

Replicating the Effects of a Teacher-Scaffolded Voluntary Summer Reading Program:
The Role of Poverty

Thomas G. White¹

The Center for Advanced Study of Teaching and Learning
University of Virginia, Charlottesville, VA, USA

James S. Kim, Helen Chen Kingston, Lisa Foster
Harvard University, Graduate School of Education
Cambridge, MA, USA

This study was made possible by an Investing in Innovation Fund (i3) grant from the U.S. Department of Education (PR/Award # U396B100195); however, the contents of this article do not represent the policy of the U.S. Department.

¹Correspondence concerning this article should be addressed to Thomas G. White, The Center for Advanced Study of Teaching and Learning, University of Virginia, 2200 Old Ivy Road, Charlottesville, VA, 22903. Email: tgw7u@virginia.edu

Abstract

A randomized trial involving 19 elementary schools (K-5) was conducted to replicate and extend two previous experimental studies of the effects of a voluntary summer reading program that provided (a) books matched to students' reading levels and interests and (b) teacher scaffolding in the form of end-of year comprehension lessons. Matched schools were randomly assigned to implement one of two lesson types. Within schools, students were randomly assigned to a control condition or one of two treatment conditions: a basic treatment condition replicating procedures used in the previous studies or an enhanced treatment condition that added teacher calls in the summer. During summer vacation, children in the treatment conditions received two lesson books and eight books matched to their reading level and interests. Overall, there were no significant treatment effects, and treatment effects did not differ across lesson type. However, there was a significant interaction between the treatment conditions and poverty measured at the school level. The effects of the treatments were positive for high poverty schools ($d = + .08$ and $+ .11$), defined as schools where 75% to 100% of the students were receiving free or reduced-price lunches (FRL). For moderate poverty schools (45%-74% FRL), the effects of the treatments were negative ($d = - .11$ and $- .12$). The results underscore the importance of looking at patterns of treatment effects across different contexts, settings, and populations.

Keywords: *comprehension, socioeconomic factors, research methodology, childhood*

Replicating the Effects of a Teacher-Scaffolded Voluntary Summer Reading Program:
The Role of Poverty

During the summer months, low-income elementary students lose ground in reading relative to their middle- and high-income counterparts (Alexander, Entwisle, & Olson, 2001; Burkam, Ready, Lee, & LoGerfo, 2004; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Heyns, 1978; Kim, 2004; Phillips & Chin, 2004). On average, summer vacation creates a 3-month gap in reading achievement between low- and middle-income children (Cooper et al., 1996). Even small differences in summer learning can accumulate across the elementary years resulting in a large achievement gap by the time students enter high school (Alexander, Entwisle, & Olson, 2007).

Socioeconomic differences in reading growth rates are larger in the summer months than in the school year (Benson & Borman, 2010; Cheadle, 2008; Downey, von Hippel, & Broh, 2004; LoGerfo, Nichols, & Reardon, 2006; McCoach, O'Connell, Reis, & Levitt, 2006). Yet in an era of shrinking tax bases and budget cuts, many large school districts do not offer any kind of summer programs to low-income students. McCombs et al. (2011) estimated the costs of six school-based summer learning programs and a home-based summer reading program. The school-based programs cost \$1109 to \$2801 per student, compared to \$245 per student for the home-based summer reading program. School-based summer learning programs improve reading achievement by .14 standard deviations on average (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). If home-based summer reading programs have roughly comparable positive effects, they are likely to be a much more cost-effective means of reducing or eliminating summer loss.

This article presents the results of a randomized control trial studying the effectiveness of a voluntary summer reading program that targeted 19 elementary schools with at least 45 percent low-income students (defined as students who are receiving free or reduced-price lunch). The program is called READS (Reading Enhances Achievement During Summer). The key features of READS are (a) providing summer books that are matched to students' reading levels and interests, and (b) providing teacher "scaffolding" and parent support for summer reading in the form of teacher lessons at the end of the school year and materials sent to students and parents in the summer (Kim & White, 2008; White & Kim, 2008, 2010).

The purpose of this study was to replicate and extend previous studies of READS. Our specific goals were to: (a) replicate, with a larger sample of schools, the positive effects of READS that were observed in two studies conducted previously in another state; and (b) extend previous studies by studying the effects of an enhanced version of READS that added teacher calls in the summer. In addition, we conducted exploratory analyses of the effects of poverty at the school level and student level and the effects of modified end-of-year lessons focusing on content-based predictions instead of comprehension strategies.

Theoretical and Empirical Support for READS

Figure 1 displays the logic model that underlies the current replication/extension study. In essence, we hypothesize that low-income students' fall reading achievement will improve if they engage in more voluntary summer reading, and that voluntary summer reading can be increased by providing access to matched and interesting books, teacher scaffolding, and parent support. The logic model rests on four pillars of evidence: (a) studies supporting theoretical explanations of why low-income students fall behind in the summer; (b) studies of the relationship between the amount of reading and growth in reading skills, including studies of leisure reading and

home-based summer reading; (c) studies indicating that students will make greater progress in reading during the summer if they read books that are well-matched to their individual reading levels and interests; and (d) experimental studies suggesting that students will make better progress in reading during the summer if they receive teacher scaffolding for summer book reading through end-of-year lessons and parent support of summer reading.

Why low-income students fall behind in the summer months. According to the “faucet theory” proposed by Entwisle, Alexander, and Olson (1997), all children gain when they are in school because the resources needed for learning are available to them. But when school is not in session, the resource faucet is turned off, and inequalities in resources exert their effects, causing children from low-income families to stop gaining or lose ground while children from middle- or high-income families maintain or improve their skills. Summer learning resources could differ in a variety of ways. Relative to middle-income children, low-income children may have fewer material resources (e.g., books) or psychological resources (e.g., family support for literacy) in the home, or they may have fewer material or psychological resources in the community in which they live (e.g., neighborhood libraries, learning from peers) (Duncan & Brooks-Gunn, 1997; Entwisle et al., 1997; Hart & Risley, 2003; Heyns, 1978; Lareau, 2003).

Research with nationally representative samples has shown that there is a strong relationship between family income and access to books and other reading materials in the home. Low-income families are far less likely than children from high-income families to have 10 or more books in their home (Bradley, Corwyn, McAdoo, & Garcia Coll, 2001). Lee and Burkam (2002) found that children in the bottom quintile of family socioeconomic status own an average of 38 books compared to an average of 108 books for children in the top quintile.

Family support for literacy is strongly associated with income. Burkam et al. (2004) found that low-income parents were significantly less likely to read a book to their child in the summer between kindergarten and first grade than high-income parents. A similar pattern was evident for taking the child to a library or bookstore. Bradley et al. (2001) found that high-income mothers were more likely than low-income mothers to read to their children three or more times per week. Phillips (2011) found income-based differences in time spent in literacy activities that were larger for children ages six and older than for three- to five-year-old children. Neuman and Celano (2006) found that in libraries in middle-income neighborhoods, adult caregivers frequently assisted preschool children in selecting challenging reading materials, whereas in libraries in low-income neighborhoods, children received little guidance from an adult.

There are qualitative as well as quantitative socioeconomic differences in family support for literacy. An ethnographic study of fourth-grade students' summer activities by Chin and Phillips (2004) found that parents of low-income children often went out of their way to obtain books for their children to use in the summer, but they were less skilled at organizing and facilitating literacy-related activities and making them appealing for their children, and they were less knowledgeable about their children's capabilities than middle-income parents. For example, a middle-income mother organized a book club for her daughter and her friends and their mothers, whereas a low-income mother bought her daughter \$45 worth of *Harry Potter* books that were too difficult for her to comprehend.

Regarding community resources, Neuman and Celano (2001) found that in low-income neighborhoods, fewer books were available in stores, childcare centers, and local elementary school and public libraries. Also, in low-income neighborhoods the available books were of lower quality.

Amount of reading and growth in reading skill. Access to print is positively related to various measures of reading skill (Lindsay, 2010). However, to the extent that it is causal, this relationship is likely to be mediated by time spent reading; that is, access to print is a necessary but not sufficient condition for improved reading skill. Time spent reading in school is positively correlated with reading achievement (National Institute of Child Health and Human Development [NICHD], 2000; Reutzel, Jones, & Newman, 2010; Stahl, 2004), as is reading outside of school (Anderson, Wilson, & Fielding, 1988). For students in Grade 1-12, overall print exposure as measured by title or author recognition checklists predicts oral language skills, word recognition, spelling, and reading comprehension, with correlations ranging from .36 to .45 (Mol & Bus, 2011).

Reading or print exposure is likely to influence growth in reading comprehension through such mechanisms as the development of orthographic knowledge and faster orthographic and phonological processing, application of letter-sound knowledge in a self-teaching process, growth in reading vocabulary from inferring the meaning of words in context, and knowledge acquisition enabling subsequent comprehension (Adams, 1990; Cunningham & Stanovich, 1997; Nagy, Anderson, & Herman, 1987; Share, 1999; Stanovich, 2000; Stanovich & Cunningham, 1993). In addition, more skilled readers who enjoy reading may choose to read more frequently, which causes their decoding and comprehension skills to grow at a faster rate than less skilled readers (Juel, 1988; Stanovich, 1986; Wigfield & Guthrie, 1997); hence, there is in all likelihood a reciprocal relationship between reading and reading skill (Mol & Bus, 2011).

In Heyns's (1978) landmark study of summer reading, hours spent reading and books read in the summer were significantly related to fall reading achievement with spring reading achievement, family income, parental education, and household size controlled. Several studies

have replicated Heyns's (1978) original findings (Burkam et al., 2004; Kim, 2004; Phillips & Chin, 2004). All of these investigators controlled for spring scores and family income and included a variety of additional covariates as controls (e.g., demographic characteristics, parents' expectations, teacher ratings, student attitude toward reading). Phillips and Chin (2004) found that students who read more than 30 minutes per day in the summer had higher reading comprehension scores in the fall. Burkam et al. (2004) found a significant relationship between fall reading achievement and a composite of seven literacy-related summer activities that included frequency of the student reading a book on his or her own and number of visits to a library or bookstore. Finally, Kim (2004) found a significant relationship between books read in the summer and fall reading comprehension scores.

Kim and Quinn (2013) conducted a meta-analysis of studies of K-8 summer reading interventions published since the Cooper et al. (2000) meta-analysis of summer programs. The results showed that home-based interventions were no less effective than school-based interventions, for total reading achievement (home-based $d = .12$ and school-based $d = .09$). Five of the 11 studies of home-based interventions employed experimental designs (Allington et al., 2010; Kim, 2006, 2007; Kim & Guryan, 2010; Kim & White, 2008;). These experimental studies, reviewed in the next two sections, are critical in establishing that the relationship between amount of voluntary summer reading and fall reading achievement is causal, because it may otherwise be interpreted as showing only that better readers read more in the summer.

Reading matched and interesting books. Two considerations are theoretically important in providing books for students' summer reading, text difficulty and interest. Regarding difficulty, efforts to assess students' reading levels and provide them with appropriately leveled texts have a long history dating back to the McGuffey readers (Cunningham et al., 2005; Fry, 2002). When

students are reading independently, especially if they are struggling readers, it is generally considered crucial to provide them with texts that are not too challenging so they do not become frustrated and stop reading (Allington & McGill-Franzen, 2013; Clay, 1985; Hiebert & Sailors, 2009). In addition, word reading accuracy, fluency, and comprehension improve when students read texts at their independent level (O'Connor, Swanson, & Geraghty, 2010; Shany & Biemiller, 1995). There is an equally strong consensus among reading scholars that to encourage voluntary reading outside of school, it is essential to provide students with books on topics that interest them (Allington & McGill-Franzen, 2013; Morrow, 2003). Research has shown that students are more likely to comprehend texts that they rate as interesting (Asher, 1979; Guthrie & Humenick, 2004). Also, students report that interesting texts are more enjoyable to read, and that they would like to continue reading those texts (Guthrie & Humenick, 2004).

Designers of voluntary summer reading programs face a significant practical challenge, however, in providing students with access to books that are both readable and interesting. The challenge arises from the fact that interest and difficulty can be at cross purposes for individual students.

Two approaches to providing students with books have been implemented and studied in randomized control trials of a voluntary summer reading program. In the first approach (Allington et al., 2010; Kim & Guryan, 2010), the researchers selected titles that were consistent with students' grade level and reading skills, then held book fairs at the end of the school year in which students self-selected the books they wanted to read in the summer.

Allington et al. (2010) randomly assigned over 1,000 first and second grade students in 17 high poverty schools to a treatment group or a control group. For three consecutive summers, students in the treatment group were invited to a book fair where they were asked to select 12

paperback trade books that they wanted to read. Selected books were delivered in individual boxes on the last day of school. The book fair included 400-500 fiction and non-fiction titles. The books represented a range of difficulty that targeted students' anticipated independent reading levels, where the independent reading level was defined as 99% word reading accuracy with good phrasing and expression. There were no end-of-year lessons or any other kind of support for summer reading. After the third summer, the treatment group scored significantly higher than the control group on the Florida Comprehensive Assessment Test (FCAT) ($d = .14$). In this study (as reported in Allington & McGill-Franzen, 2013), 100 randomly selected students were asked to review their book choices and read aloud a brief excerpt from five books. Struggling readers often selected books that they could not read accurately. Their choices seemed to be influenced by the cover of the book or their tendency to mimic the choices of better readers.

Kim and Guryan (2010) randomly assigned 370 low-income fourth grade students to a treatment group in which children received 10 self-selected books during summer vacation, a family literacy group in which children received 10 self-selected books and were invited with their parents to attend 3 summer literacy events, or a control group. Neither of the treatment groups performed better than the control group on the Gates–MacGinitie Reading Test. The lack of treatment effects may have been due in part to poorly matched books, because 67% of the students selected books that were above their independent reading level. As in the Allington et al. (2010) study, poor readers were especially prone to choosing books that were too difficult.

The second approach to providing books for a voluntary summer reading program is to assess students' reading levels and interests and use the data from these assessments to select a customized set of books for each student to read during the summer (Kim, 2006, 2007; Kim &

White, 2008). For example, Kim and White (2008) randomly assigned 400 children in grades 3 to 5 to one of four experimental conditions: control, matched books only, matched books with oral reading scaffolding, and matched books with oral reading and comprehension scaffolding. The Iowa Tests of Basic Skills (ITBS) reading comprehension test was administered as a pretest and posttest to measure gains over the summer. To assess student's interests, teachers administered a reading preference survey that asked students how much they enjoyed reading books from 25 categories. A two-step computer algorithm selected, from among 240 available titles for which a text difficulty (Lexile) rating was available, a set of 8 books for each student. The selected books matched both (a) the student's interests based on the reading preference survey, and (b) the student's independent reading level based on the comprehension pretest, where the independent reading level was defined as Lexile range (i.e., +50 Lexiles to -100 Lexiles around the observed pretest Lexile score). For students in the treatment groups, one matched book was mailed each week for eight successive weeks from early July until the end of August. Kim and White (2008) found that students in the books with oral reading and comprehension scaffolding condition made significantly larger comprehension gains than the control group ($d = .14$).

Allington et al.'s (2010) procedure for selecting books "privileges" self-selection of books over text difficulty (Allington & McGill-Franzen, 2013, p. 104), but there is no compelling evidence to suggest that self-selection is necessary. First, Guthrie and Humenick's (2004) meta-analysis of experimental studies of reading motivation found that providing students' with a choice of reading materials had significant positive effects on reading comprehension and achievement ($d = .95$), but the effect of interesting texts was considerably larger ($d = 1.64$). Thus if students can be provided with books that interest them, it may not be necessary to give them a

choice. Second, studies of voluntary reading in school also indicate that poor readers need guidance in text selection (Donovan, Smolkin, & Lomax, 2000; Juel, 1996; Reutzel et al., 2010). Finally, the comprehension gains made by students during the course of a single summer in the Kim and White (2008) study ($d = .14$) were identical to the gains made by students over three summers in Allington et al.'s (2010) study ($d = .14$). However, these studies varied on other dimensions, and there have been no direct comparisons of the two matching procedures, so the question of which approach is best for a voluntary summer reading program remains open.

Teacher scaffolding and parent support. Our logic model for READS (Figure 1) asserts that low-income students' volume of summer reading will increase and their fall reading achievement will improve if they receive teacher *scaffolding* in addition to having access to matched and interesting books. Following Graves and Graves (2003), we use the term "scaffolding" to refer to "a set of prereading, during reading, and postreading opportunities and experiences designed to assist a particular group of students in successfully reading, understanding, learning from, and enjoying a particular selection" (p.2), such as instructional procedures for teaching reading comprehension strategies and instructional frameworks that foster content learning (Clark & Graves, 2005).

Kim (2007) studied the effects of a voluntary summer reading program for first- through fifth-grade students. In the late spring, the students took the reading portion of the Stanford Achievement Test (SAT) as a pretest and also completed a 20-item survey of their reading preferences. Then they were randomly assigned to a control condition or a treatment condition in which they received 10 books for summer reading that were matched to their interests and reading levels using a procedure that was similar to Kim and White's (2008) book matching procedure. Teachers informed students that they would be receiving books over the summer

accompanied by postcards asking whether they read the book, whether they liked it, and whether it was easy to read. In the fall, students took the SAT reading test as a posttest and completed a survey that included questions about summer reading activity. Although students in the treatment group reported reading significantly more books over the summer than students in the control group, there was no difference between the groups on the reading comprehension posttest. Kim (2007) suggested that to strengthen the efficacy of summer reading programs, teachers could scaffold reading activities by instructing children in the use of reading comprehension strategies ([NICHD], 2000; Pressley, 2002; Rosenshine & Meister, 1994).

In the Kim and White (2008) study of READS, the effective treatment condition—matched books with oral reading and comprehension scaffolding—involved three teacher-directed lessons at the end of the school year. In these lessons, the teacher modeled fluent oral reading and five comprehension strategies (reread, predict, ask questions, make connections, and summarize). Students practiced using the comprehension strategies while reading silently on their own and practiced fluent oral reading in a paired reading format. In the summer, students received matched books with postcards asking them to write down the book title, check any of the five comprehension strategies they used to better understand the book, and indicate whether they finished the book and how many times they read it. The postcards also asked children to tell someone in their family what the book was about, and to select a 100-word passage from the book and read the passage aloud to a family member. Finally, the postcard included a request that the parent sign and return it with an optional comment. Thus there was some parent support as well as teacher scaffolding, although parent support was quite limited.

In a similar experiment by Kim (2006), students in the READS treatment condition received matched books in the summer and the same end-of-year lessons (teacher scaffolding) and

postcards (parent support). The difference between the treatment group and the control group on the reading comprehension posttest, $d = .08$, was comparable to the treatment-control difference in Kim and White (2008), although it fell short of statistical significance ($p = .059$).

Wilkins et al. (2012) conducted a large-scale, multi-district randomized control trial of the effects of a summer reading program that included minimal support for summer reading. Students in the treatment group ($n = 1,571$) received books that were matched using our procedure that takes both interest and difficulty into account, but there were no lessons at the end of the year. The results showed that while students in the treatment group did report reading more books, there were no significant treatment effects on the reading comprehension posttest given in the fall ($d = .02$). This finding is consistent with the null results of Kim (2007) and Kim and White's (2008) finding that students who received matched books only did not differ from students in the control group ($d = .02$).

Magnitude of the Effects in Relation to Summer Loss for Low-income Students

In the two experimental studies of a voluntary summer reading program that showed statistically significant positive effects on reading comprehension (Allington et al., 2010; Kim & White, 2008), the overall effect sizes were .14, but the effects were larger for students receiving free or reduced-price lunch, $d = .21$ (Allington et al., 2010) and $d = .28$ (Kim & White, 2008; see re-analysis in White & Kim, 2008). Kim (2006) also found larger effects for Black students ($d = .22$) Latino students ($d = .14$), less fluent readers ($d = .17$), and students who reported owning fewer than 50 children's books ($d = .13$).

Importantly, in Allington et al.'s (2010) study and in Kim and White's (2008) study, the effects on reading comprehension for low-income students were large enough to offset most or all of the summer loss shown by low-income students in Cooper et al.'s (1996) meta-analysis, $d =$

-.27 (Table 7, p. 257). Additionally, the effects were comparable to effects for school-based and home-based summer programs in meta-analyses by Cooper et al. (2000) and Kim and Quinn (2013). Effect sizes should be interpreted with reference to the related literature, not arbitrary benchmarks (Sun, Pan, & Wang, 2010). Thus summer voluntary reading interventions have shown effects that are practically important considering what typically occurs in the population of interest when there is no intervention and what is known about the impact of similar interventions.

Summary of the Literature

There is emerging evidence that it may be possible to reduce or eliminate summer loss for low-income students by implementing a summer voluntary reading program (Allington & McGill-Franzen, 2013; White & Kim, 2008, 2010). The available evidence also suggests that the effectiveness of summer voluntary reading programs may be enhanced by matching books to students' reading levels and interests, and effectiveness is likely to be enhanced by incorporating teacher scaffolding in the form of end-of-year lessons, materials accompanying books sent to students' homes during the summer, or activities that are aimed at increasing parent involvement in students' summer reading. However, more research is needed to: (a) replicate the positive effects of book matching and teacher scaffolding, and (b) develop and test new and better ways to support children's summer reading.

Rationale for the Current Study

Again, the main purpose of this study was to replicate and extend previous studies of READS.

Replication. We sought to replicate our previous studies with a different student population in a different geographic location and different instructional context. The students, teachers,

school communities, and district reading curricula were all different. In the Kim and White (2008) study and Kim (2006) study, READS was implemented in a large and ethnically diverse suburban school district in the mid-Atlantic region. For the 12 participating schools in those studies, the mean percentage of students receiving free or reduced-price lunch was 41 ($SD = 22$), and only one school (with a mean percentage of 88) met the 75 percent criterion for a “high poverty” school as defined by the National Center for Education Statistics (Aud et al., 2010). In the current study, the school mean percentage free or reduced-price lunch was 72 ($SD = 18$), and 10 of the 19 schools met the 75 percent criterion for high poverty. Although the student population and context in this study was different from our earlier studies, we maintained the two key features of READS, matched books and teacher scaffolding for summer reading.

Educators and education policymakers need evidence from multiple methodologically rigorous studies to make informed decisions about interventions including summer reading programs (Gersten et al., 2005; Institute of Education Sciences, 2013). The goal of a replication study is to determine whether effects observed in one setting, with particular teacher and student participants and a particular set of conditions, generalize to other settings with different participants and different conditions. In a typical replication study (e.g., Coyne et al., 2013), the investigator deliberately varies one or more conditions while key features of the intervention are retained. These types of investigations are called “systematic replication and extension” (Sidman, 1960) or “varied” replications (Van IJzendoorn, 1994). Replication studies play a critical role in strengthening the generalizability and external validity of experimental research.

Extension: Teacher calls. Considering the demonstrated benefits of teacher scaffolding in the form of end-of-year lessons, we reasoned that ongoing teacher support during the summer months could enhance comprehension outcomes. We thought that verbal encouragement from a

teacher might increase children's motivation to read, and that opportunities to orally recall text would support children's comprehension of matched summer books (Reed & Vaughn, 2012; Reese, Suggate, Long, & Schaughency, 2010). Accordingly, we developed a procedure in which teachers called children three times in the summer and prompted an oral recall of a matched book.

Exploration 1: Content-based prediction routine. In the matched books with oral reading and comprehension scaffolding treatment conditions in Kim (2006) and Kim and White (2008), the end-of-year lessons included both fluency practice and a *multiple strategy routine* in which teachers instructed children to use multiple comprehension strategies (making connections, predicting, asking questions, re-reading) with a narrative and informational text in three classroom lessons. In the current study, we sought to replicate these lesson effects while simultaneously exploring the effects of new end-of-year lessons incorporating what we call a *content-based prediction routine*. Half of the schools implemented lessons with fluency practice and a multiple strategy routine and half of the schools implemented lessons with a content-based prediction routine but no fluency practice. We did not include fluency practice in the lessons incorporating the content-based routine because Kim and White (2008) found that end-of-year lessons with fluency practice alone were not effective. However, we retained the fluency component in the lessons focusing on a multiple strategy routine.

In the content-based prediction routine, the end-of-year lessons were designed to focus children's attention on text-based content and to teach them how to apply the procedure before, during, and after reading a summer book. Different content-based routines were used for narrative and informational text. The content-based prediction routine with narrative text was based on a story impression activity using key words (McGinley & Denner, 1987). The content-

based prediction routine with informational text was based on a K-W-L activity using charts to scaffold comprehension of informational text before, during, and after reading (Carr & Ogle, 1987; Ogle, 1986; Williams et al., 2005). More detailed descriptions of these routines are provided in the Methods section.

Exploration 2: Effects of poverty at the student level. The faucet theory (Entwisle et al., 1997) suggests that low-income students lose ground in the summer relative to middle-income students because there are fewer resources available to them in the summer months, including fewer books in the home and family support for literacy. It follows that providing low-income students with books in the summer and encouraging them to read might prevent summer loss. In contrast, middle-income students may exhibit little or no summer loss, and they may already have enough books and family support. For these reasons, middle-income students might benefit less from receiving books, teacher scaffolding, and summer calls.

Exploration 3: Effects of poverty at the school level. At the student level, receipt of free or reduced-price lunch provides a measure of the material and psychological resources available in the home environment in the summer. At the school level, school percent receiving free or reduced-price lunch is a proxy measure for the material and psychological resources available in the community during the summer. Alternatively, school poverty may carry part of the effect of poverty that may be measured inadequately at the student level as the dichotomous variable, receiving or not receiving free or reduced-price lunch (ryk & Raudenbush, 1992; van Ewijk & Slegers, 2010). Either way, school-level poverty may moderate the effects of READS.

Research Questions and Hypotheses

We posed five research questions: a replication question, an extension question, and three exploratory questions.

Replication question: Were the effects of READS, a voluntary summer reading program providing matched books and teacher scaffolding, replicated in a new sample of schools? We expected positive effects of the basic READS intervention without teacher calls, regardless of whether the end-of-year lessons included fluency instruction and regardless of whether the lessons focused on a multiple strategy routine or a content-based prediction routine.

Extension question: Did the addition of teacher phone calls in the summer enhance the effects of the program on reading comprehension? We viewed these calls as providing more support for summer reading and predicted stronger effects through such mechanisms as motivating students to read more books or increasing student's understanding of, and engagement with the books they were reading. During the summer phone calls, teachers prompted children to produce an oral recall of text (i.e., "tell me about the book you are reading"), and to continue reading during the summer months. Our expectation was that the teacher phone call would encourage oral recall of texts and improve children's comprehension of books that were mailed home.

Exploratory question 1: Was the impact of READS on comprehension outcomes stronger in schools implementing the content-based prediction routine than in schools implementing the multiple strategy routine with fluency practice? We hypothesized that the content-based routine would be more effective, particularly for narrative texts because they were delivered with a book-specific (content-based) set of story impression words. For both the narrative and informational books that were delivered in the summer, students received a postcard that prompted a specific comprehension routine with a pre-reading and post-reading component

instead of multiple general strategies that they had to remember to apply when appropriate during reading.

Exploratory question 2: Was the impact of READS on comprehension outcomes stronger for students who are receiving free or reduced-price lunch than students who are not receiving free or reduced-price lunch? Differential effects based on family income have been reported previously (Allington et al., 2010; Kim & White, 2008).

Exploratory question 3: Was the impact of READS on comprehension outcomes greater in high poverty schools than moderate poverty schools, controlling for income status at the individual student level (i.e., receipt of free or reduced-price lunch)? Consistent with the criterion used by the National Center for Education Statistics (Aud et al., 2010), high poverty schools were schools where 75 to 100 percent of the students received free or reduced-price lunch and moderate poverty schools were schools where 45 to 74 percent of the students received free or reduced-price lunch.

Methods

Setting and Participants

This study was implemented in a mid-sized urban school district in North Carolina. In spring 2011, we successfully recruited 19 schools from among the 29 elementary schools (K-5) in the district. A total of 1,421 Grade 3 students from the 19 schools received parental consent to participate in this study and completed reading comprehension pretests in spring 2011. The overall consent rate was 81%. Low-income children (i.e., 72% were receiving free lunch), Black children (51%) and Hispanic children (30%) comprised a clear majority of the sample, and 29% of the children were identified as limited English proficient (LEP). Baseline scores on the Iowa

Tests of Basic Skills comprehension test ($M = 45$ percentile rank, $SD = 29$) indicate that the sample mean was below the national norm.

Design and Treatment Conditions

The design of the study was based on the two-stage random assignment procedure that is illustrated in Figure 2. This design enabled us to generate internally valid estimates of the effects on reading comprehension of: (a) the treatment conditions, poverty measured at the student level, and the interaction of student-level poverty and treatment conditions; (b) the lesson conditions and their interaction with treatment conditions; and (c) poverty measured at the school level, and the interaction of school-level poverty with treatment conditions.

We first formed matched pairs of schools on test scores from the North Carolina End of Grade (EOG) test in Grade 3 reading (primary criterion) and school size (secondary). Specifically, we used the average of the school's mean EOG score across the past three years and total enrollment to form the pairs. Two small schools were treated as one member a pair. Next, as the first stage of the randomization procedure, the members of each pair of schools were assigned at random to receive either multiple strategy lessons or content-based prediction lessons.

In stage two, within each of the 19 schools, third grade students were randomly assigned to one of three conditions: control, basic treatment with teacher lessons and summer books only (TL-SB), or the enhanced treatment with teacher lessons, summer books, and teacher phone calls (TL-SB-TC). As in our previous experimental studies, participating teachers were also randomly assigned to one of the three treatment conditions. A total of 81 teachers were assigned in equal numbers ($n = 27$) to the TL-SB condition, TL-SB-PC condition, or control condition.

The student and teacher random assignment procedures enabled us to eliminate any confounds related to student and teacher characteristics.

Students in the control condition participated in “business-as-usual” instruction and summer reading activities. The control teachers were asked to use their basal texts to provide typical reading instruction for their students, and students received no matched books during the summer.

For students in the basic treatment condition, teacher lessons and summer books (TL-SB), teachers implemented end-of-year lessons using either (a) a multiple strategy routine that replicated lessons from our earlier studies or (b) a content-based prediction routine schools using lessons that incorporated a story impression activity with narrative text and a K-W-L activity with informational text. Students received 10 books (and 10 postcards) in the mail each week of summer, including two lesson books and eight matched books. Students also received postcards that prompted the use of multiple comprehension strategies or content-based predictions. For students in the enhanced treatment condition, teacher lessons, summer books, and teacher phone calls (TL-SB-TC), in addition to participating in lessons and receiving summer books, students received up to three phone calls in which teachers prompted an oral recall of a mailed summer book.

Measures

Student demographic variables. We obtained student demographic data from district administrative files, including gender, ethnicity, the primary language spoken at home, income status (i.e., receipt of free or reduced-price lunch), and whether the child had limited English proficiency.

Iowa Tests of Basic Skills, Reading Comprehension (ITBS). The pretest and posttest measure was the ITBS reading comprehension test. Level 9, Form A was administered in spring 2011 when children were in third-grade, and Form C was administered in fall 2011. Teachers administered the reading tests following a training session based on the ITBS administration manuals. The ITBS is a 37-item comprehension test based including 20 items from narrative passages and 17 items from informational passages. The ITBS is a highly reliable assessment with reported KR-20 coefficients above .93 and equivalent form estimates of .86 or higher (Hoover et al., 2003). The total comprehension scores are vertically equated through Item Response Theory scaling to yield a continuous measure of reading ability, the Developmental Standard Score. The ITBS also yields a Lexile range that represents each child's independent reading level (for this sample $M = 592$, $SD = 188$, $Min = 110$, $Max = 1,100$).

Students' reading preferences. Students completed a reading preference survey in the spring. They used a 4-point scale (I don't like it, I like it, It's okay, I really like it) to rate their interest in reading books from 18 categories (e.g., adventure, biographies, historical fiction, science, sports). The 18 categories were drawn from published surveys of children's reading preferences and included the categories used in previous studies (Kim, 2006; Kim & White, 2008).

Independent book reading based on fall survey. When they returned to school in the fall, students completed a survey that asked them to report the number of books they had read in the summer and the number they had checked out from a public library. The scale for each item ranged from 0 to 10 books.

The fall survey also included an item to assess students' perceptions of the difficulty level of the books they read during summer. Students were asked to respond to a prompt ("The books I

read this summer were?") by choosing one of three options: (a) too hard, (b) just right, or (c) too easy. Eighty percent reported reading books that were "just right" during the summer.

Classroom lesson fidelity. To assess fidelity of implementation of the lessons, we developed an adherence checklist for each of the six lessons that focused on content-based prediction and each of the six lessons that focused on multiple strategies. The checklist included elements representing adherence to the script. For example, the checklist for the content lessons assessed whether teachers implemented the story impressions activity (e.g., teacher reads story impression words, encourages students to write a story guess), and the K-W-L activity (e.g., teacher asks students what they know about a topic, what they want to know, and what they learned). For the multiple strategy lessons, the checklist assessed whether teachers explained and modeled the use of four comprehension strategies. Two lessons from each treatment teacher were randomly selected to videotape and then coded for adherence to the lesson script. To assess inter-rater agreement, we double coded a random sample of 20% of the lessons. Reliability exceeded 80%. We also selected a random sample of 15% of the control teachers ($n = 12$) and reviewed a random sample of two videotaped lessons from each of these teachers.

Teacher phone calls. Teachers kept records of their phone calls for summer follow-up. We assessed implementation of the teacher call condition in terms of the number of calls each student received at home ($M = 1.12$, $SD = 1.05$, $Min = 0$, $Max = 3$) and the number of books students reported reading when they were successfully contacted ($M = 2.93$, $SD = 1.59$, $Min = 0$, $Max = 10$).

Total postcards returned for treatment students. We used postcard data to check implementation and assess children's reading engagement with their matched summer books. We coded whether or not a child returned a postcard for each matched book sent home, and

computed a mean postcard return for children in the content-based prediction and multiple strategy conditions. Among children in the treatment group, 55% returned at least 1 postcard.

Procedures

Book selection, matching, and distribution to students. To develop an initial book list for this study, the third author identified suitable texts from a Scholastic catalog for Grades 2-4. She began with titles that had been popular with students in previous studies, consulted with teachers and librarians, and paid particular attention to the third grade units from the district's Reading Street basal. The final list included 484 narrative and informational books. Narrative books were high-interest series books with a story grammar (Stein & Glenn, 1979) involving characters, settings, plots, problems, resolutions, and themes (e.g., Amber Brown, Captain Underpants, Goosebumps). Informational books were books about animals (e.g., Polar Bears, Penguins), or books on natural science topics (e.g., Awesome Ocean Records), or biographies (e.g., Duke Ellington, Roberto Clemente). All of the informational books included one or more of the following text structures: description, sequence, cause and effect, compare and contrast, or problem and solution (Meyer, 1985).

As in previous studies, we used a computer algorithm to identify a set of eight books matched to each student's interests and reading level. The algorithm merged data from two files. One file included a level and preference category for each of the 484 book titles, and the second file included each student's preference scores from the spring survey and Lexile scores from the spring administration of the ITBS.

Lesson books to be used in the end-of-year lessons were selected with several criteria in mind: high-interest, not commonly known and read by students, appropriate text difficulty, and richness and complexity of content to facilitate the teaching of multiple-strategy or content-based

comprehension routines. The narrative text chosen was *Look Out, Jeremy Bean!* (Schertle, 2011), and the two informational texts were *Polar Bears* (Gibbons, 2002) and *The Sweaty Book of Sweat* (Barnhill, 2010). The same three texts were used in both lesson conditions.

For children in the treatment groups, the lesson books, *Look Out, Jeremy Bean!*, and *Polar Bears* were mailed to students in the first two weeks of the summer. In each of the next eight weeks, one matched book was mailed to treatment group students. Every effort was made to ensure the delivery of all 10 books to each student during the summer. When a book package was returned as undeliverable, we attempted to obtain a correct address and resend it.

Teacher training. Teachers attended a 2-hour training session that differed by lesson type (multiple strategy or content-based prediction) and condition (treatment or control). The multiple strategy training was led by two veteran teachers from a Virginia school district who had developed the lessons and led training sessions in previous studies (Kim, 2006; Kim & White, 2008); the content-based prediction training was led by two veteran teachers from the participating North Carolina district, including an instructional coach and a National Board Certified teacher who had implemented a pilot version of the lessons; and the control group training was led by a staff member of Communities in Schools of North Carolina.

Teachers who were assigned to the control condition participated in a professional development seminar on classroom management. At the end of the session, control teachers were instructed to conduct "business as usual" reading lessons on the days when the treatment teachers were conducting multiple strategy or content-based comprehension lessons.

Teachers who were assigned to the treatment condition received a lesson plan for each of the six lessons, accompanying materials, and training to conduct one of the lesson types. The treatment group trainers began by explaining the goals and purpose of the lessons, walked

teachers through the lesson procedures, modeled the lessons, answered questions, and offered the teachers an opportunity to ask follow-up questions via email or phone up prior to and during the week when the lessons were scheduled. The stated purposes of the lessons were (a) to teach students to apply comprehension strategies to narrative and informational texts and recognize fluent oral reading (multiple strategies) or (b) to draw students' attention to structural differences between narrative and informational texts, prompt the use of a different routine with each type of text, and motivate engagement with text during the reading activities at school and home (content-based prediction).

End-of-year lessons. The multiple strategy lessons have been described in the introduction and previously published work (Kim & White, 2008; White & Kim, 2008). These lessons consisted of three lessons using the narrative text, *Look Out, Jeremy Bean!*, followed by two lessons using the informational texts, *Polar Bears* and *The Sweaty Book of Sweat*. The six lessons focused on making predictions, making connections, re-reading, and asking questions. The first 3 lessons also included a fluency component as described previously.

The content-based prediction lessons incorporated different comprehension routines for narrative and informational texts. In the first three lessons with narrative text, the routine was based on a story impression activity (McGinley & Denner, 1987). The lesson materials included key words selected from the story and listed in the order they appeared. In the pre-reading activity, the teacher read aloud the story impression words and phrases and prompted students to use the words and phrases to make a story guess based on common text structures such as character, setting, goal, problem, plot or action, resolution, or theme. The words and phrases were selected specifically to direct students' attention to such text structures and to frame their story guesses within these text structures. In the during-reading activity, the teacher read the

story aloud and asked literal and inferential text-based questions that helped students activate and integrate text-based ideas. In the post-reading activity, the teacher led a discussion comparing the story guess with the actual content of the story (National Assessment Governing Board, 2008; Shanahan et al., 2010). Thus the teacher provided scaffolding for the content routine by helping students perform each part of the routine prior to summer vacation (Graves & Graves, 2003; Rogoff, 1990).

In the next three lessons with informational text, the comprehension routine was based on a K-W-L activity. Before reading the text, teachers activated background knowledge by prompting students to ask, “What do I Know, and what do I Want to know?” Next, teachers read aloud from the text and helped students identify important text-based information needed to address the questions in the before reading activity. After reading the text, the teacher and students wrote down answers to the question, “What did I Learn?”

Postcards delivered with summer books. To support use of the routines at home, we delivered a postcard with each book that was mailed to students. For students who participated in multiple strategies lessons, there were two different postcards, one for narrative and one for informational texts. For narrative texts, the postcard asked the student to write the book title, indicate whether they finished the book and how many times they read it, and to check each comprehension strategy they used to better understand the book, from a list including re-reading, making predictions, asking questions, and making connections. The postcard also prompted the student to do the following: tell someone in their family what the book was about; select a 100-word excerpt from the book and to read aloud to a family member; read the excerpt a second time and ask the family member whether they read the text more smoothly, knew more words, and read with more expression; and ask for a signature from the family member. The postcard

for informational texts was identical, except that it prompted the student to tell a family member two things they learned from the book instead of simply telling them what the book was about.

For students who participated in content-based prediction lessons, we delivered, with each of their summer books, a postcard with book-specific prompts, either a story impression postcard (for narrative texts) or a K-W-L postcard (for informational texts). The story impression postcard prompted the child to read the book-specific story impression words, make a prediction about the story, read the book, and compare the prediction with the actual story after reading it. The K-W-L postcard prompted children to think about what knew about a topic before reading a text, articulate what new topics they wanted to learn about, and then to talk about new things that they learned with a family member.

Teacher phone calls. Teachers used a phone call script to call students. They were instructed to attempt to contact each student on their list three times during the summer months. During each call, the teacher asked students how many mailed books they had read, prompted an oral recall using the probe “tell me about the book you read,” encouraged continued reading of the books, and requested return of the postcards in the mail.

Results

Implementation Checks

To check implementation of the treatment components, we analyzed data from the lessons, teacher phone calls, and summer postcards.

Lessons. On average, treatment teachers adhered to 72% of the lesson script elements ($M = 72\%$, $SD = 16\%$). In addition, the mean adherence rates were statistically equivalent in schools implementing a multiple strategy routine ($M = 70\%$) and content-based prediction routine ($M = 74\%$), $F(1, 18) = .58$, $p = .46$, and also equivalent for moderate poverty schools ($M = 73\%$) and

high poverty schools ($M = 70\%$), $F(1, 18) = .37, p = .55$. In sum, there was no difference in teachers' adherence rates based on lesson condition or school poverty. Mean adherence rates were also similar to those reported in previous studies of comprehension instruction in the elementary grades (Guthrie et al., 2004; Williams et al., 2005). Review of videos of the control teacher lessons revealed a wide range of instructional activities including reading circles, social studies lessons, read alouds, and other common instructional methods used during reading blocs.

Teacher calls. Among the 484 students in the TL-SB-TC condition, 187 (39%) received 0 calls, 108 received 1 call (22%), and 134 (28%) received 2 calls. Just 55 students (11%) received the full dosage of 3 calls. For this reason, the effect of the TL-SB-TC condition should be viewed as an intention-to-treat analysis.

Postcards returned. We examined return data for students in the two treatment conditions (TL-SB and TL-SB-TC). Overall, 55% of the students in the treatment groups returned at least one postcard and 18% returned more than half of their postcards (5 to 10). These postcard return rates are consistent with our previous studies. As shown in Table 1, we also measured postcard returns for the 8 matched books, since the first 2 lesson books were not matched to each child's reading level and interest. For the 8 matched books, students in content-based prediction routine schools ($M = 3.24, SD = 2.58$) mailed one more postcard to school during the summer than students in the multiple strategy routine schools ($M = 2.14, SD = 2.36$), $d = .45, p < .001$. The advantage favoring content-based schools over multiple strategy schools was due to narrative books ($d = .53$), not informational books ($d = .01$).

Data Analysis Approach

To assess the comparability of the experimental conditions at pretest, we examined group equivalence on student-level measures. As shown in Table 2, there were no statistically

significant student-level differences on pretest measures of reading comprehension or in the percentage of low-income or minority students, or students with limited English proficiency. From pretest ($N = 1,421$) to posttest ($N = 1,189$), 16% of the sample was lost to attrition. Attrition rates for the control group (16%), the TL-SB group (17%), and the TL-SB-TC group (15%) were statistically equivalent $F(2, 1,418) = 1.09, p = .34$. Among children who were included in the analysis of treatment effects, there was no difference in pretest reading scores among children in the control group ($M = 184, SD = 25$), the TL-SB group ($M = 185, SD = 25$), and the TL-SB-TC group ($M = 184, SD = 24$), $F(2, 1186) = .33, p = .72$. We reduced attrition rates by administering posttests to children who moved from one of the 19 study schools to a non-study school within the district and by administering re-tests for children who were absent during the first round of posttests.

To address our first two research questions (replication and extension), we used ordinary least squares (OLS) regression models to estimate the main effects for the basic and enhanced treatment conditions at the student level. We standardized all pretest and posttest measures as z-scores in our models and included the pretest reading score as the covariate. Thus, the coefficient for the two student-level effects can be interpreted as the standardized covariate-adjusted posttest difference between groups.

To address the exploratory research questions 3 to 5, we included cross-level interactions involving the student-level treatment conditions and each of the school-level variables, including school poverty and lesson condition. In addressing questions 4 and 5, we blocked schools by poverty strata, including (a) high poverty schools (75% to 100% receiving free or reduced-price lunch, FRL), and (b) moderate poverty schools (45% to 74% FRL).

Effects on Reading Comprehension

Descriptive analyses of covariate adjusted posttests. Table 3 presents covariate-adjusted posttest means and standard deviations for the three student-level experimental conditions overall (for all schools) and by school lesson condition (for schools implementing a particular comprehension routine).

Exploratory question 1: Effects of lesson type and comprehension routine. Treatment effects (i.e., treatment-control differences) were not different for schools implementing multiple strategies or content-based prediction. This analysis is described in Appendix A (see Table A treatment by lesson condition interaction coefficients). Therefore, in the models (below) testing the effects of the TL-SB and TL-SB-TC treatment conditions, we pooled the lesson conditions, including students from both the multiple strategies and content-based prediction conditions.

Treatment effects: Replication and extension questions. Table 4 reports the results of the models that were used to estimate the main effects of treatment conditions and the treatment by student FRL and treatment by school poverty interaction effects. As shown in Table 4, Model 1, there was no difference between students in the teacher lesson and summer book condition (TL-SB) and control group (Coefficient = $-.04$, $SE = .05$); and there was no difference between the teacher lesson, summer book, and teacher call condition (TL-SB-TC and the control condition (Coefficient = $-.02$, $SE = .05$). In addition, the difference between students in the teacher lesson, summer book, and teacher call condition (TL-SB-TC) and the basic treatment was not significant (Coefficient = $.02$, $SE = .05$).

Treatment by Poverty Interaction Effects

Our second exploratory question asked whether treatment effects were moderated by student-level poverty (FRL status). The results of Model 2 in Table 4 indicate that student-level income status predicted treatment effects weakly but not significantly. The direction of the treatment

effect was negative for non-FRL and positive for FRL students, consistent with other research on home-based summer reading programs (Allington et al., 2010; Kim & Quinn, 2013).

The third and last exploratory question asked whether treatment effects were moderated by school poverty. There was clear evidence that school poverty moderated the effects of both the TL-SB and TL-SB-TC treatments. As shown in Model 3 in Table 4, the interaction effects of the TL-SB by school poverty (Coefficient = .19, $p = .06$) was marginally significant and TL-SB-TC by school poverty (Coefficient = .22, $p = .03$) was statistically significant. The predicted effect sizes for both the TL-SB condition ($d = .07$) and the TL-SB-TC condition ($d = .11$) were positive for high poverty schools, relative to the control condition. In sharp contrast, for moderate poverty schools, the predicted effect sizes for the TL-SB condition ($d = -.12$) and for the TL-SB-TC condition ($d = -.11$) were negative relative to the control condition. As a sensitivity analysis (see Appendix B, Table B), we replicated the results of the ordinary least squares regression models using hierarchical linear models that included school-level variables (i.e., high poverty school and school lesson condition), student by school-level interactions (i.e., student-level treatments by high poverty schools), and a school-level variance component.

Differences in Control Group Performance

The disparate results for high versus moderate poverty schools were due in large part to differences in the performance of students in the control condition. Figure 3 displays summer gains or losses as effect sizes (i.e., standardized differences between the fall posttest and spring pretest). For high poverty schools, treatment students performed at about the same level in the spring and fall ($d = -.02$ or $+.02$) and control students lost ground in reading comprehension during the summer months ($d = -.09$), as might be expected. For moderate poverty schools,

treatment students made small summer gains ($d = + .04$ or $+ .05$), whereas control students enjoyed a larger and surprising gain in reading comprehension during the summer ($d = + .16$).

In an effort to deepen our understanding of the control group results, we compared fall-spring standardized differences for control students who were or were not receiving FRL in moderate poverty and high poverty schools. As shown in Figure 4, FRL students in moderate poverty schools scored slightly higher on the fall posttest than on the spring pretest; unlike FRL students in high poverty schools, they did not exhibit summer loss. In both high and moderate poverty schools, students not receiving free or reduced-price lunch made summer gains.

We also compared the performance of control group students in moderate and high poverty schools in this study with the two studies of READS that we were attempting to replicate. As shown in Figure 5, in the initial studies, as in high poverty schools in this study, control group students lost ground in the summer, $d = - .07$ (Kim, 2006) and $d = - .08$ (Kim & White, 2008).

Differential Treatment Effects on Student-Reported Book Reading

To further explore possible reasons for the disparate treatment effects for high and moderate poverty schools, we examined (a) treatment effects on students' book reading, as measured by the survey of summer reading that was given to all children in the fall, and (b) treatment effects on students' reports of whether the books they read in the summer were too easy, too hard, or "just right." Differences in either the quantity or the quality of book reading may account for different effects on reading comprehension in high and moderate poverty schools if, as our logic model (Figure 1) suggests, treatment effects on reading comprehension are mediated by the amount of reading of matched books that treatment group students do relative to the amount of reading that control group students do.

Table 5 presents descriptive statistics on the total number of books students reported reading during the summer by treatment condition, and by school poverty. We report treatment-control differences as a standardized mean difference (Cohen's d) and raw difference (raw d). Treatment group students reported reading about the same number of books in high and moderate poverty schools, and for both high and moderate poverty schools, the pooled treatment group students (i.e., combining TL-SB and TL-SB-TC) reported reading significantly more books than the control group students; however the difference favoring the treatment group was considerably larger for high poverty schools ($d = + .54$ versus $d = + .28$). Interestingly, for moderate poverty schools only, treatment students reported checking out significantly fewer books from the public library than control students ($d = - .22$). Thus in moderate poverty schools, the treatment conditions appear to have caused a decline in the number of books students checked out of the public library in the summer months. As a result, treatment-control differences in the total number of books read were smaller in moderate poverty schools than in high poverty schools.

As shown in Figure 6, in high poverty schools, there was no difference across experimental conditions in the percentage of students who reported that the books they read in the summer were just right, $\chi^2(486) = .13$, n.s. However, in moderate poverty schools, a significantly larger percentage of students in the control children (90%) reported that their books were just right than students in the TL-SB condition (76%) and TL-SB-TC condition (69%), $\chi^2(592) = 24.97$, $p < .001$.

Discussion

The main purpose of this experimental study was to replicate and extend prior research on the effects of a teacher-scaffolded voluntary summer reading intervention that provided students

with books that were matched to their reading levels and interests (READS). We were interested in determining whether the positive impact of READS in two initial randomized control trials would replicate in a different school district. We also wanted to know whether the effects of READS could be enhanced by adding teacher phone calls in the summer. In addition, we asked two exploratory questions about the moderation of treatment effects by poverty measured at the individual student level or the school (aggregate) level. Finally, we explored whether modified end-of-year lessons focusing on content-based prediction would be more effective. In what follows, we begin with content-based prediction and teacher calls before discussing the question of why the impact of READS differed across our earlier studies and this study.

Content-Based Prediction

We found no evidence in this study that the effects of the basic and enhanced treatments on comprehension outcomes were stronger in schools implementing content-based prediction lessons than in schools implementing multiple strategy lessons. However, students in the content-based prediction condition returned significantly more narrative postcards than children in the multiple strategy condition (Table 1). This finding suggests that the story impression activity may have enhanced children's engagement with narrative texts. Clearly, more research is needed to examine how different comprehension routines affect students' reading of different text types or genres in the summer. It may be that students should use different approaches depending on the type of text they are reading.

Teacher Phone Calls

In this study, the teacher phone call condition did not differ significantly from the basic treatment condition providing matched books and teacher scaffolding in end-of-year lessons. An intention-to-treat analysis indicated that students assigned to receive up to three teacher calls had

similar posttest outcomes to students who received no calls. However, implementation data showed that students received fewer calls than we intended: 39% of the children received one call and only 11% received the target number of three phone calls. Thus it would be premature to conclude that teacher phone calls do not improve students' summer reading. The low percentages of students receiving calls reflect the difficulty that teachers experienced in reaching students. A common problem, as noted by teacher log data, was that the parent's home phone was disconnected. It is possible that another approach would be more effective, such as sending parents text messages. Additionally, in future work we plan to target students who are not returning postcards rather than all students.

Non-Replication in the Full Sample of Moderate and High Poverty Schools

In the initial studies of READS, students received teacher scaffolding in the form of end-of-year comprehension lessons and summer books that were matched to their reading levels and interests, and there was a statistically significant effect (Kim & White, 2008) or a nearly significant effect (Kim, 2006) on reading comprehension for all students. In this study as well, students in the basic treatment group (TL-SB) received matched books and teacher scaffolding in the form of end-of-year comprehension lessons, but the difference between the basic treatment group and the control group on the reading comprehension posttest was not significant for the overall sample (combining lesson types). Also, in schools implementing lessons that incorporated the multiple strategy routine and fluency practice, the basic treatment condition (TL-SB) in this study was identical with the treatment condition in Kim and White (2008). Yet the covariate-adjusted posttest means in the top panel of Table 3 show that the basic treatment condition was, if anything, less effective with multiple strategy/fluency lessons than with content-based prediction lessons. Thus the effect of matched books and teacher scaffolding,

either broadly defined or more narrowly defined as multiple strategy/fluency lessons exactly as it was in Kim and White (2008), was not replicated in the full sample of schools. However, the overall null or near-zero treatment effects we observed ($d = -.04$ and $-.02$ for the TL-SB and TL-SB-TC conditions, respectively) were not due to effects of $d = -.04$ or $-.02$ in all schools but rather, to significant variability in treatment effects across the 19 schools. It is important to subject this heterogeneity of treatment effects to close examination and attempt to understand why it occurred.

Positive Effects in High Poverty Schools

For the 10 high poverty schools, the covariate-adjusted reading comprehension posttest treatment group means were higher than the control group mean, $d = +.08$ for both the teacher lesson and summer book (TL-SB) condition and the enhanced treatment condition with teacher calls (TL-SB-TC), $d = +.11$. These positive effects were consistent with our logic model, theoretical explanations of summer loss, and the effects that were observed in the studies we were attempting to replicate, Kim and White (2008, $d = +.14$) and Kim (2006, $d = +.06$), and they compared favorably to the effects of providing books to students in high poverty schools over a three-year period in Allington et al.'s (2010) study, $d = +.14$. Additionally, the effects for high poverty schools were similar to effects in meta-analyses of school-based summer programs by Cooper et al. (2000), $d = +.14$, and home-based summer programs by Kim and Quinn (2013), $d = +.12$ for total reading achievement. Thus the results for high poverty schools make conceptual sense and align with previous theory and research.

What Explains the Negative Treatment Effects in Moderate Poverty Schools?

For the 9 moderate poverty schools, the covariate-adjusted reading comprehension posttest treatment group means were lower than the control group mean, $d = -.12$ for both the teacher

lesson and summer book (TL-SB) condition and the enhanced treatment condition with teacher calls (TL-SB-TC), $d = -.11$. These negative effects were reasonably consistent across the moderate poverty schools. Of the 9 schools, 6 showed negative effects ($d < 0$); they were not the result of something that happened in one or two schools.

As shown in Figure 3, we found that, for moderate poverty schools where 45 to 74% of the students were receiving free or reduced-price lunch, control students overall enjoyed a gain in comprehension during the summer ($d = +.16$). In addition, low-income (FRL) students in the control group did not lose ground over the summer. Because Cooper et al.'s (1996) meta-analysis of the effects of summer vacation on reading achievement found an average summer loss of $d = -.27$ for low-income students and $d = -.14$ for middle-income students, and because more than half of the students in most of our moderate poverty schools were low-income students, these findings were unexpected. Further, summer loss was observed in control group students in our previous studies where 11 of the 12 participating schools were moderate poverty schools as defined here (fewer than 75% of the students receiving free or reduce-price lunch).

Although the surprisingly strong performance of control group students for moderate poverty schools in this study helps to explain why the results differed by school poverty and why our earlier studies did not replicate in the overall sample, it does not explain why control group students outperformed treatment group students for moderate poverty schools, that is, why the treatment conditions had negative effects in relation to the control group.

One possible explanation for the negative effects is that the READS program of matched book delivery caused students to read fewer books in the summer than they ordinarily would have. Consistent with this, we found that treatment group students in moderate poverty schools checked out fewer library books from the public library than control students (Table 5).

However, despite checking out fewer books from the public library, treatment students still read more books than controls, so there would be no reason to expect negative effects. Negative effects might still occur if, in addition, the books that treatment group students read were less well suited to their interests and abilities than they would have been in normal circumstances without READS. There is some support for this “negative substitution” hypothesis in our data. We found that a higher percentage of control students than treatment students reported reading “just right” books (Figure 6).

An alternative explanation for negative effects in moderate poverty schools is that control group students may have checked out more library books than they typically would have if they were not participating in an experiment. We know that the parents of the participating students signed a consent form explaining that their child would be randomly assigned to receive books in either the summer or the following fall. After the first few weeks of summer passed, having deduced that their child was in the control group, some parents may have taken their child to the public library or encouraged her or him to go to the library for the purpose of checking out books to read. Research methods texts refer to such compensatory behavior on the part of the control group as a “John Henry” effect (Cook & Campbell, 1976; see Kocakaya, 2011 for a recent attempt to study John Henry effects).

We still are faced with a challenge in explaining why there were no negative effects for our high poverty schools where, apparently, there was no negative substitution or control group compensation. High poverty schools presumably served students who received less parental guidance in choosing and reading appropriate books (Neuman & Celano, 2006) or served communities where there were fewer books or lower quality books (Neuman & Celano, 2001). We speculate that under these conditions of poverty, (a) our books were not less well matched to

students' reading levels than the books students would normally have read, and/or (b) control parents did not or could not compensate for their child's non-receipt of books from us.

Limitations

It is possible that the negative effects for moderate poverty schools and the statistically significant interaction of the pooled treatment conditions with school poverty represent an anomaly. These findings must be corroborated by future research. We are presently conducting two follow-up studies, an experimental study in a larger sample of high poverty and moderate poverty schools and a mixed-methods study that may shed light on the mechanisms underlying the treatment by poverty interactions.

Another important limitation of this study is the way we measured poverty at the individual student level (i.e., as a dichotomous variable, receipt or non-receipt of free or reduced-price lunch) and at the school level (i.e., as school percent receiving free or reduced-price lunch). These measures are crude. In conducting replication studies across different contexts, researchers must do a better job of defining and measuring poverty. As shown in Figure 4, when "low-income" students are defined by their receipt of free or reduced-price lunch, unexpected differences in reading comprehension may be observed across different contexts. Future studies might employ, for example, a composite measure of socioeconomic status based on family income, parent education, and occupational status. This kind of composite measure may do a better job of both (a) identifying low-SES children who are most at-risk of summer loss and (b) predicting the likelihood of positive treatment effects.

Conclusions and Implications

Our results underscore the importance of conducting well-designed replication studies of reading interventions. Our data suggest that for summer reading programs, school poverty may

determine whether a program replicates across districts and schools. Other factors besides school poverty can influence the results obtained in different settings. Coyne et al. (2013) found that a supplemental beginning reading intervention was effective in six different school districts in Connecticut and Texas but not in a Florida district that had a better coordinated and systematic approach to professional development focusing on evidence-based instructional strategies. In the Coyne et al. (2013) study, as in our study, differences in the response of comparison group students explained the failure to replicate.

There is nothing in our results to discourage policymakers who wish to reduce or eliminate summer loss in high poverty schools. At the same time, if our results for moderate poverty schools can be reproduced in a larger and more diverse sample of districts and schools, it would imply that education policymakers should be exceedingly cautious about implementing summer voluntary reading programs in moderate poverty schools that do not meet the 75% free or reduced-price lunch criterion.

Author Notes

This study was made possible by an Investing in Innovation Fund (i3) grant from the U.S. Department of Education (PR/Award # U396B100195); however, the contents of this article do not represent the policy of the U.S. Department.

¹Correspondence concerning this article should be addressed to Thomas G. White, The Center for Advanced Study of Teaching and Learning, University of Virginia, 2200 Old Ivy Road, Charlottesville, VA, 22903. Email: tgw7u@virginia.edu

Author Statement

Thomas G. White is a Senior Research Scientist at The Center for Advanced Study of Teaching and Learning, University of Virginia, USA; email: tgw7u@virginia.edu (corresponding author).

James S. Kim is an Associate Professor at Harvard University, Graduate School of Education; email: james_kim@gse.harvard.edu

Helen Chen Kingston is a Research Associate at Harvard University, Graduate School of Education; email: helen_kingston@gse.harvard.edu

Lisa Foster is a Postdoctoral Fellow at Harvard University, Graduate School of Education; email: lisa_foster@gse.harvard.edu

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Appendix A: Multilevel Models Used to Estimate Treatment by School Lesson Interactions

Appendix B: Multilevel Models Used to Estimate Treatment by School Poverty Interactions

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171–191. doi: 10.3102/01623737023002171
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180. doi: 10.1177/000312240707200202
- Allington, R. L., & McGill-Franzen, A. M. (Eds.). (2013). *Summer reading: Closing the rich/poor reading achievement gap*. New York, NY: Teachers College Press.
- Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., ... Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411-427. doi: 10.1080/02702711.2010.505165
- 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. <http://www.jstor.org/stable/748043>
- Asher, S. R. (1979). Influence of topic interest on Black children's and White children's reading comprehension. *Child Development, 50*(3), 686-690. doi: 10.1111/1467-8624.ep7250653
- Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., ... Drake, L. (2010). *The condition of education 2010* (NCES 2010-028). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

- Barnhill, K. R. (2010). *The sweaty book of sweat*. Mankato, MN: Capstone Press.
- Benson, J. G., & Borman, G. D. (2010). Family, neighborhood, and school settings across seasons: When do socioeconomic context and racial composition matter for the reading achievement growth of young children? *Teachers College Record*, *112*(5), 1338-1390.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*(6), 1844-1867. doi: 10.1111/1467-8624.t01-1-00382
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004). Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education*, *77*(1), 1-31. doi: 10.1177/003804070407700101
- Carr, E., & Ogle, D. (1987). K-W-L Plus: A strategy for comprehension and summarization. *Journal of Reading*, *30*(7), 626-631. <http://www.jstor.org/stable/40031872>
- Cheadle, J. E. (2008). Educational investment, family context, and children's math and reading growth from kindergarten through the third grade. *Sociology of Education*, *81*(1), 1-31. doi: 10.1177/003804070808100101
- Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of Education*, *77*(3), 185-210. doi: 10.1177/003804070407700301
- Clark, K. F., & Graves, M. F. (2005). Scaffolding students' comprehension of text. *The Reading Teacher*, *58*(6), 570-580. doi: 10.1598/RT.58.6.6

Clay, M. M. (1985). *The early detection of reading difficulties*. Portsmouth, NH: Heinemann.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223-336). Chicago, IL: Rand McNally & Co.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, 65(1), 1-118. doi: 10.1111/1540-5834.00059

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(3), 227-268. <http://www.jstor.org/stable/1170523>

Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M.,... Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, 6(1), 1-23. doi: 10.1080/19345747.2012.706694

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934-945. doi: 10.1037/0012-1649.33.6.934

Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly*, 40(4), 410-427. doi: 10.1598/RRQ.40.4.2

Donovan, C. A., Smolkin, L. B., & Lomax, R. G. (2000). Beyond the independent-level text: Considering the reader? Text match in first graders' self-selections during recreational reading. *Reading Psychology*, 21(4), 309-333. doi: 10.1080/027027100750061949

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(5), 613-635. doi: 10.1177/000312240406900501

Duncan, G. J., & Brooks-Gunn, J. (Eds.). (1997). *Consequences of growing up poor*. New York,

NY: Russell Sage Foundation.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*.

Boulder, CO: Westview Press.

Fry, E. (2002). Readability versus leveling. *The Reading Teacher*, 56(3), 286-291.

<http://www.jstor.org/stable/20205195>

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005).

Quality indicators for group experimental and quasi-experimental research in special

education. *Exceptional Children*, 71(2), 149-164.

Gibbons, P. (2002). *Polar Bears*. New York, NY: Holiday House.

Graves, M., & Graves, B. (2003). *Scaffolding reading experiences: Designs for student success*.

Norwood, MA: Christopher-Gordon Publishers, Inc.

Guthrie, J. T., & Humenick, N. M. (2004). Motivating students to read: Evidence for classroom

practices that increase reading motivation and achievement. In P. McCardle & V.

Chhabra (Eds.), *The voice of evidence in reading research* (pp. 329-354). Baltimore,

MD: Brookes Publishing.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., ...

Tonks, S. (2004). Increasing reading comprehension and engagement through concept-

oriented reading instruction. *Journal of Educational Psychology*, 96(3), 403-423. doi:

10.1037/0022-0663.96.3.403

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3.

American Educator, 27(1), 4-9.

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

Hiebert, E. H., & Sailors, M. (Eds.). (2009). *Finding the right texts: What works for beginning and struggling readers*. New York, NY: Guilford Press.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J.,...

Shannon, G. P. (2003). *The Iowa Tests, guide to research and development, Forms A and B*. Itasca, IL: Riverside Publishing.

Institute of Education Sciences. (2013). *What Works Clearinghouse: Procedures and standards handbook (version 3.0)*. Washington, DC: U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437-447. doi: 10.1037/0022-0663.80.4.437

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, 31(3), 268- 289. doi: 10.1598/RRQ.31.3.3

Kim, J. S. (2004). Summer reading and the ethnic achievement gap. *Journal of Education for Students Placed at Risk*, 9(2), 169-188. doi: 10.1207/s15327671espr0902_5

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*, 28(4), 335–355. doi: 10.3102/01623737028004335

Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology, 99*(3), 505-515. doi:

10.1037/0022-0663.99.3.505

Kim, J. S., & Guryan, J. (2010). The efficacy of a voluntary summer book reading intervention for low-income Latino children from language minority families. *Journal of Educational Psychology, 102*(1), 20-31. doi: 10.1037/a0017270

Kim, J. S. & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research, 83*(3), 386-431. doi:

10.3102/0034654313483906

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading, 12*(1), 1–23. doi:

10.1080/10888430701746849

Kocakaya, S. (2011). An educational dilemma: Are educational experiments working? *Educational Research and Reviews, 6*(1), 110-123. doi: 10.5897/ERR10.219

Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Berkeley, CA: The University of California Press.

Lee, V. E. & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Lindsay, J. (2010). *Children's access to print material and education-related outcomes: Findings from a meta-analytic review*. Naperville, IL: Learning Point Associates.

- LoGerfo, L., Nichols, A., & Reardon, S. F. (2006). *Achievement gains in elementary and high school*. Washington, DC: The Urban Institute.
- McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, 98(1), 14-28. doi: 10.1037/0022-0663.98.1.14
- McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B., Lichter, D. S., Cross, A. B. (2011). *Making summer count: How summer programs can boost children's learning*. Santa Monica, CA: RAND Corporation.
- McGinley, W. J., & Denner, P. R. (1987). Story impressions: A prereading/writing activity. *Journal of Reading*, 31(3), 248-253. <http://www.jstor.org/stable/40029853>
- Meyer, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text: A theoretical and practical handbook for analyzing explanatory text* (pp. 11-64). Hillsdale, NJ: Erlbaum.
- Mol, S. E. & Bus, A. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137(2), 267-296. doi: 10.1037/a0021890
- Morrow, L. M. (2003). Motivating lifelong voluntary readers. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 857-867). Mahwah, NJ: Erlbaum.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270. <http://www.jstor.org/stable/1162893>

- National Assessment Governing Board. (2008). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: US Department of Education.
- National Institute of Child Health and Human Development [NICHD]. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00- 4769). Washington, DC: U.S. Government Printing Office.
- Neuman, S. B., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly*, 36(1), 8-26. doi: 10.1598/RRQ.36.1.1
- Neuman, S. B., & Celano, D. (2006). The knowledge gap: Implications of leveling the playing field for low-income and middle-income children. *Reading Research Quarterly*, 41(2), 176-201. doi: 10.1598/RRQ.41.2.2
- O'Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology*, 102(1), 1-19. doi: 10.1037/a0017488
- Ogle, D. M. (1986). K-W-L: A teaching model that develops active reading of expository text. *The Reading Teacher*, 39(6), 564-570. <http://www.jstor.org/stable/20199156>
- Phillips, M. (2011). Parenting, time use, and disparities in academic outcomes. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 207-228). New York: Russell Sage Foundation.
- Phillips, M., & Chin, T. (2004). How families, children, and teachers contribute to summer learning and loss. In G. D. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs* (pp. 255-278). Mahwah, NJ: Erlbaum.

Pressley, M. (2002). *Reading instruction that works: The case for balanced teaching* (2nd ed.).

New York, NY: Guilford Press.

Reed, D. K. & Vaughn, S. (2012). Comprehension instruction with reading disabilities in grades 4 through 12. *Learning Disabilities -- A Contemporary Journal*, 10(1), 17-33.

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing*, 23(6), 627-644. doi: 10.1007/s11145-009-9175-9

Reutzel, D. R., Jones, C. D., & Newman, T. L. (2010). Scaffolded Silent Reading (ScSR): Improving the conditions of silent reading practice in classrooms. In E. H. Hiebert & D. R. Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers* (pp. 129-150). Newark, DE: International Reading Association.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York, NY: Oxford University Press.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479-530. <http://www.jstor.org/stable/1170585>

Schertle, A. (2009). *Look Out, Jeremy Bean!* Chronicle Books: San Francisco, CA.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov/publications/practiceguides.

- Shany, M. T., & Biemiller, A. (1995). Assisted reading practice: Effects on performance for poor readers in Grades 3 and 4. *Reading Research Quarterly*, 30(3), 382–395.
<http://www.jstor.org/stable/747622>
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95–129. doi: 10.1006/jecp.1998.2481
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books, Inc.
- 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, MD: Brookes Publishing.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407.
<http://www.jstor.org/stable/747612>
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford Press.
- Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, 85(2), 211-229. doi: 10.1037/0022-0663.85.2.211
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. D. Freedle (Ed.), *Advances in discourse processes: Vol. 2. New directions in discourse processing* (pp. 53-119). Norwood, NJ: Ablex.

- Sun, S., Pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology, 102*(4), 989-1004. doi: 10.1037/a0019507
- van Ewijk, R., & Slegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134-150. doi: 10.1016/j.edurev.2010.02.001
- van IJzendoorn, M. H. (1994). *A process model of replication studies: On the relation between different types of replication*. In R. van der Veer, M.H. van IJzendoorn, & J. Valsiner (Eds.), *On reconstructing the mind: Replicability in research on human development* (pp. 57-70). Norwood, NJ: Ablex.
- White, T. G., & Kim, J. S. (2008). Teacher and parent scaffolding of voluntary summer reading. *The Reading Teacher, 62*(2), 116-125. doi: 10.1598/RT.62.2.3
- White, T. G., & Kim, J. S. (2010). Can silent reading in the summer reduce socioeconomic differences in reading achievement? In E. H. Hiebert and D. R. Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers* (pp. 67-91). Newark, DE: International Reading Association.
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*(3), 420- 432. doi: 10.1037/0022-0663.89.3.420
- Wilkins, C., Gersten, R., Decker, L. E., Grunden, L., Brasiel, S., Brunnert, K., and Jayanthi, M. (2012). *Does a summer reading program based on Lexiles affect reading comprehension?* (NCEE 2012-4006). Washington, DC: National Center for Education

Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & deCani, J. S. (2005).

Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, 97(4), 538-550. doi: 10.1037/0022-0663.97.4.538

Appendix A: Multilevel Models to Estimate Treatment by School Lesson Interactions

Results from this multi-level model are displayed in Table A below. At level 1, we fit the following student-level equation

$$Y_{ij} = \beta_{0j} + \beta_{1j} (\text{ITBSPRE}) + \beta_{2j} (\text{TL-SB}) + \beta_{3j} (\text{TL-SB-TC}) + \beta_{4j} (\text{FRL}) + \beta_{5j} (\text{Male}) + e_{ij}$$

where

Y_{ij} is the ITBS posttest score for student i in school j ,

β_{0j} is the mean ITBS posttest for students in school j adjusted for the other predictors,

β_{1j} is the slope of the ITBS pretest of students in school j ,

β_{2j} (TL-SB) is the main effect of the teacher lesson and summer books condition of students in school j ,

β_{3j} (TL-SB-TC) is the main effect of the teacher lesson, summer books, and teacher phone call condition of students in school j ,

β_{4j} (FRL) is the main effect for the indicator of receiving free or reduced-price lunch of students in school j ,

β_{5j} (MALE) is the main effect for the indicator for male students of students in school j ,

and e_{ij} is the student-level residual (level-1 random effect).

The level 2 equation used to address the treatment by lesson condition was written as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (\text{High poverty school}) + \gamma_{02} (\text{Lesson condition}) + \gamma_{03} (\text{TL-SB} * \text{Lesson condition}) + \gamma_{04} (\text{TL-SB-TC} * \text{Lesson condition}) + \mu_{0j}$$

where

β_{0j} is the mean ITBS posttest for students in school j adjusted for the other predictors,

γ_{00} is the intercept associated with level-1 predictors across schools,

γ_{01} is the main effect of high poverty school across schools,

γ_{02} is the main effect of the lesson condition across schools,

γ_{03} is the interaction effect of TL-SB and lesson condition across schools,

γ_{04} is the interaction effect of TL-SB-TC and lesson condition across schools, and

μ_{0j} is the school level random effect.

Table A

Results of Multilevel Models Used to Estimate Treatment Effects, and Cross-Level Interactions with Lesson Condition

Fixed Effect	Model 1			Model 2		
	Coefficient	SE	z	Coefficient	SE	z
Intercept	0.226	0.06	3.58**	0.227	0.07	3.31**
Student variables						
ITBS pretest (β_{1j})	0.861	0.023	37.25**	0.861	0.023	37.24**
TL-SB (β_{2j})	-0.032	0.049	-0.65	-0.042	0.068	-0.61
TL-SB-TC (β_{3j})	0.000	0.049	0.00	0.004	0.068	0.05
FRL (β_{4j})	-0.255	0.053	-4.84**	-0.255	0.053	-4.83**
Male (β_{5j})	-0.005	0.040	-0.13	-0.006	0.040	-0.14
School variables						
High poverty (γ_{01})	-0.080	0.055	-1.46	-0.080	0.055	-1.46
Lesson condition (γ_{02})	0.087	0.052	1.67~	0.083	0.077	1.08
TL-SB * Lesson condition (γ_{03})				0.021	0.098	0.21
TL-SB-TC * Lesson condition (γ_{04})				-0.009	0.098	-0.09
Random Effect						
Student-level (e_{ij})	0.686			0.686		
School-level (μ_{0j})	0.066			0.066		

Note: ITBS = Iowa Test of Basic Skills, Reading Comprehension. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls. FRL = Student eligible for free- or reduced-price lunch.
~ $p < .10$. * $p < .05$. ** $p < .01$.

Appendix B: Multilevel Models Used to Estimate Treatment by School Poverty Interactions

As a sensitivity test, we also used multilevel models to examine the treatment by high poverty school interaction. The level 1 student model was identical to the previous analysis involving the treatment by school lesson condition cross-level interaction. The level 2 equation was specified as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (\text{High poverty school}) + \gamma_{02} (\text{Lesson condition}) + \gamma_{03} (\text{TL-SB} * \text{High poverty school}) + \gamma_{04} (\text{TL-SB-TC} * \text{High poverty school}) + \mu_{0j}$$

where

β_{0j} is the mean ITBS posttest for students in school j adjusted for other predictors,
 γ_{00} is the intercept associated with level-1 predictors across schools,
 γ_{01} is the main effect of high poverty school across schools,
 γ_{02} is the main effect of the lesson condition across schools,
 γ_{03} is the interaction effect of TL-SB and high poverty school across schools,
 γ_{04} is the interaction effect of TL-SB-TC and high poverty school across schools, and
 μ_{0j} is the school level random effect.

Table B

Results of Multilevel Models Used to Estimate Treatment Effects, and Cross-Level Interactions with Student FRL and School Poverty

Fixed Effect	Model 1			Model 2		
	Coefficient	SE	z	Coefficient	SE	z
Intercept	0.297	0.076	3.92**	0.281	0.067	4.17**
Student variables						
ITBS pretest (β_{1j})	0.861	0.023	37.33**	0.861	0.023	37.32**
TL-SB (β_{2j})	-0.179	0.090	-1.99~	-0.116	0.065	-1.79~
TL-SB-TC (β_{3j})	-0.073	0.093	-0.79	-0.084	0.065	-1.30
FRL (β_{4j})	-0.358	0.081	-4.45**	-0.252	0.053	-4.78**
Male (β_{5j})	-0.003	0.040	-0.08	-0.007	0.040	-0.18
TL-SB * FRL (β_{6j})	0.209	0.107	1.95~			
TL-SB-TC * FRL (β_{7j})	0.105	0.109	0.96			
School variables						
High poverty (γ_{01})	-0.080	0.055	-1.47	-0.210	0.080	-2.66**
Lesson condition (γ_{02})	0.086	0.052	1.66~	0.087	0.052	1.68~
TL-SB * High poverty (γ_{03})				0.195	0.099	1.98*
TL-SB-TC * High poverty (γ_{04})				0.194	0.099	1.96*
<i>Random Effect</i>						
Student-level (e_{ij})	0.685			0.685		
School-level (μ_{0j})	0.067			0.066		

Note: ITBS = Iowa Test of Basic Skills, Reading Comprehension. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls. FRL = Student eligible for free- or reduced-price lunch.
 ~ $p < .10$. * $p < .05$. ** $p < .01$.

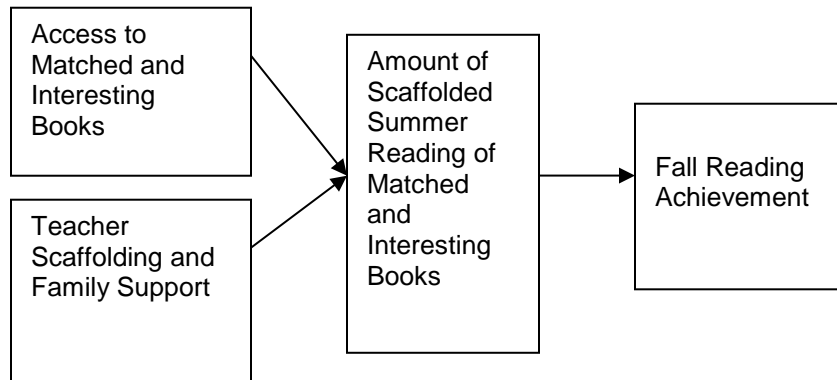


Figure 1. Logic model for studies of scaffolded voluntary summer reading.

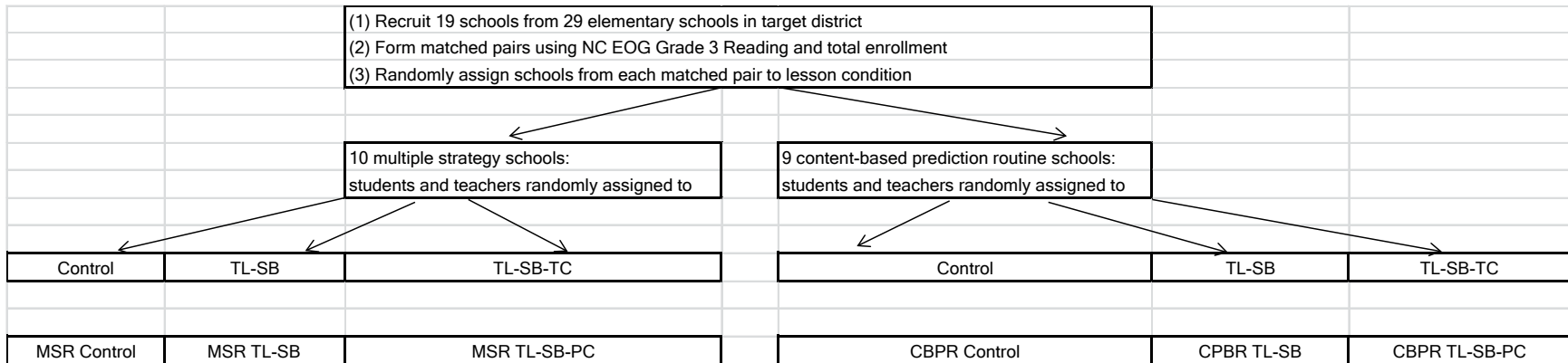


Figure 2. Two-stage procedure used to randomly assign schools to lesson conditions and students to treatment or control conditions.

NC EOG = North Carolina End-of-Grade Test, Grade 3 Reading. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher

lessons, summer books, phone calls. FRL = free or reduced price lunch status.

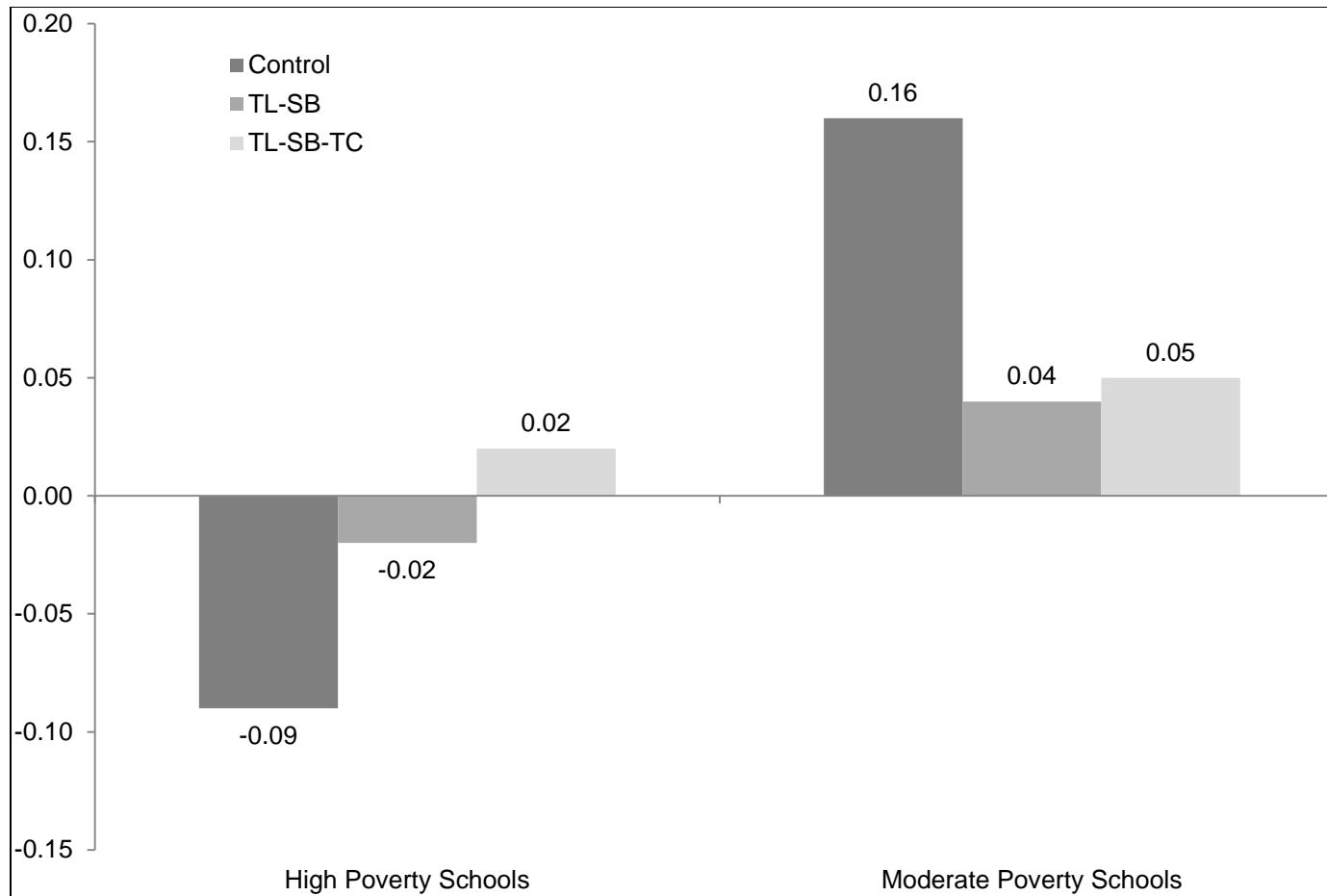


Figure 3. Summer reading loss (or gain) on reading comprehension, by treatment condition and school poverty. Vertical axis displays summer loss (or gain) as a standardized mean difference (spring to fall change in reading comprehension scores in standard deviation units). Horizontal axis displays magnitude of summer loss (or gain) in 10 high poverty schools (75% to 100% receiving free or

reduced-price lunch, FRL) and 9 moderate poverty schools (45% to 74% FRL), by treatment condition. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls.

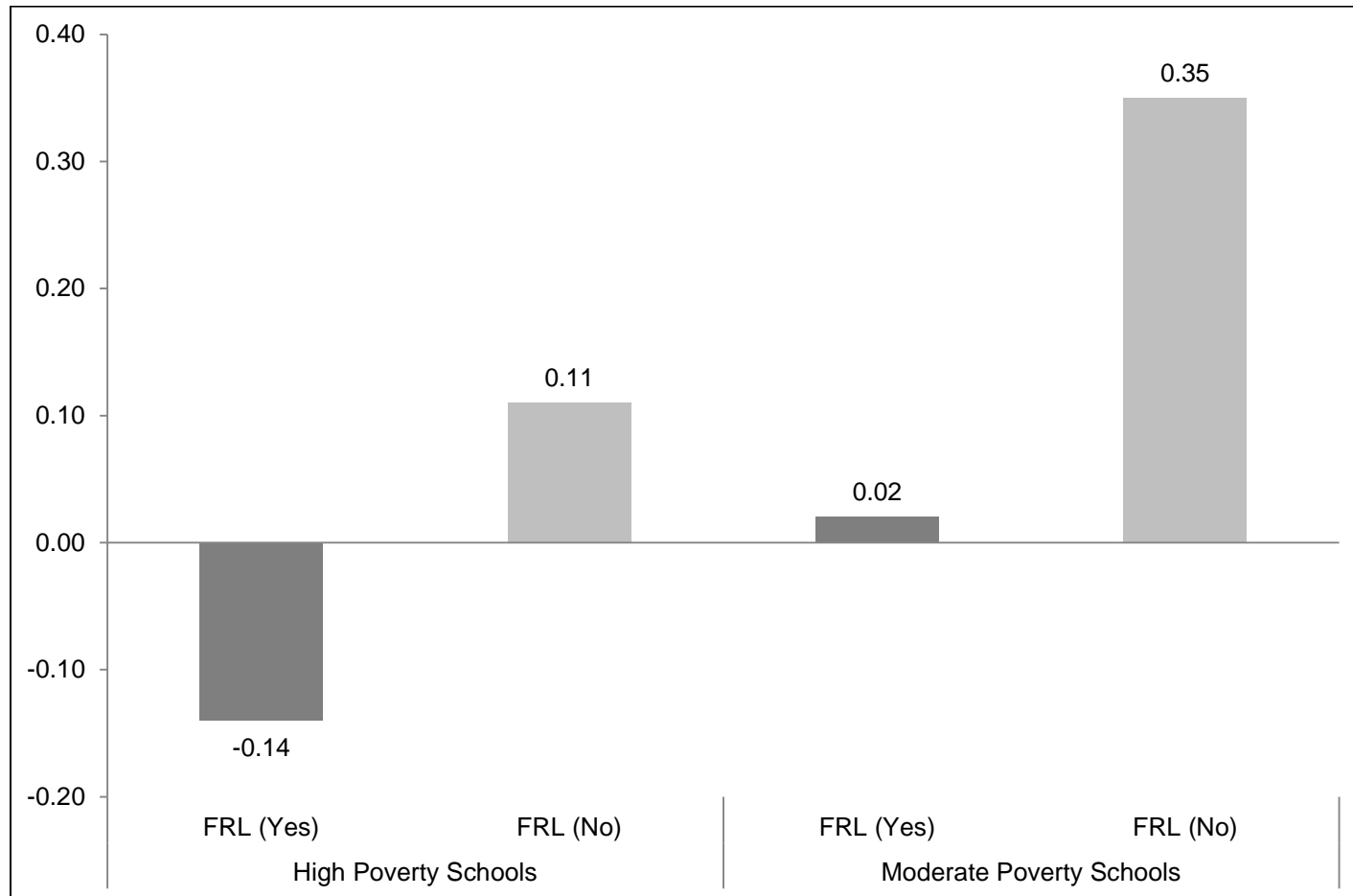


Figure 4. Summer reading loss (or gain) on reading comprehension for control students only, by student FRL status (free or reduced price lunch) and school poverty. Vertical axis displays summer loss (or gain) as a standardized mean difference (spring to fall change in reading comprehension scores in standard deviation units). Horizontal axis displays magnitude of summer loss (or gain) in 10 high

poverty schools (75% to 100% receiving free or reduced-price lunch, FRL) and 9 moderate poverty schools (45% to 74% FRL), by treatment condition.

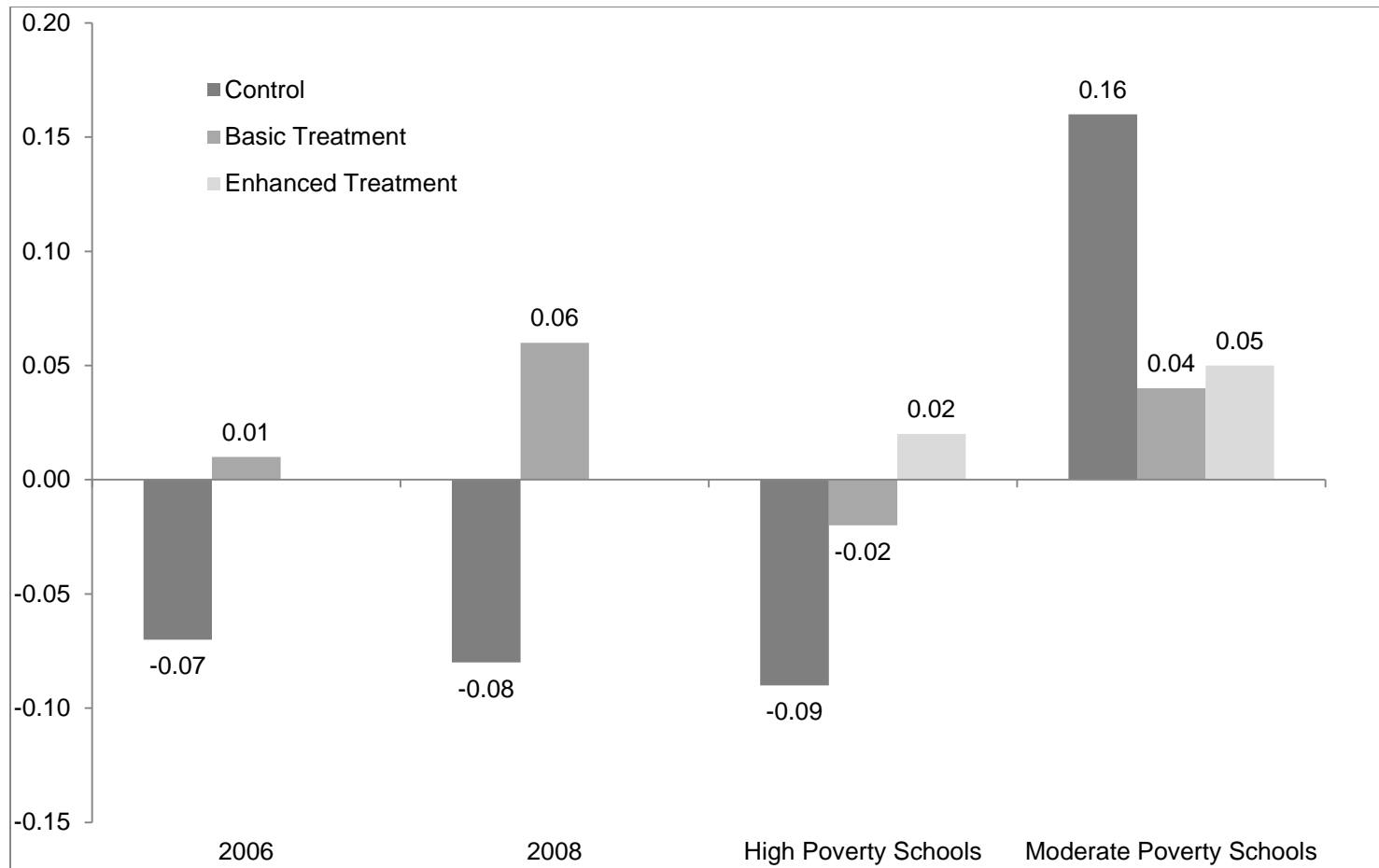


Figure 5. Comparison of summer reading loss (or gain) on reading comprehension for control and treatment students in two previous studies and the current study, by school poverty. Vertical axis displays summer loss (or gain) as a standardized mean difference (spring to fall change in reading comprehension scores in standard deviation units). Horizontal axis displays magnitude of summer loss (or gain) for control students and treatment students in Kim (2006) and Kim and White (2008) and the current study in 10 high

poverty schools (75% to 100% receiving free or reduced-price lunch, FRL) and 9 moderate poverty schools (45% to 74% FRL), by treatment condition.

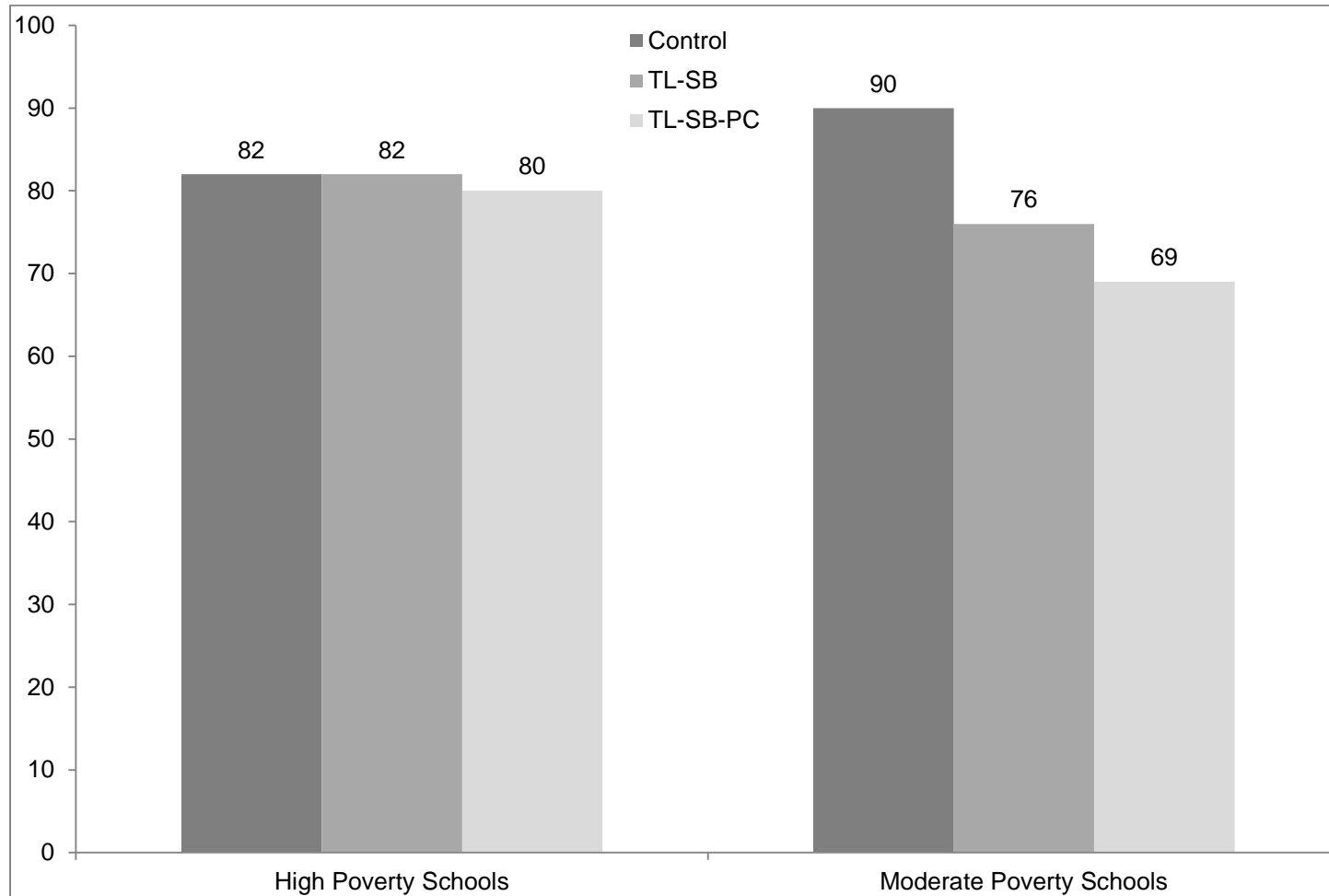


Figure 6. Percentage of students reporting that the books they read in the summer were just right, by treatment condition and school poverty. Vertical axis displays the percentage of students (%). Horizontal axis displays % of students by treatment condition and in 10 high poverty schools (75% to 100% receiving free or reduced-price lunch, FRL) and 9 moderate poverty schools (45% to 74%

FRL), by treatment condition. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls.

Table 1

Postcard Return Rates for 8 Matched Books, by Lesson Condition

Variables	Content-based Prediction Routine (n = 252)		Multiple Strategy Routine (n = 293)		<i>t</i>	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Matched books total	3.24	2.58	2.14	2.36	5.20	<.001	0.45
Narrative matched books	2.19	2.22	1.11	1.87	6.17	<.001	0.53
Informational matched books	1.06	1.66	1.04	1.71	0.12	n.s.	0.01

Table 2

Descriptive Characteristics of Children in Control and Treatment Groups on Pretest Measures, (N = 1,421)

Characteristic	Control		TL-SB		TL-SB-TC		<i>F</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
ITBS Reading Comprehension	184	25	184	25	184	24	0.01	0.99
Low-income (%)	70	46	72	45	75	43	1.48	0.23
Black (%)	51	50	51	50	52	50	0.05	0.95
Latino/a (%)	27	45	30	46	34	47	2.59	0.08
Limited English proficient (%)	26	44	31	46	30	46	1.48	0.23

Note. ITBS = Iowa Tests of Basic Skills, Reading Comprehension Scaled Score. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls. FRL = free or reduced-price lunch status.

Table 3

Descriptive Statistics for Covariate-Adjusted Posttest Scores by School Lesson Routine

School sample	Control				TL-SB				TL-SB-TC			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min-Max</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min-Max</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min-Max</i>
All Schools (CBPR + MSR)	396	184.95	29.73	131-260	395	184.01	28.88	123-260	397	184.55	28.29	123-260
CBPR Schools (n = 9)	196	184.14	29.82	131-260	184	183.43	29.15	123-260	182	185.27	27.9	123-260
MSR Schools (n = 10)	200	185.73	29.70	135-260	211	184.51	28.72	138-260	215	183.90	28.66	131-255

Note. MSR= Multiple Strategy Routine. CBPR = Content-Based Prediction Routine. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, teacher phone calls.

Table 4

*Results of Ordinary Least Squares (OLS) Regression Analyses Used to Estimate Treatment**Effects and Treatment by Student FRL Status and Treatment by High Poverty School Interaction**Effects*

Variable	Model 1	Model 2	Model 3
ITB pretest	0.927*** (0.020)	0.869*** (0.023)	0.912*** (0.021)
TL-SB	-0.043 (0.050)	-0.179* (0.091)	-0.125~ (0.066)
TL-SB-TC	-0.017 (0.050)	-0.084 (0.094)	-0.113~ (0.066)
FRL		-0.380*** (0.079)	
TL-SB * FRL		0.206~ (0.108)	
TL-SB-TC * FRL		0.119 (0.110)	
High poverty school			-0.239*** (0.072)
TL-SB * High poverty school			0.189~ (0.100)
TL-SB-TC * High poverty school			0.220* (0.100)
Constant	0.048 (0.035)	0.315*** (0.065)	0.152** (0.047)
<i>N</i>	1189	1182	1189
<i>R</i> ²	0.637	0.647	0.641

Note. Standard errors are in parentheses. ITBS = Iowa Test of Basic Skills, Reading Comprehension. TL-SB = Teacher lessons and summer books. TL-SB-TC = Teacher lessons, summer books, phone calls. FRL = Student eligible for free- or reduced-price lunch.

~ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 5

Student-Reported Independent Book Reading during the Summer

Survey question	Control			TL-SB			TL-SB-TC			Treatment Effects			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>	<i>Cohen's d</i>	<i>raw d</i>
All Schools (n = 19)													
Total summer books read	360	6.16	3.52	367	7.49	2.89	365	7.36	2.94	20.05	0.00	0.39	1.27
Total books checked out of the library	357	3.74	3.41	364	3.47	3.53	364	3.36	3.44	1.19	0.30	-0.09	-0.33
High Poverty schools (n = 10)													
Total summer books read	161	5.61	3.44	165	7.48	2.76	166	7.13	2.92	17.25	0.00	0.54	1.70
Total books checked out of the library	160	2.95	2.96	165	3.32	3.44	165	3.07	3.39	0.55	0.58	0.08	0.25
Moderate Poverty schools (n = 9)													
Total summer books read	199	6.60	3.53	202	7.50	3.00	199	7.55	2.94	5.70	0.00	0.28	0.93
Total books checked out of the library	197	4.39	3.62	199	3.59	3.61	199	3.59	3.46	3.32	0.04	-0.22	-0.80

Note. *Cohen's d* = standardized mean difference. *Raw d* = raw difference in the number of books. Treatment effects = Mean differences in book reading measures for the control group and the combined treatment groups (TL-SB and TL-SB-TC).