Davis, Z. T. (1988). A comparison of the effectiveness of sustained silent reading and directed reading activity on students' reading achievement. *The High School Journal, 72*, 46-48.
- Donahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). *The nation's report card: Fourth-grade reading 2000, NCES 2001-499*. Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
- Downey, D. B., von Hippel, P. T., & Broh, B. A. (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: Westview Press.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2000). Summer learning and home environment. In R. D. Kahlenberg (Ed.), *A notion at risk* (pp. 9-30). New York: The Century Foundation Press.
- Ferguson, R. F. (2001). Test-score trends along racial lines, 1971 to 1996: Popular culture and community academic standards. In N. J. Smelser, W. J. Wilson & F. Mitchell (Eds.), *American becoming: Racial trends and their consequences* (Vol. 1, pp. 348-390). Washington, D.C.: National Academy Press.
- Ferguson, R. F. (2002). *What doesn't meet the eye: Understanding and addressing racial disparities in high-achieving suburban schools*. Naperville, IL: North Central Regional Educational Laboratory.
- Filgreen, J. (2000). *The SSR handbook: How to organize and maintain a sustained silent reading program*. Portsmouth, NJ: Heinemann.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 241-258.
- Galda, L., Ash, G. E., & Cullinan, B. E. (2000). Children's literature. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research, vol. 3* (pp. 361-380). Mahwah, NJ: Lawrence Erlbaum Associates.

- Good, R. H., & Kaminski, R. A. (2003). *Dynamic indicators of basic early literacy skills, 6th edition*. Eugene, OR: Institute for the Development of Educational Achievement, University of Oregon.
- Heyns, B. (1978). *Summer learning and the effects of schooling*. New York: Academic Press, Inc.
- Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development, 58*, 1151-1160.
- Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics, 1*, 69-88.
- Holt, S. B., & O' Tuel, F. S. (1989). The effect of sustained silent reading and writing on achievement and attitudes of seventh and eighth grade students reading two years below grade level. *Reading Improvement, 26*, 290-297.
- Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., et al. (2003). *The Iowa tests, guide to research and development*. Itasca, IL: Riverside Publishing.
- Ivey, G., & Broaddus, K. (2001). "just plain reading": A survey of what makes students want to read in middle school classrooms. *Reading Research Quarterly, 36*(4), 350-377.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- Klibanoff, L. S., & Haggart, S. A. (1981). *Summer growth and the effectiveness of summer school (rep. No. 8)*. Mountain View, CA: RMC Research Corporation.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.
- Langford, J. C., & Allen, E. G. (1983). The effects of u.S.S.R. On students' attitudes and achievement. *Reading Horizons, 23*, 194-200.
- Lareau, A. (1989). *Home advantage: Social class and parental intervention in elementary education*. New York: The Falmer Press.
- Lauer, P., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275-313.
- Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis (2nd edition)*. Thousand Oaks, CA: Sage Publications.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.
- Manning, G. L., & Manning, M. (1984). What models of recreational reading make a difference. *Reading World, 23*, 375-380.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*, 3-11.
- McClure, A. A., & Kristo, J. V. (2002). *Adventuring with books, a booklist for pre-k-grade 6*. Urbana, IL: National Council for Teachers of English.
- McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for educators. *The Reading Teacher, May*, 626-639.
- Monson, D. L., & Sebesta, S. (1991). Reading preferences. In J. Flood (Ed.), *Handbook of research on teaching the English language arts* (pp. 664-673). New York: Macmillan.

- Morrow, L. M., & Weinstein, C. S. (1986). Encouraging voluntary reading: The impact of a literature program on children's use of library centers. *Reading Research Quarterly, 21*, 330-346.
- Murnane, R. J. (1975). *The impact of school resources on the learning of inner-city children*. Boston: Ballinger Press.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- National Research Council (Ed.). (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.
- Paris, S. G., Pearson, P. D., Cervetti, G., Carpenter, R., Paris, A. H., DeGroot, J., et al. (2004). Assessing the effectiveness of summer reading programs. In G. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs* (pp. 122-161). Mahwah, NJ: Lawrence Erlbaum Associates.
- Peak, J., & Dewalt, M. W. (1994). Reading achievement: Effects of computerized reading management and enrichment. *ERS Spectrum, 12*, 31-34.
- Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics (NCES 2005-464)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 229-272). Washington, D.C.: Brookings Institution Press.
- RAND Reading Study Group. (2002). *Reading for understanding*. Santa Monica, CA: RAND.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 181-221.
- Shanahan, T. (2004). Critiques of the national reading panel report: Their implications for research, policy, and practice. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 235-266). Baltimore: Paul H. Brookes Publishing Co.
- Stahl, S. A. (2004). What do we know about fluency? Findings of the national reading panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 187-212). Baltimore: Paul H. Brookes Publishing Co.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford Press.
- Sturm, B. W. (2003). The information and reading preferences of north carolina children. *School Library Media Research* Retrieved April 20, 2005, from <http://www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume62003/readingpreferences.htm>
- Summers, E. G., & Lukasevich, A. (1983). Reading preferences of intermediate-grade children in relation to sex, community, and maturation (grade level): A canadian perspective. *Reading Research Quarterly, 18*, 347-360.
- Summers, E. G., & McClelland, J. V. (1982). A field-based evaluation of sustained silent reading (SSR) in intermediate grades. *Alberta Journal of Educational Research, 28*, 100-112.

- Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement (research report 109)*. Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Topping, K. J. (1987). Paired reading: A powerful technique for parent use. *The Reading Teacher*, 40, 608-614.
- U. S. Department of Education. (1998). *Summer activities of students enrolled in grades 1-12*. Washington, DC: U. S. Department of Education, National Center for Education Statistics.
- U. S. Department of Education. (2001). *Assessing the lexile framework: Results of a panel meeting, NCES 2001-08*. Washington, DC: U. S. Department of Education, National Center for Education Statistics.
- U. S. Department of Education. (2003a). No child left behind summer reading achievers pilot program. Retrieved July 7, 2003, from <http://www.ed.gov/inits/summerreading/index.html>
- U. S. Department of Education. (2003b). *Supplemental education services, non-regulatory guidance (final draft)*. Washington, DC: Author.
- U. S. Department of Education. (2005). *The condition of education 2005 (NCES 2005-94)*. Washington, DC: National Center for Education Statistics, U. S. Government Printing Office.
- Vollands, S. R., Topping, K. J., & Evans, R. M. (1999). Computerized self-assessment of reading comprehension with the accelerated reader: Action research. *Reading and Writing Quarterly*, 15, 197-211.

Notes

¹ I wish to thank editors Ellen Goldring and Kenneth Wong and three anonymous reviewers for providing excellent feedback on earlier drafts of this manuscript.

² According to analyses of the Early Childhood Longitudinal Survey, Kindergarten Cohort of 1998 (ECLS-K), virtually all children do well on basic reading tasks, including the ability to recognize letters, beginning and end sounds, and sight words. However, there are significant racial/ethnic disparities related to more difficult reading comprehension items (e.g., literal inference, extrapolation, and evaluation). See Fryer and Levitt (2005), Table 7.

³ Pseudonyms are used to maintain the anonymity of the school district.

⁴ There are over 100 elementary schools Lake County and schools are organized into smaller administrative and geographic sub-units, each with its own assistant superintendent. The assistant superintendent of one sub-unit agreed to be the official research sponsor and provided administrative support to carry out the experiment. This sub-unit had both Title I and non-Title schools. The non-Title I schools were ethnically diverse and had sufficiently large Black and Latino enrollments whose achievement counted for AYP calculations. We decided not to target only Title I schools because they were located in 7 different sub-units and managed by 7 different assistant superintendents. In short, it was administratively easier to conduct the study in 1 sub-unit. Thus, the decision to target the one sub-unit with both Title I schools and racially and ethnically diverse non-Title I schools facilitated the administration of the experiment and the analyses of subgroup analyses.

⁵ Due to administrative and financial costs, we decided not to assign schools at random to experimental and treatment conditions. Instead, we randomly assigned students to conditions. Based on Cooper et al.'s (2000) meta-analysis of summer school programs and a previous pilot experiment, we anticipated a small effect size between .10 and .20 standard deviations. Using this information, we conducted a power analysis by assuming an effect size of between .10 and .20, and a pretest that was highly correlated with the posttest ($r = .80$). We estimated that approximately 500 students would be needed to have an 80% chance of finding a significant ($\alpha = .05$) difference between the treatment and control group on a posttest reading measure.

⁶ All 4 Title I schools administered a schoolwide program, which, under the Elementary and Secondary Education Act, permits these schools to use their federal dollars to support programs serving all students regardless of their income status.

⁷ Relative to the public school enrollment (K-12) in the United States in fall 2003, the study sample has a slightly larger percentage of minority students. Figures for the nation are as follows: White students = 58.3%, Black students = 16.1%, Latino students = 18.6%, and other = 7.0%. See table 4.1 in the Condition of Education 2005 (U. S. Department of Education, 2005).

⁸ Information on summer school attendance was obtained through the fall reading survey, which asked whether they attended a summer school program in a public school. About 17% of the sample reported attending summer school, which is similar to figures on summer school attendance in nationally representative surveys of elementary and secondary school students (Burkham *et al.*, 2004; U. S. Department of Education, 1998).

⁹ As noted in the manual for administration for the DIBELS oral fluency assessment (Good & Kaminski, 2003), all words omitted, substituted, and hesitations of more than 3 seconds were scored as errors while self-corrections within 3 seconds were scored as correct. Each examiner adhered to scripted directions, stating: "Please read this (point to passage) out loud. If you get stuck, I will tell you the word so you can keep reading. When I say "stop," I may ask you to tell me about what you read, so do your best reading. Start here (point to the first word of the passage). Begin." After the passage was read, students were asked: "Please tell me all about what you just read. Try to tell me everything you can. Begin."

¹⁰ For example, Summers and Lukasevich (1983) conducted a literature review of preference surveys and found 5-30 common themes in surveys measuring the reading interests and preferences of intermediate grade students. The 14 categories from their survey are similar to those in the NCTE list and include: adventure, history/geography, animals, children/family, sports, humor, science, poetry, fantasy, travel, romance, nature study, and mystery.

¹¹ For the reading preference, the following categories were used: (1) science nonfiction (earth science, space, technology), (2) prehistoric life, dinosaurs, (3) animal nonfiction (zoology, mammals, marine life, reptiles, et.), (4) historical nonfiction and biography, (5) historical fiction, (6) African and African American stories (real and realistic fiction), (7) Asian and Asian American stories (real and realistic fiction), (8) Latino and Latino Americans (Spanish speaking) and stories (real and realistic fiction), (9) Native American stories (real and realistic fiction), (10) family and everyday life stories, (11) school life stories, (12) struggle and survival stories (realistic adventure, heroes, heroines, etc.), (13) poetry, (14) sports biographies, (15) sports fiction, (16) animal fantasy, (17) science fiction fantasy, (18) time travel fantasy, (19) heroic and adventure fantasy, (20) supernatural tales and fantasy (ghosts, magic, and monsters), (21) realistic animal stories, (22) mystery stories, (23) stories of other girls my age (real and realistic

fiction), (24) stories of other boys my age (real and realistic fiction), (25) traditional literature (fables, myths and legends, tall tales, etc..).

¹² We decided not to use a book fair to match books to readers for two reasons. First, principals and teachers wanted to implement an administratively easy and simple procedure for matching books to children's reading preferences. Educators at school sites thought a book fair would be too burdensome to manage given existing administrative tasks associated with the end of the school year. Second, we wanted to employ a uniform procedure for ensuring that books were within each student's independent reading level. Prior research indicated that allowing children to self-select texts would produce poor matches between texts and readers.

¹³ The computer program (Python) was used to create two input: (1) student information on reading preferences and reading levels (Lexile level), and (2) book information on reading categories and reading levels (Lexile level). For each student, the student file contains the Survey ID, lexile range determined by the ITBS test, and the results of the reading preference survey. The book file contains a list of the 240 titles available to match to each student, along with the lexile level, reading level, and the list of categories. The output is a list of students matched to a specific book and a score for the match between the student and the book. A high score indicates a good match between child and book, and books with the 8 highest scores were matched to each child.

¹⁴ We retained all components with an eigenvalue over 1.00. The first component had an eigenvalue of 2.4 (48% of variance) and loadings ranged from .87, .73, .86 (items 1, 2, 3), and -.03, and .33 (items 4, 5). Thus, a student who scored high on this first component would report reading frequently for fun at home, reading books or stories at bedtime, and reading books in general. Moreover, a student would have to report reading infrequently with their parents and reading aloud to a family member. The eigenvalue for the second principal component was 1.1 (22% of variance) with loadings of .14, .06, and .17 (items 1, 2, 3) and .89 and .72 (items 4, 5). A student who scored high on the second component would report less independent reading of books and more oral reading of text with parents or family members.

¹⁵ The family letter was drafted by teachers and principals: Dear Parent (Family Member), Please encourage your child to read this book and complete the postcard. It will help your child if he or she reads out loud to you, or to an older brother or sister. After you listen to your child reading out loud a second time, tell him or her how they improved. There is also a place for your signature. Please sign the postcard indicating that you listened to your child read a part of the book. The postcard does not require a stamp; all you need to do is put it in the mail. It is important to return the postcard even if your child has not finished the book. After the postcard has been returned, you may certainly encourage your child to finish the book, read it again, or re-read favorite parts of it. The information on the postcard will help us understand the results and improve the program next year. Thank you for your time and effort towards making your child's summer reading a successful experience.

¹⁶ All students were told that they would receive books and postcards either in the summer or fall. Teachers followed a uniform script in which they explained that, regardless of when they received the books, children should read their books and then follow the direction for the completing the postcards and reading activities.

¹⁷ While there were no significant differences between the treatment and control group, the magnitude of the differences on the overall literacy habits score is suggestive of the idea that children in the experimental condition engaged in more literacy-related activities than control

students. With the exception of Asian students, average scores on the 4-point literacy habits measure are higher, on average, for the treatment group than the control group. In particular, the mean score on the literacy habits survey is at or above 2.3 for treatment students in each of the four subgroups of students. Since a score of 2.3 corresponds to a self-reported behavior that occurred between “once or twice a month” (mean score = 2.0) and “once or twice a week” (mean score = 3.0), this figure implies that students in the treatment group engaged in literacy related activities approximately once a week (i.e., read for fun, read books, read at bedtime, read with parents, and read aloud with a family member).

¹⁸ As noted by Burkham et al. (2004) in their analysis of summer learning using ECLS-K data, “this study did not investigate children’s progress in school, but instead investigated academic learning when children were out of school during the summer months. Thus, the nested nature of research questions that is typical of school-effects studies is not applicable here. Instead of the multilevel methods that are typically required for school-effects studies (i.e., hierarchical linear models, or HLM), we used the major analytic method appropriate for multivariate analyses of individual children: ordinary least-squares (OLS) regression” (p. 9).

¹⁹ Among Asian students, the control group included a significantly larger percentage of females than males. In general, females are more likely to read more on their own and do better on reading tests than males (Blackburn, 2003; Donahue et al., 2001). In our sample of Asian students, the z-score for females was .08 compared to -.19 for males on the spring ITBS.

²⁰ To estimate the main effect of the treatment and to examine interactions between the treatment and student background characteristics, we used an analysis of covariance (ANCOVA) with two- and three-way interactions to replicate the results from the OLS regressions for all students and for each ethnic subgroup. In particular, the ANCOVA was based on the same 463 students (White = 160, Black = 93, Latino = 125, Asian = 85) used in the OLS regressions and included both two-way (treatment and student ethnicity; treatment and income status) and three way interactions (treatment, ethnicity, income status).

²¹ There were no significant two-way interaction between the treatment and whether or not a student reported owning 50 or more kids books, $F(1, 451) = .72, p = .395$, or between the treatment and whether or not a student reported owning 100 or more books in general, $F(1, 452) = .15, p = .70$.

²² The BELL summer program provides a range of academic and enrichment services to students in grades 1 to 7 in Boston, New York, and Washington, DC. The curriculum also includes phonics instruction aligned with recommendations from the National Reading Panel and wide reading of multicultural literature. The Teach Baltimore program serves low-income children in the early grades and includes phonics-based instruction, read alouds, and silent reading practice. Although these programs focus on improving a broad range of youth outcomes, they are similar to the current intervention in three ways: (1) there is a strong emphasis on improving reading skills, (2) participation is voluntary, and (3) evaluations have been subjected to a randomized experiment. A recent evaluation of the BELL summer program revealed effect sizes ranging from .08 to .14 (Gates-MacGinitie reading test), and a three-year longitudinal evaluation of the Teach Baltimore reported effect sizes near .30 (Comprehensive Test of Basic Skills/4th Edition) for participating students who had above average attendance rates. The effect sizes from the current study ranged from .10 to .20 on a standardized reading test (Iowa Test of Basic Skills) and are similar to recent evaluations of more intensive voluntary summer programs involving elementary school students.

²³ The present intervention cost less than \$100 per student, which includes the price of books, postage, and labor. An experimental study testing the costs and effects of different summer program would shed light on whether a voluntary summer program is cost-effective relative to more intensive summer school programs.