NCEE 2012-4006

U. S. DEPARTMENT OF EDUCATION

Does a summer reading program based on Lexiles affect reading comprehension?

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Does a summer reading program based on Lexiles affect reading comprehension?

Final Report

March 2012

Authors:

Chuck Wilkins, Ph.D.
Edvance Research, Inc.

Russell Gersten, Ph.D.
Instructional Research Group

Lauren E. Decker, Ph.D.
Edvance Research, Inc.

Leslie Grunden, Ph.D.
Edvance Research, Inc.

Sarah Brasiel, Ph.D.
Edvance Research, Inc.

Kim Brunnert, Ph.D.
Edvance Research, Inc.

Madhavi Jayanthi, Ed.D.
Instructional Research Group

Project Officer:

Karen Armstrong
Institute of Education Sciences

NCEE 2012-4006

U.S. Department of Education

U.S. Department of Education

Arne Duncan

Secretary

Institute of Education Sciences

John Q. Easton

Director

National Center for Education Evaluation and Regional Assistance

Rebecca A. Maynard

Commissioner

March 2012

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06-CO-0017 with Regional Educational Laboratory Southwest administered by Edvance Research.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Disclosure of potential conflicts of interest

The Lexile Framework® for Reading, developed by MetaMetrics,® Inc., was used in this study to match books, supplied by Scholastic, Inc., with student reading level. MetaMetrics, Inc., as part of its normal business practices, provided Lexile measures for the books selected for this study. Regional Educational Laboratory (REL) Southwest at Edvance Research, Inc. has no direct relationship with MetaMetrics, Inc. and takes no position on the utility of the Lexile Framework compared with other measures. In addition, REL Southwest has no financial interest in MetaMetrics, Inc., the use of the Lexile Framework, or Scholastic. None of the authors or other staff involved in the study from Instructional Research Group, MetaMetrics, Inc., or Scholastic, Inc., has financial interests that could be affected by the content of this report.[1]

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Acknowledgments

This study was a collaborative effort by many individuals and organizations worthy of acknowledgment. The study authors would like to thank all the district personnel, principals, teachers, and students for their assistance during the summer and school year. Without their participation, this evaluation would not have been possible.

Through the involvement and guidance of a number of technical experts, the study team was able to clarify and solve many complex technical and analytical questions. The team of experts included Russell Jackson with Decision Information Resources, Inc. (DIR) and David Myers with American Institutes for Research (AIR), a member of the study's technical working group.

The study authors would also like to thank the many individuals who participated in the study. Stephanie Adams, project manager, directed and participated in all aspects of the study from initial study planning to final report preparation. Anna Baldwin, project coordinator, oversaw site-management activities and provided a wide range of assistance throughout the evaluation. Laural Logan-Fain led recruitment efforts with remarkable persistence and patience. Meg Grant contributed to the recruitment activities and directed all study dissemination efforts with insightful understanding of constituents. Debrale Graywolf provided comprehensive and essential editorial services and guidance. Lois Gregory, Noelle Howland, Marshall Garland, and Debora Carmichael, along with Herb Turner of Analytica, Inc., provided expert quality reviews and invaluable advice throughout report preparation and multiple review cycles. Drs. Ellie Sanford-Moore and Heather Koons of MetaMetrics contributed to the categorization and Lexile assignments of the books for students, Lexile documentation, and report preparation. Eric Schmitt of Scholastic, Inc. provided support and direction for processing and managing the orders and shipment of books to students.

Contents

Tables

Figures

Executive summary

To successfully engage in today's global market, students need advanced literacy skills (Snow, Burns, and Griffin 1998). A lack of proficiency in reading is more widely found in children from economically disadvantaged families (Alexander, Entwisle, and Olson 2007; Lee, Grigg, and Donahue 2007); in fact, by grade 4, only 46 percent of students from economically disadvantaged families achieve reading proficiency above the basic level (Perie, Grigg, & Donahue, 2005). One reason that these students tend to have lower reading proficiency is that they experience a decline in reading comprehension over the summer months, known as summer reading loss (Cooper et al. 1996; David 1979). This disproportionate reading loss for economically disadvantaged students may, in part, be explained by the limited access to books and literacy-related activities in the home environment that many of these students experience (Allington, Guice, Baker, Michelson, and Li 1995; Fryer and Levitt 2002; McGill-Franzen, Lanford, and Adams 2002; McGill-Franzen et al. 2005).

This decline in reading skills over the summer months has led some researchers to attempt to improve reading comprehension through implementation of summer reading programs. Although research regarding summer reading programs that include providing books to students in the home environment is limited (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008), the research does provide some evidence that summer reading programs may have the potential to raise the reading comprehension of economically disadvantaged students over the summer.

Kim (2006) recommended two lines of research focusing on the effect of a summer reading program. One approach would be to investigate the effects of a voluntary summer reading intervention versus a mandatory summer school program. The second approach would be to investigate the effectiveness of individual components of a summer reading program, including: a) various teacher or parent instruction/encouragement activities prior to or during the summer, and b) actually providing students with books for summer reading.

Kim and other researchers have pursued the second approach (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008). These studies, including randomized controlled trials (RCT; Allington et al. 2010; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008) and studies that used quasi-experimental designs (Butler 2010; Crowell and Klein 1981) have examined the effects of summer reading programs that supply students with books (matched on reading level, interest area, or both) on student reading achievement. Three of the five RCTs found statistically significant effects of a summer reading program on reading achievement (Allington et al. 2010, Kim 2006, and Kim and White

2008). The other two RCTs did not find a statistically significant effect on reading achievement, but did find a statistically significant effect on the number of books read (Kim 2007, Kim and Guryan 2010).

All but one of these previous studies regarding summer reading took place in a single district (Allington et al. 2010 examined two districts). Only one study (Butler 2010) required student enrollment in the free or reduced-price lunch (FRPL) program for study eligibility, and none of the previous studies targeted students reading below the 50th percentile nationally (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008). Besides providing matched books to students, four of the five RCT studies evaluated a summer reading program that included one or more parent components (such as family literacy events) and/or one or more teacher components (such as training and classroom instruction) that could be burdensome to implement (Kim 2006, Kim 2007, Kim and Guryan 2010, Kim and White 2008). Allington (2010) studied a summer reading program that only provided students with books, without additional teacher or parent components, and estimated the cumulative effects of students receiving books (provided by the study) for three years.

Because different studies used different combinations of parent/teacher components, it was unclear which of these would facilitate student summer reading and subsequent reading achievement, and which would not. Therefore, to begin to understand which components are beneficial, the program examined in the current study removed certain components of the intervention developed by Kim and colleagues (Kim 2006, Kim 2007, Kim and Guryan 2010, Kim and White 2008), particularly ones that involved cost and time allocated to parent or teacher training and teachers spending instructional time preparing students for summer reading. The current study investigated the effects of providing books to students, along with sending them reminder postcards during the summer, but without any other encouragement components or teacher instruction. If this streamlined approach was found to be effective, the study team reasoned that this would indicate that the intervention Kim studied could be implemented without imposing any burden on teachers and schools, such as the costs associated with teacher-training and guidance. This streamlined approach could also be replicated by charitable organizations and other groups independent of schools, who could manage the program and even provide the needed books, further reducing costs.

This report presents estimates from a large-scale, multi-district RCT on the effectiveness of a summer reading program on improving student reading comprehension for economically disadvantaged grade 3 students reading below the 50th percentile nationally. This study focused on the summer between grades 3 and 4 for three reasons: (1) independent reading demands increase dramatically in grades 3 and 4 (Chall 1983; National Research Council 1998); (2) the grade 3 to grade 4 transition was not a focus of previous studies; and (3) Texas state assessment data are available for the first time for

students beginning in grade 3, and those data were used to control for baseline differences in this study. Each student in the treatment group was sent a single shipment of eight books matched to his or her reading level and interest area during the first part of the summer (June/July 2009), followed by a reminder postcard each week for six weeks. Eight books were chosen because this was the number used in two of the summer reading programs shown to have statistically significant positive effects on reading comprehension for specific subgroups (Kim 2006) or for the entire sample (Kim and White 2008).

Students and books were matched using the Lexile Framework® for Reading, a linguistic, theory-based method of matching reading level and books that was developed by MetaMetrics, Inc. (MetaMetrics, Inc. n.d.d), and student interest surveys. According to a 2001 report from a panel of reading experts working with the National Center for Education Statistics, the Lexile Framework has been found to have solid psychometric properties and has been validated across a wide range of populations (White and Clement 2001). Lexile measures are reported for both readers and texts using a common scale unit called a Lexile (L), which ranges from 0L for emerging readers and beginning texts to 1700L for advanced readers and texts. A book's Lexile measure is calculated by parsing the entire text into slices of 125 words and using a proprietary regression equation to assign a reading difficulty value to each slice based on word frequency and sentence length. The results are combined across slices to obtain the overall Lexile measure for the book. To match students to books on interest area, students were asked to select areas of interest from a list of such topics as art, adventure, and mystery..

Study questions

The study was designed to answer the following confirmatory research question:

- For economically disadvantaged students reading below the 50th percentile nationally, does being sent eight free books in the first part of the summer (June/July 2009) matched to reading level and interest area, along with six reminder postcards, result in significantly better reading comprehension scores in the fall?

and two exploratory research questions:

- Did students in the summer reading program (treatment group) report reading more books over the summer than did students in the control group?

- Did the summer reading program have differential effects on reading comprehension, depending on baseline reading proficiency?

Study design and methodology

Based on research using the same book-matching method (Kim 2006, 2007; Kim and White 2008) and on feedback from this study's technical working group, the study team concluded that it should design the study to be able to detect a mean difference in reading achievement of 0.12 standard deviations or larger. Four Texas school districts were recruited to meet the target sample size of 1,516 students needed to satisfy the power demands of the study. Three districts invited all schools to participate in the study; the fourth district invited a subset of schools.

Students had to meet three criteria to be eligible for the study: (1) be enrolled in the FRPL program (to identify economically disadvantaged students); (2) have a Lexile measure (provided as part of the Texas assessment program) at or below 590L on the spring 2009 grade 3 English language reading assessment, which represents the 50th percentile for grade 3 on national norms linked to the Lexile scale (Scholastic, Inc. 2007); and (3) be physically and mentally able to read independently (as determined by their teachers) since students were not going to receive support during the summer. Consent forms were distributed to students in participating schools who met the eligibility criteria, and students with parent consent participated in the study.

Random assignment took place separately in each school. Participating students were randomly assigned to the treatment group or the control group. Treatment group students received a single shipment of eight matched books in July followed by six weekly reminder postcards during summer 2009. Control group students were not sent books or postcards during the summer. The baseline sample consisted of 1,785 students (896 treatment and 889 control) in 112 schools.

The Scholastic Reading Inventory (SRI; Scholastic, Inc. 1999), administered to both treatment and control group students, was used as the posttest measure of reading comprehension. Students were also asked to fill out a summer reading survey describing their reading activities during the summer. Eighty-eight percent of students in both groups (791 treatment and 780 control) completed the SRI, yielding a final analytic sample of 1,571 students. Eighty-four percent of students in the treatment group (n=750) and 83 percent in the control group (n=734) completed the summer reading survey. Due to logistics, for most posttested students (n=1,002; 63.8 percent), these outcome data were collected in October; for others, data were collected in September (n=246; 15.7 percent), November (n=299; 19.0 percent), or early December (n=24; 1.5 percent).

Analysis and findings

Ordinary least squares regression was used to examine the impact of the summer reading program on reading comprehension. Baseline Lexile scores and school indicators

were included as covariates in the analytic model. Missing data were addressed using casewise deletion.

The confirmatory analysis found that the summer reading program did not have a statistically significant impact on student reading comprehension (effect size=0.02, p=0.62). Seven sensitivity analyses showed the confirmatory impact result to be robust to different analytic approaches. The first exploratory analysis identified a statistically significant effect of the summer reading program on the number of books students reported reading over the summer (effect size=0.11, p=0.01). On average, this is equivalent to students in the treatment group having reported reading 1.03 more books over the summer (mean=8.76 books) than did students in the control group (mean=7.73 books). The second exploratory analysis did not find a statistically significant differential effect of the summer reading program on reading comprehension for students at three different levels of baseline reading proficiency (p-values for comparisons of different levels ranged from 0.64 to 0.98).

Conclusions

Seven previous studies examined summer reading programs, and five found a statistically significant improvement in reading achievement following implementation of a reading program (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006; Kim and White 2008). Of the five studies that used an RCT design, three found a statistically significant effect on reading achievement (Allington et al. 2010; Kim 2006; Kim and White 2008). The current study's confirmatory finding did not replicate the findings from these studies.

Two of the five RCT studies found that students sent books over the summer reported reading more books than did students who were not sent books (Kim 2007; Kim and Guryan 2010); an exploratory analysis in the current study found similar results.

The summer reading program examined in this study did not include teacher support, instructional components, or parent involvement, which several previous studies had included to varying degrees—four RCTs (Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008) and one quasi-experiment (Butler 2010). These other components could potentially account for differences in observed effects across studies. Also, the program examined in the current study spanned a single summer, whereas the program examined in Allington et al. (2010) spanned three summers. Further, the current study sample consisted of economically disadvantaged students reading below the 50th percentile nationally, while the samples in the studies with statistically significant results consisted of students with economically diverse backgrounds (Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008) and were not composed exclusively of students

reading below the 50th percentile nationally (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008).

One possible inference to draw from this study, and the more recent work of Kim and colleagues (Kim and Guryan 2010; Kim and White 2008), is that some of the components that Kim and his colleagues added—in particular, personalized teacher encouragement of each student to read the books during the summer and brief, small group lessons on strategies for reading—may be essential components to success. Although such additions may be costly and time intensive for the teaching staff, many teachers find this type of activity a rewarding part of their jobs. Future scale-up research could continue to examine the issue of varied types of teacher and parent support components that Kim included (Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008).

Allington (2010) found that when students were provided books over a period of three summers, even without any additional support components, student reading significantly improved. Therefore, it may be that teacher and parent support components are necessary for a summer reading program to be effective during a single summer, but may be less important if students participate in summer reading programs over a longer time period.

Study limitations and suggestions for future research

This study has several limitations. First, the sample was not selected randomly from a larger population, but rather from four Texas school districts that met qualifying criteria and agreed to participate; therefore, the results from this study would not necessarily apply to students in other districts.

Second, parent consent required for student participation, ranged from 42 to 74 percent across the four participating districts. Because households in which parents are more likely to consent may differ from households in which parents are less likely to consent, the results might not generalize to all students meeting study criteria in participating districts.

Third, because the study only focused on economically disadvantaged students whose reading skills were below the 50th percentile nationally, the results of this study may not generalize to other groups of students. This exclusion of students reading at or above the 50th percentile could further lead to statistical artifacts due to restriction of range, such as underestimating correlations.

Fourth, matched books were sent to students in the treatment group, but there was no follow-up or survey evidence to determine whether students received the books or read them, or even which portions of the books they read. Therefore, the study cannot say whether treatment students who received and read the books were affected differently by

the summer reading program than treatment students who may not have received or read the books.

Fifth, scheduling constraints delayed shipment of the books until July, shortening the time students had to read the books.

Sixth, there was a lag between the end of the summer reading program and administration of the posttest and summer reading survey. Both treatment and control students received classroom instruction in the fall semester, and effects of the program may have diminished over time; thus, the lag might have made any effect harder to detect. In addition, responses to the summer reading survey, completed in the fall, may have been influenced by students' varying ability to accurately recall their summer reading activities.

Seventh, because the random assignment was conducted within schools, it is possible that there was some contamination between treatment and control students or their parents, and this could have affected the behavior of control students or parents.

Eighth, only a single outcome measure, the SRI, was used in this study, which is a different measure than used in some other summer reading studies, potentially making it difficult to compare the results across studies

Finally, the survey instruments in this study were modified from instruments used in previous research. Although those instruments were well-developed, the modifications were minor, and these were not used for evaluating the confirmatory research question, it should be noted that there was no attempt to evaluate the psychometric properties of the modified instruments.

Future research could explore variations of the summer reading program examined in this study. For example, the impact of a summer reading program over a single summer (such as the current study) could be compared with the impact of a program spanning several summers, such as in Allington et al. (2010), to explore dosage effects. In addition, the current study did not include the teacher instruction and parent involvement components that were part of summer reading programs found to have statistically significant positive impacts on student reading achievement (Butler 2010; Crowell and Klein 1981; Kim and White 2008). Future research might examine what additional components might result in statistically significant effects for a summer reading program.

Chapter 1. Introduction and study overview

To successfully engage in today's global market, students need advanced literacy skills (Snow, Burns, and Griffin 1998). A lack of proficiency in reading is more widely found in children from economically disadvantaged families (Alexander, Entwisle, and Olson 2007; Lee, Grigg, and Donahue 2007). By grade 4, only 46 percent of students from economically disadvantaged families achieved reading proficiency above the basic level (Perie, Grigg, & Donahue, 2005). One reason that economically disadvantaged students tend to have lower reading proficiency is they experience a decline in reading comprehension over the summer months, known as summer reading loss (Cooper et al. 1996; David 1979). This disproportionate reading loss for economically disadvantaged students may, in part, be explained by many of these students having limited access to books and literacy related activities in their home environments (Allington, Guice, Baker, Michelson, and Li 1995; Fryer and Levitt 2002; McGill-Franzen, Lanford and Adams 2002; McGill-Franzen et al. 2005). This reading loss may also be due to less exposure to complex language at home (Hoff 2003; Hart and Risley 1995), fewer materials in the home that stimulate learning (Conger and Donnellan 2007; Constantino 2005; Neuman et al. 2001), and limited summer reading activity (Anderson and Stokes 1984; Carlson 1998; Storch and Whitehurst 2001; Vernon-Feagans et al. 2001).

This decline in reading over the summer months has led some researchers to attempt to improve reading comprehension through the implementation of summer reading programs. Although research about summer reading programs that include providing books to students in the home environment is limited (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008), the research does provide some evidence that summer reading programs may have the potential to raise the reading comprehension of economically disadvantaged students over the summer.

Kim (2006) recommended two lines of research focusing on the effect of a summer reading program. One approach called for investigating the effects of a voluntary summer reading intervention versus a mandatory summer school program. The second approach called for investigating the effectiveness of individual components of a summer reading program including: a) various teacher or parent instruction/encouragement activities prior to or during the summer, and b) actually providing students with books for summer reading. Kim pursued the second approach, researching the effects of changes in teacher instruction/encouragement activities while maintaining a consistent provision of books for summer reading (Kim 2007; Kim and Guryan 2010; Kim and White 2008). Allington (2010) studied a summer reading program that only provided students with books, without additional teacher or parent components, and focused on the cumulative effects of students receiving books for three years.

The purpose of this study is to investigate the effectiveness of a summer reading program that provides books to students at an appropriate level of difficulty in areas of interest, along with

reminder postcards sent to them during the summer, but without any other teacher or parent components. If effective, this approach could be replicated by charitable organizations and others independent of schools or certified teacher trainers. The selected summer reading program for this study did not have an administrative impact on schools or teachers, yet supported student reading over the summer.

This study considers the potential value of a short-term summer reading program that might have an impact on reading comprehension, at little to no cost to schools or teachers—a streamlined version of Kim (2006). Common summer activities related to reading comprehension, such as summer school, include labor and resource costs, as well as other concerns for schools, most of which are financially strapped at the current time. Trips to libraries, though advantageous, may present logistical needs in terms of transportation and scheduling (McGill-Franzen and Allington 2003). Another potential limitation of library trips with volunteers is that these trips typically encourage students to voluntarily select books and as such, students may select books that are not appropriate for their reading levels. Therefore, the program examined in this study removed certain components of the intervention developed by Kim and colleagues, in particular ones that involved parent or teacher training and teachers spending instructional time preparing students for summer reading. An advantage of such a program is that it can be implemented by a charitable group, publisher, or some other form of voluntary group independent of schools or certified teacher trainers. The lack of prior investigations into the impacts of varying degrees of teacher and parent components on student reading comprehension, mixed with the interest in cost-saving and burden-reducing efforts, led us to focus on the key components (books matched on reading level and interest) of the intervention developed by Kim and colleagues (Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008).

This report presents estimates from a large-scale, multi-district RCT on the effectiveness of a summer reading program on improving student reading comprehension for economically disadvantaged grade 3 students reading below the 50th percentile nationally. Each student in the treatment group was sent a single shipment of eight books matched to his or her reading level and interest area during the first part of the summer (June/July), followed by a reminder postcard each week for six weeks..

Reading comprehension and economically disadvantaged students

A decline in student reading comprehension over the summer months has been well documented, beginning with analyses of the initial evaluation data from Title I (David, 1979) and continuing for the next several decades (Allington and McGill-Franzen 2003; Bracey 2002; Heyns 1987; Luftig 2003; Malach and Rutter 2003). Cooper et al. (1996) reviewed 39 studies of summer reading loss and conducted a meta-analysis of 13 studies with sufficient data for analyses; 11 of these studies reported effect sizes. Those 11 studies examined the relationship between gender, ethnicity, and economic status and summer reading loss for more than 40,000

students. Neither gender nor ethnicity was shown to be related to summer reading loss, but family economic status was. While middle-income students showed gains in some measures of reading achievement over the summer, economically disadvantaged students consistently showed losses. More recent work (Alexander, Entwisle, and Olson 2007) parallels these findings.

The disproportionate effect on economically disadvantaged students has been posited as one reason for the persistence of the reading achievement gap between economically disadvantaged and nondisadvantaged students (Alexander, Entwisle, and Olson 2007; Cooper et al. 1996; David 1979). Academic difficulties among economically disadvantaged students may be due to less exposure to complex language at home (Hoff 2003; Hart and Risley 1995), fewer materials in the home that stimulate learning (Conger and Donnellan 2007; Constantino 2005; Neuman et al. 2001), and limited summer reading activity (Anderson and Stokes 1984; Carlson 1998; Storch and Whitehurst 2001; Vernon-Feagans et al. 2001). Multiple studies have further demonstrated that economically disadvantaged students have limited access to books (Allington, Guice, Baker, Michelson, and Li 1995; Anderson and Stokes 1984; Constantino 2005; Fryer and Levitt 2002; Heyns 1978; McGill-Franzen, Lanford and Adams 2002; McGill-Franzen et al. 2005; Neuman 1986; Neuman et al. 2001).

The association between economic disadvantage, limited access to books, and lower reading comprehension makes it important to examine the effects of a summer reading program targeting economically disadvantaged students.

Economically disadvantaged students are often identified by enrollment in the National School Lunch Program, frequently referred to as the free or reduced-price lunch (FRPL) program (Harlwell, Maeda, and Lee 2004) [2]. The percentage of students eligible for FRPL rose from 2003/04 to 2007/08 nationally and in all five states served by Regional Educational Laboratory (REL) Southwest (Arkansas, Louisiana, New Mexico, Oklahoma, and Texas; National Center for Education Statistics n.d.). In addition, in all five states served by the REL, FRPL rates have consistently been above the national average (table 1-1). Summer reading loss among economically disadvantaged students may, therefore, be particularly important to educators and policymakers in the Southwest Region.

[2] To receive FRPL benefits, a household must submit an application and its eligibility must be verified by local education officials (U.S. Department of Agriculture 2008). Although some households that apply and are approved may not actually receive benefits, the households are still counted as part of the program and are identified as "enrolled" in this report. Because of differences in how states gather and represent these data, the National Center for Education Statistics reports and labels FRPL data as "eligibility" rates (the percentage of students in a school who are either eligible or actually enrolled in the FRPL program aggregated at the state and national level). But because of the inherent difficulty in identifying students who are eligible for FRPL but have not applied, it is likely that—regardless of the source—the data reflect enrolled students.

Table 1-1. Percentage of students eligible for the free or reduced-price lunch program nationally and in Southwest Region states, 2003/04–2007/08

Location	*2003/04*	*2004/05*	*2005/06*	*2006/07*	*2007/08*
United States	36.2	37.2	41.4	41.2	40.4
Southwest Region					
Arkansas	49.8	51.9	52.8	58.6	56.2
Louisiana	61.4	61.6	61.2	61.6	63.2
New Mexico	58.2	58.1	55.7	60.4	60.6
Oklahoma	53.0	53.9	54.5	55.2	55.2
Texas	46.7	47.6	48.2	47.2	47.7

Source: National Center for Education Statistics n.d.

Research evidence on the effect of a summer reading program

Increasing access to books and reading material over the summer has long been advocated for reducing summer reading loss (Elley and Mangubhai 1983; Krashen 1995; Neuman et al. 2001). Since at least the early twentieth century, public libraries have had summer reading programs that bring books to children, both nationally (Woodward 1901) and in the Southwest Region (Anonymous 1910). Nearly 90 years later, the U.S. Department of Education found that 95 percent of public libraries were still offering some form of summer reading program (Fiore 1998).

In the Southwest Region, both library- and school-based summer reading programs are in operation. Between 2002 and 2009, state libraries in four of the five Southwest Region states joined the Collaborative Summer Reading Program, a national consortium of states whose goal is to provide summer reading materials to children through public libraries. In 2009, the Albuquerque/Bernalillo County (New Mexico) Library System sponsored an eight-week summer reading program offering weekly prizes for students who kept a log of the books they read and showed it to library staff (Albuquerque/Bernalillo County Library System 2010). Other libraries have had similar programs (for example, Dallas Public Library 2010; New Orleans Public Library 2010). School districts and state departments of education have also instituted school-based summer reading programs that range from providing students with a list of suggested books (Dallas Independent School District n.d.; Louisiana Department of Education n.d.) to coordinating with local libraries to "keep [their] students reading over the summer" (Houston Independent School District 2010, para. 1).

Matching books to student reading level and interest areas

The central component of summer reading programs is providing access to books, but methods vary for determining which books to provide (Fountas and Pinnell 1996, 2001; Leslie and Caldwell 2006). Books are often matched to students on two dimensions: book reading level and student interest (Fountas and Pinnell 1996).

For more than a century, researchers have sought to match students and books at an appropriate reading level to enable students to read independently (Burns, Griffin, and Snow 1999; Fry 2002; McGill-Franzen and Allington 2003). According to Fry (2002), the first widely used set of leveled readers (readers ordered by reading difficulty) was developed by McGuffey in 1836, and the first formula for leveling (determining the difficulty level of books) was developed by Lively and Pressey in 1923.

A recent method of matching students and books by reading level is the Lexile Framework® for Reading, developed by MetaMetrics, Inc. The Lexile theory operationalizes the semantic and syntactic components of reading by using mean sentence length and $\log_{10}$ [3]word frequency (Stenner et al. 1987). More specifically, the semantic component is gauged using proxies for word frequency—average number of letters or syllables per word. These proxies capitalize on the high negative correlation between length of words and frequency of word usage. Long words and polysyllabic words are used less frequently than short monosyllabic words, making word length a good proxy for the likelihood of an individual being exposed to them. Syntactic complexity is gauged using sentence length.

This linguistic, theory–based method measures student reading comprehension and the reading difficulty of texts using a common scale unit called a Lexile (L), which ranges from 0L for emerging readers and beginning texts to 1700L for advanced readers and texts (MetaMetrics, Inc. n.d.d). The Lexile measure of a book is calculated by parsing the text into 125-word slices and using a proprietary regression equation to assign a reading difficulty value to each slice based on word frequency and sentence length. Combining results across slices yields the overall Lexile measure for the book. According to a 2001 report from a panel of reading experts working with the National Center for Education Statistics, the Lexile Framework has been found to have solid psychometric properties and has been validated across a range of populations (White and Clement 2001). Lexile measures are reported as a measure of student reading ability on state assessments in four of the five Southwest Region states (Arkansas, New Mexico, Oklahoma, and Texas) and on a number of nationally normed tests such as the Stanford Achievement Test Series and the TerraNova series (MetaMetrics, Inc. n.d.b). (See appendix A for more information about the Lexile Framework, including examples of books at various Lexile measures.)

[3] Each word in the MetaMetrics database of scanned texts has a frequency, which represents the proportion of times the word is used in the entire database. For each word used in the Lexile analysis, the $\log_{10}$ function is applied to each word's frequency.

Researchers argue that, in addition to matching students and books on reading level, providing students with books that interest them is also important (Allington et al. 2010; Burns, Griffin, and Snow 1999; Kim 2006; McGill-Franzen and Allington 2003; McQuillan and Au 2001; Mraz and Rasinski 2007; Schiefele 1999). For example, Burns, Griffin, and Snow (1999, p. 10) state that "throughout the early grades, time, materials, and resources should be provided to support daily independent reading of texts selected to be of particular interest for the individual student." One method of matching students to books that appeal to them is to survey students about their interests. Another method is to allow students to select their own books. A study of 501 students ages 5–17 and their parents in 25 U.S. cities found that 89 percent of students reported that their favorite books are those they choose themselves (Scholastic, Inc. 2008).

Previous studies of summer reading programs using book matching

Several recent studies, including ones that used randomized controlled trials (RCT; Allington et al. 2010; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2008) and others that used quasi-experimental designs[4] (Butler 2010; Crowell and Klein 1981) have examined effects of summer reading programs that supply students with books matched on reading level, interest area, or both, on student reading achievement. The book-matching approaches varied across all studies. Some studies matched students and books on both reading level and interest area, whereas others matched students and books only on one of these dimensions. The studies also differed in the student populations examined and in the additional components provided to students. Table 1-2 summarizes the studies discussed in this section; detailed results for each study, and information on formulas used to recalculate study findings as effect sizes, are provided in appendix B.[5]

[4] For the purposes of this report, a conservative approach to classifying studies was used, and Butler's (2010) study was classified as quasi-experimental. This study design could be referred to as a hybrid since random assignment was used for some (but not all) students in the sample. As such, causal inferences cannot be made for all comparisons between groups. For the purposes of this report, Crowell and Klein's study (1981) is considered quasi-experimental because of the way students were assigned to either the treatment or "comparison" group. Students were rank ordered based on their score on a criterion-referenced reading test and teacher judgment of their reading levels. Students were then systematically assigned in an alternating fashion based on their rank order rather than being randomly assigned.

[5] An effect size is a standardized measure of the strength of an outcome. Many of the studies described here do not provide effect sizes; for these, effect sizes were estimated from the information provided. See appendix B for the formulas used to estimate a standardized mean difference (the difference in standard deviation units). For example, an effect size of 0.5 implies that the difference in group means is one-half of the pooled standard deviation.

Table 1-2. Comparison of studies of summer reading programs using matched books

Component	*Allington et al. (2010)*	*Butler (2010)*	*Crowell and Klein (1981)*	*Kim (2006)*	*Kim (2007)*	*Kim and Guryan (2010)*	*Kim and White (2008)*	*Current study*
Study design	Randomized controlled trial	Quasi-experimental	Quasi-experimental	Randomized controlled trial	Randomized controlled trial	Randomized controlled trial	Randomized controlled trial	Randomized controlled trial
Student grade level[a]	3–5	2–4	1–2	4	1–5	4	3–5	3
Total analytic sample size	1,330	94	48	486	279	324	400	1,581
Statistically significant findings[b]	Reading comprehension (full sample and economically disadvantaged students)	Oral reading fluency and oral retelling (full sample)	Reading vocabulary (full sample)	Reading achievement (Black students, below median readers, and those with <50 books in home)	Number of books read (full sample)	Number of books read (full sample)	Total reading achievement (full sample)	Number of books read (full sample)
Non-significant findings[b]	NA	Oral reading fluency and retelling (home visits vs. books only, race vs. assignment, reading level vs. assignment)	Reading vocabulary (2^{nd} grade students)	Reading achievement (treatment, treatment* income, treatment* ethnicity* income, Asian, Hispanic, White, above median readers), Reading Fluency (above median readers), <100 books at home,	Number of books read (treatment* grade, treatment* FRPL), reading achievement (treatment, treatment* grade, treatment* FRPL)	Number of books read (treatment vs. family literacy), total reading achievement, reading vocabulary, reading comprehension	Reading achievement (books with oral reading and comprehension scaffolding versus books only, books with oral reading and comprehension scaffolding versus books with oral reading	Reading achievement

Component	*Allington et al. (2010)*	*Butler (2010)*	*Crowell and Klein (1981)*	*Kim (2006)*	*Kim (2007)*	*Kim and Guryan (2010)*	*Kim and White (2008)*	*Current study*
				>100 books at home, > 50 books at home			scaffolding)	
Race/ethnicity (%)								
Asian/Pacific Islander	0	7	100	17	0	0	8	4
Black	89[c]	0	0	19	0	0	25	21
Hispanic	0	57	0	26	0	90	29	67
White	5	36	0	33	42	0	31	7
Other	6	0	0	5	58	10	7	1
Free and reduced-price lunch program (%)	65–99 across schools	100	Not provided	39	23	96	38	100
Baseline reading levels	Not provided	Varied	Varied	Varied	Varied	Varied	Varied	Low
Books matched on								
Reading level	X	√	√	√	√	√	√	√
Interest area	√	√	X	√	√	√	√	√
Number of books	12	10	10	8	10	10	8	8
Postcards	X	X	X	√	√	√	√	√
Parent letters	X	X	√	√	√	√	√	X
Teacher support								
Postcard direction	X	X	X	√[d]	√[d]	√[d]	√[e]	X
Reading encouragement	X	√	X	√[d]	√[d]	√[d]	√[e]	X
Teacher professional	X	X	X	√[d]	X	√[d]	√[d]	X

Component	*Allington et al. (2010)*	*Butler (2010)*	*Crowell and Klein (1981)*	*Kim (2006)*	*Kim (2007)*	*Kim and Guryan (2010)*	*Kim and White (2008)*	*Current study*
development								
Reading instruction								
Oral reading	X	X	X	√[d]	X	√[d]	√[f]	X
Comprehension	X	X	X	√[d]	X	√[d]	√[f]	X
Paired reading	X	X	X	√[d]	X	√[d]	√[f]	X

√: component was included in a study; X: component was not included in a study; NA: Not Applicable

a. The grade in which a student was enrolled prior to the summer intervention.

b. See appendix B for a complete description of the findings.

c. The sample is described as 89 percent African-American or Hispanic.

d. These components were provided to treatment and control groups.

e. The study included three treatment groups and one control group; students in all three treatment groups received postcard direction and reading encouragement.

f. Students in all four study groups (three treatment groups and one control group) received some type of reading instruction, which varied based on group assignment.

Source: Authors' analysis.

Matching on reading level only

Crowell and Klein (1981) evaluated a summer reading program that matched students to books on reading level but not interest area. The 50 grade 1 and grade 2 students were ranked according to a combination of information provided by the teacher including their judgment of each student's reading level and the student's reading level as measured by a periodic criterion-referenced test. From this rank order, students were sorted alternately into either the treatment or comparison group (25 treatment and 23 comparison students in all), resulting in treatment and comparison groups with the same proportion of grade 1 and grade 2 students. For the purposes of this report, Crowell and Klein's study (1981) is considered quasi-experimental because of the way students were assigned to either the treatment or "comparison" group. Students were rank ordered based on their score on a criterion-referenced reading test and teacher judgment of their reading level. Instead of using random assignment, students were systematically assigned in an alternating fashion based on their rank order.

Treatment students were sent 10 books "sorted by level of difficulty so that children would receive books appropriate to their reading levels" (p. 562). Parents of treatment students were instructed to encourage their children to read over the summer. Two outcomes were investigated: vocabulary and reading comprehension scores. The study examined grade 1 and grade 2 treatment groups separately and combined.

The analysis of covariance showed statistically significant improvements in vocabulary scores for students who received books for both the combined treatment group (effect size >0.67,[6] p <0.05) and grade 1 treatment group (effect size >1.03, p<0.05). There was not a statistically significant difference in vocabulary scores between the treatment and comparison students in the grade 2 sample.[7] No statistically significant impacts were observed for the reading comprehension outcome measure for any group.

Matching on interest area only

Allington et al. (2010) evaluated a summer reading program that matched students to books on interest area but not reading level. The study spanned three years and included 1,330 students (852 treatment, 478 control). Students were in grades 1 and 2 at the beginning of the study and most of the students were in 4th or 5th grade at the end of the study, although a small number were in 3rd or 6th grade because of differences in retention and promotion of students were. Treatment students were sent 12 self-selected books per summer; no additional components were provided. Control students did not receive books. Reading comprehension was the outcome of interest.

[6] Because of the nature of the information provided in Crowell and Klein (1981), it was possible to estimate only a lower bound for the effect size.

[7] Crowell and Klein (1981) provide results only for significant outcomes, so it is not possible to estimate the effect size for the other outcomes.

A statistically significant effect was identified for the treatment students at the end of the third year (effect size=0.13, p=0.015). There was also a positive and statistically significant effect reported for FRPL treatment students (effect size=0.20, p=0.001).

Matching on both reading level and interest area

Several recent studies have examined summer reading programs that provided students with books matched on both reading level and interest area (Butler 2010; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008). One study (Butler 2010) was conducted in a single school. The others (Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008) were conducted in a single district.

In Butler (2010), 94 grade 2–4 students enrolled in FRPL were assigned[8] to one of three groups (53 in two treatment groups and 41 in a control group). The first treatment group of students (n=27), received 10 leveled books of interest as well as teacher visits over the summer, were assigned based on proximity to the teacher travel routes, not randomly. The remaining students (n=67) were randomly assigned to either the second treatment group (n=26), which received 10 leveled books of interest over the summer, or the control group (n=41), which received no books. For the purposes of this report, a conservative approach to classifying studies was used, and Butler's (2010) study was classified as quasi-experimental. This study design could be referred to as a hybrid since random assignment was used for some, but not all, students in the sample. Students in all three groups received reading logs and were encouraged to record the dates they read and amount of time they spent reading. Students in the two treatment groups were provided books matched on reading level and interest area. For students currently in guided reading groups, books were matched to students' reading level using Fountas and Pinnell's (1996) guided reading levels;[9] otherwise, teacher input and books students were currently reading in the classroom were used in matching. A self-selection method in which students choose from a collection of books in the range of their reading level was used to match books to students' interest area. Students in the first treatment group (n=27) chose 10 books and received all of them at the end of the school year. Students in the second treatment group (n=26) were visited once a week by a school staff member and chose one or two books per visit from a selection already matched on reading level; their reading logs were also reviewed during these visits. The outcomes of interest in this study were oral reading fluency and oral retelling.

Based on a two-way analysis of covariance, Butler (2010) reported on six main comparisons, four of them comparing students who received books with those who did not. These four comparisons showed that, after adjusting for student baseline scores, the students who

[8] Students in one treatment group were selected nonrandomly because their home addresses placed them on convenient driving routes for school staff who delivered books weekly to that group. The students not on those routes were randomly assigned to either the other treatment group (whose members took books home at the end of the school year) or the control group (who received no books).

[9] Fountas and Pinnell (1996) provide a list of books by reading level for guided reading. Guided reading aims to move students toward successful independent reading by teaching them a set of reading strategies, which students then apply to texts of increasing difficulty.

were part of the first treatment group (matched books only) had significantly higher scores in oral reading fluency (effect size=0.65, $p<0.0001$) and oral retelling (effect size=0.77, $p<0.0001$) than did students in the control group. Similarly, after adjusting for student baseline scores, students who were part of the other treatment group (matched books and weekly staff visits) had significantly higher scores in oral reading fluency (effect size=0.59, $p<0.001$) and oral retelling (effect size=0.65, $p<0.001$) than did students in the control group. (Details of all of the comparisons are provided in appendix B.)

The series of studies by Kim and colleagues examining summer reading programs that matched books on both reading level and interest area shared several features: (1) they were each conducted in a single district, (2) they used the same method to match students and books on reading level (Lexile measures), and (3) they included postcards mailed to students, letters to parents explaining the program, and teacher support encouraging students to read over the summer.

In the first of these summer reading studies, Kim (2006) examined the impact of a summer reading program on 486 grade 4 students (252 treatment and 234 control). The control group was not a business-as-usual condition. Students in both treatment and control groups received teacher support and teacher instruction as part of the study-provided curriculum, and students in the control group received books and postcards in the fall rather than in the summer. Classroom teachers participated in a two-hour training session on teaching reading techniques to students and motivating them to read the books they would be sent during the summer (treatment group) or the fall (control group). During the last two weeks of school, students in both treatment and control groups received lessons in which teachers modeled comprehension strategies, practiced paired reading,[10] and assigned homework that gave students practice in filling out postcards. During the summer, students in the treatment group were sent eight matched books, followed by postcards that reminded them to practice the oral- and silent-reading strategies they were taught and encouraged them to read aloud with a parent. Students were matched to books on interest level using student interest surveys. The outcome of interest was reading achievement (a comprehension and vocabulary composite). Student subgroups (race/ethnicity, FRPL status, access to books, summer school attendance, baseline reading fluency, and baseline reading comprehension) were defined prior to randomization.

The summer reading program was not found to have a statistically significant effect on reading achievement across the entire treatment sample (effect size=0.17, $p=0.059$), but it did have a statistically significant effect on the scores of three subgroups: Black students (effect size=0.24, $p=0.011$), students with baseline reading fluency scores below the median (effect size=0.35, $p=0.008$), and students with fewer than 50 children's books in their home (effect size=0.29, $p=0.020$). Because the control group also received teacher support and instruction,

[10] Topping (1987) defines paired reading (sometimes called *repeated reading*) as an oral fluency strategy that involves a student reading sections (of approximately 100 words) of a book to an adult (parent or family member). Each section is to be read twice, with a discussion about the student's improvements in fluency during the second reading.

this study estimated the effect of a summer reading program only consisting of matched books and letters/postcards.

In the second study, Kim (2007) examined the impact of a summer reading program on 279 students (138 treatment and 141 control) in grades 1–5. As in the earlier Kim study (2006), students in both the treatment and control groups received teacher support. Teachers were told that during the last week of school, they should briefly explain the study to students, encourage all students to read over the summer, teach students how to fill out and return postcards, and encourage them to return the postcards. Students in the treatment group were sent 10 matched books, along with postcards, over the summer. Students in the control group were sent matched books and postcards in the fall, after posttest data collection was complete. Students were matched to books on interest area using student interest surveys. The key outcomes evaluated were reading achievement (a comprehension and vocabulary composite) and number of books read.

This summer reading program had no statistically significant effect on reading achievement (effect size=0.10, p=0.36) but did have a statistically significant positive effect on the number of books read (effect size=0.76, p=0.001). Students in the treatment group reported reading approximately three more books on average than students in the control group.

A third study (Kim and White 2008) randomly assigned teachers and 400 grade 3–5 students to one of four conditions (293 in three treatment groups and 107 in one control group). Students in each treatment group were sent eight matched books and postcards over the summer. Students in the control group were sent matched books and postcards in the fall, after posttest data collection was complete. Students were matched to books on interest area using student interest surveys. The three treatment groups varied in the kind of reading strategy lessons the teachers provided. Treatment group 1 (n=93) received one scripted lesson, treatment group 2 (n=100) received two scripted lessons, and treatment group 3 (n=100) received three scripted lessons. Control group students received no scripted lessons. (Details about the different types of lessons are in appendix B.) Students in all three treatment groups were sent one book a week, for a total of eight books, accompanied by a parent letter and student postcard; the parent letters for group 1 suggested that parents encourage student reading; the parent letters for groups 2 and 3 suggested that parents listen while their student read aloud. The key outcome evaluated was reading achievement (a comprehension and vocabulary composite).

The study reported on four comparisons, only one focused on students who received books and students who did not. That comparison showed that the students in treatment group 3 (matched books and three scripted lessons) had a statistically significant difference in reading achievement (after adjusting for students' baseline scores) than did students in the control group (effect size=0.24, $p<0.03$). The other three comparisons were of treatment groups 1 versus 3, treatment groups 2 versus 3, and treatment groups 2 and 3 combined (the groups with scaffolding combined)versus treatment group 1 and the control group (the groups without scaffolding combined). Details are in appendix B.

In the fourth study (Kim and Guryan 2010), 324 low-income grade 4 Hispanic students were randomly assigned to one of three groups (two treatment and one control). Students in all three groups received teacher instruction in three scripted lessons. (See appendix B for details.) Students in both treatment groups (n=214) self-selected 14 books that interested them and were sent 10 based on their Lexile measure. The self-selection resulted in 67 percent of treatment students (across both groups) receiving books below their Lexile measure. Students in treatment group 1 (n=102) received no additional components. Students in treatment group 2 (n=112) were invited to attend three two-hour family literacy events. Students in the control group (n=110) received 10 books in the fall, after posttest data collection was complete. Four outcomes were investigated: reading comprehension, vocabulary, reading achievement (a comprehension and vocabulary composite), and number of books read.

The summer reading program did not have a statistically significant impact on student reading comprehension (effect size=0.10, $p>0.05$), vocabulary (effect size=0.24, $p>0.05$), or reading achievement (effect size=0.10, $p>0.05$). It did have a statistically significant effect on the number of books students reported reading over the summer (effect size=0.39, $p<0.001$).

Summary of previous studies of summer reading programs using book matching

Of these seven studies, three found no statistically significant relationships with reading achievement for the full sample (Kim 2006, 2007; Kim and Guryan 2010), and four found statistically significant improvements (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim and White 2008). Two (Allington et al. 2010; Kim 2006) found that particular subgroups—such as economically disadvantaged students and students reading below the median—may be more affected by summer reading programs than other subgroups. Five of these studies used RCT designs (Allington et al. 2010; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008), three of which found statistically significant effects on reading comprehension for the full sample (Kim 2006, 2007; Kim and Guryan 2010). Two of the five studies utilizing an RCT design found that students in the treatment group read significantly more books than did those in the control group (Kim 2007; Kim and Guryan 2010); however there was no significant difference in reading achievement.

Although these seven studies reported mixed results, it is possible that summer reading programs using matched books might improve reading comprehension. These studies were

limited in several ways. Two of the four studies that identified statistically significant improvements for the full sample used quasi-experimental designs (Butler 2010; Crowell and Klein 1981). Six were single-district studies (Allington et al. 2010 included two districts). The summer reading programs examined in five of the studies (1 quasi-experiment and 4 RCTs) included teacher components (such as training and classroom instruction) or parent components (such as family literacy events) that could be costly in terms of time, training, and resources to implement on a large-scale (Butler 2010; Kim 2006, 2007; Kim and White 2008; Kim and Guryan 2010). Even though research indicates that economically disadvantaged students are disproportionately affected by summer reading loss (Cooper et al. 1996; Alexander, Entwisle, and Olson 2007), only one study required student enrollment in FRPL for study eligibility (Butler 2010). Finally, none of the studies limited the sample to students reading below the 50th percentile nationally (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008).

Current study

The current study is an RCT examining the effect of a summer reading program on student reading comprehension after sending eight books matched on reading level and interest area in a single shipment[11] to each student in the treatment group during the summer, followed by a reminder postcard each week for six weeks. Eight books were chosen because this was the number used in two of the summer reading programs shown to have statistically significant positive effects on reading comprehension for specific subgroups (Kim 2006) or for the entire sample (Kim and White 2008); the choice of six postcards was made because six weeks remained before school started once the books were shipped.

The components selected for the current study did not require much involvement from school personnel and could be administered, for example, by a charitable or volunteer group working with a public library. Schools distributed parent consent forms and provided a roster of consenting students, their addresses, and a method for gauging their reading level (such as a score from a standardized assessment) and interest areas (such as a student interest survey). Teachers were not required to attend training, encourage students to read over the summer or return postcards, or provide instruction during the school day that focused on reading strategies beyond what was included in the standard district curriculum.

Study sample

The study was conducted in the summer of 2009. It focused on economically disadvantaged students in four Texas school districts who were reading below the 50th percentile

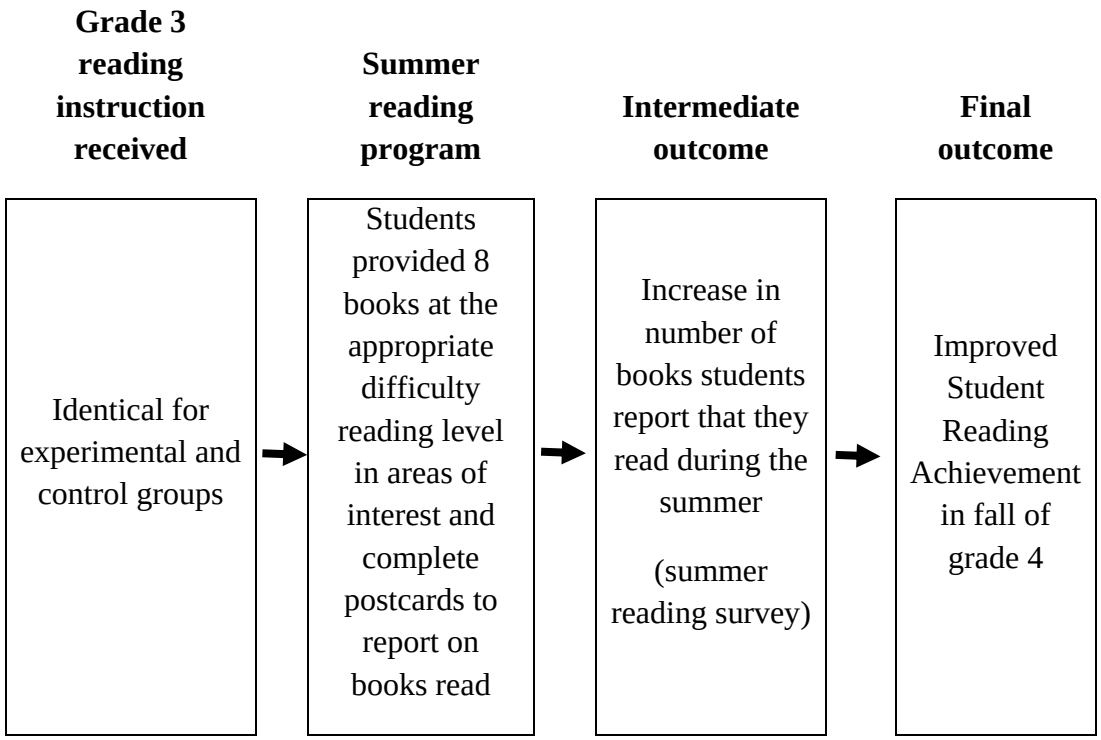
[11] The shipment included a letter describing the summer reading program and encouraging the student to read the books and return the postcards. The postcards, which were pre-addressed and had return postage, asked about students' summer reading activities and encouraged them to read. Copies of the letter and postcard are in appendix C.

nationally. These design decisions responded to recommendations by Kim (2006, 2007) that economically disadvantaged students be the focus of future research and that more than one district be examined. The final analytic sample, which included 1,571 students in 110 elementary schools, was larger than that of any of the previous seven studies, which ranged from 48 students to 1,330. This study focused on students transitioning between grade 3 and 4 for three reasons: (1) independent reading demands increase dramatically in grades 3 and 4 (Chall 1983; National Research Council 1998), (2) the grade 3 to grade 4 transition was not a focus of previous studies, and (3) Texas state assessment data is available for the first time for students beginning in grade 3, which was used to control for baseline differences in this study. The National Research Council report (1998, p. 207) on beginning reading notes that in "fourth grade and up, it is taken for granted that [students] are capable—independently and productively—of reading to learn."

Theory of change

The proposed theory of change, illustrated in Figure 1-1, suggests that students' reading both generally (Burns, Griffin, and Snow 1999; Manning, Lewis, and Lewis 2010; Hiebert and Reutzel 2010) and at their reading level (Allington et. al 2010; Butler 2010; Kim 2006; Kim 2007; Kim and Guryan 2010; Kim and White 2010) increases reading achievement. This theory is also informed by research, which demonstrates a relationship between having limited access to books and summer reading loss (Cooper et al. 1996; Alexander, Entwisle, and Olson 2007), as well as the potential positive effects of a summer reading program on student reading achievement (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim and White 2008).

Figure 1-1. Theory of change



Because summer reading loss disproportionately affects economically disadvantaged students (Cooper et al. 1996; Alexander, Entwisle, and Olson 2007), it was determined that the selected summer program should focus specifically on potential impacts for economically disadvantaged students. The selected students also performed below the 50th percentile in reading comprehension nationally at the end of grade 3, as measured by their Lexile score on the TAKS-Reading assessment. Our theory is that providing books for economically disadvantaged students reading below the 50th percentile nationally may help improve their reading comprehension.

Students in the study sample were randomly assigned within each school to the summer reading (treatment) group or to the business as usual (control) group. Thus, the actual quality of reading instruction in the spring of grade 3 and the fall of grade 4 is statistically similar for students in the treatment and control groups. The impact of the summer program on reading comprehension was estimated using the grade 4 fall reading assessment data. Grade 3 reading comprehension was controlled for in order to obtain a more precise impact estimate

Research questions

The current study was designed to answer the following confirmatory research question:

- For economically disadvantaged students reading below the 50th percentile nationally, does being sent eight free books during the first part of the summer (June/July) matched to reading level and interest area, along with six reminder postcards, result in significantly better reading comprehension scores in the fall?

Reading comprehension at posttest was measured using Scholastic Reading Inventory (SRI; Scholastic, Inc. 1999) reading comprehension scores based on administration during the fall semester of grade 4. SRI is a computer-administered reading assessment for students in grades K–12 that measures reading comprehension on the Lexile scale.

In addition to the primary confirmatory research question, the study addressed two exploratory research questions:

- Did students in the summer reading program treatment group report reading more books over the summer than students in the control group?

- Did the summer reading program have differential effects on reading comprehension depending on baseline reading proficiency?

Structure of the report

Chapter 2 focuses on the design and methodology of the current study, including recruitment and sampling, random assignment, data collection measures, and data analysis. A study timeline is also provided. Chapter 3 focuses on fidelity of the summer reading program implementation and includes information on material costs of the program. Chapter 4 provides the empirical findings for the confirmatory research question and the results of sensitivity analyses. Chapter 5 provides the empirical findings for the exploratory research questions. Chapter 6 summarizes key findings, discusses study limitations, and suggests directions for future research.

Chapter 2. Study design and methodology

This study is a multidistrict randomized controlled trial (RCT) evaluating the effect of a summer reading program on the reading comprehension scores of economically disadvantaged third grade students reading below the 50th percentile nationally. The analytic sample included 1,571 students in 110 elementary schools in four Texas school districts.

Participating students were randomly assigned to treatment or control groups within schools (half to each condition within each school). During the summer of 2009, students in the treatment group were sent eight books matched to their reading level and interest areas (matched books) in one shipment along with an explanatory letter, followed by a postcard each week for six weeks. Neither books nor postcards were sent to students in the control group during the summer, but eight matched books were sent to them after posttesting was completed in the fall. Participating students completed the Scholastic Reading Inventory (SRI; a reading comprehension measure) and a summer reading survey during the following fall semester.

The impacts of the summer reading program on reading were estimated using ordinary least squares (OLS) regression

Study timeline

The summer reading program was designed and conducted as a single summer intervention. Table 2-1 (following page) shows start and completion dates for project milestones.

Table 2-1. Study timeline, 2009

Event	*Start*	*Completion*
Eligible students identified by districts	May	May
Parent consent and student interest survey collected	May	June
Random assignment conducted[a]	June	June
Books shipped to students in the treatment group (one shipment)	July	July
Postcards sent (6 mailings, 1 per week)	July	August
Scholastic Reading Inventory (SRI) and summer reading surveys administered[b]	September	December
Books shipped to students in the control group (one shipment)	December	December

Note: Students in the control group were not sent matched books until all posttest data were collected.

a. Random assignment was conducted after parent consent and study interest survey collection was complete.

b. For logistical reasons (purchasing, installing, and administering the SRI across 128 schools), SRI administration occurred at different times in different schools in the fall semester. For most posttested students (*n*=1,002; 63.8 percent), the SRI was administered in October; others were tested in September (*n*=246; 15.7 percent), November (*n*=299; 19.0 percent), or early December (*n*=24; 1.5 percent).

Source: Study records collected May 2009–December 2009.

Target sample

The target sample size for the current study was established based on a target minimal detectable effect size (MDES), which represents the smallest *true* program impact in standard deviation units that can be detected with high probability (Bloom 2005). Based on prior research (Kim 2006, 2007; Kim and White 2008) and feedback from the current study's technical working group members, an effect size of approximately 0.12 was determined to be appropriate for testing the confirmatory hypothesis (see appendix D). We expected low parent consent rates, because consent forms were distributed at the end of the school year, with little time for teacher follow-up. This led to a target recruitment sample size of 3,032 students to yield an analysis sample of 1,516 (758 treatment and 758 control students).

Table 2-2. Participation criteria by district, school, and student

Level	*Participation criteria*
District	• Minimum of 25,000 students • Minimum of 300 students districtwide who meet the economically disadvantaged students criterion (free or reduced-price lunch, FRPL) • Signed letter of intent
School	• Minimum of two eligible students with parent consent
Student	• Enrollment in FRPL • Lexile measure on spring 2009 grade 3 English-language Texas Assessment of Knowledge and Skills (TAKS)–Reading at or below 590L • Physical and mental ability to engage in independent reading as determined by teacher or school • Parent consent

Source: Study records.

Participating districts and schools

To ensure all participating districts would share the same baseline measure, the study was implemented in a single state. Texas was selected because the Texas Assessment of Knowledge and Skills (TAKS), a statewide assessment given to all Texas public school students in grades 3–11, has been linked to the Lexile scale.[12] This study used Lexile measures to determine whether students were reading below the 50th percentile nationally (an eligibility criterion for study participation) and to identify books at the appropriate reading level for participating students.

Recruitment was restricted to Texas school districts with 25,000 or more enrolled students to limit the number of school districts needed to attain the large student sample, minimize project management effort, and contain costs. Forty-two districts that met these enrollment criteria were solicited to participate in this study; 8 expressed interest, 29 did not respond, and 5 declined for various reasons or were excluded because they did not have enough students who met eligibility requirements. Of the eight interested districts, four were eliminated based on either failure to complete a formal letter of intent—the next stage in the recruitment

[12] The Texas Education Agency requested that MetaMetrics, Inc. link TAKS scores to Lexile measures (MetaMetrics, Inc. 2005). As a result, student score reports include Lexile measures corresponding to their TAKS scaled scores (Texas Education Agency 2005a). For additional information about this linking study, see appendix E.

process—or other factors, including responsiveness to communications and concurrent participation in other research studies. Four districts were selected for participation. (Details on recruitment are in appendix F; district profiles are in appendix G.)

Districts determined which elementary schools would participate in the study using district-generated criteria not disclosed to the study team. The percentage of elementary schools selected for participation ranged from 34 percent to 100 percent across the four districts. Consent forms were provided to all selected schools in the four districts, but only a subset of schools selected for participation returned parental consent forms. Altogether, 117 schools had eligible students who received consent to participate in the study. However, within-school student-level random assignment meant that each school had to have at least two participating students. Five schools had only one eligible student each who received consent to participate, so the baseline study sample included 112 elementary schools; district participation rates ranged from 27 percent of elementary schools to 100 percent. It should be noted that this is not a random sample of schools or districts.

Student eligibility criteria

Students were required to meet three eligibility criteria: enrollment in the free or reduced-price lunch (FRPL) program, a Lexile measure on the spring 2009 grade 3 English-language TAKS–Reading at or below 590L,[13] and physical and mental ability to read independently. Enrollment in FRPL was used to identify economically disadvantaged students. A Lexile measure of 590L or below represents the 50th percentile for grade 3 on national norms linked to the Lexile scale (Scholastic, Inc. 2007). This Lexile measure is equivalent to a score of 2262 on the grade 3 TAKS–Reading (Texas Education Agency 2005b) and represents students in the 45th percentile for Texas students who took the spring 2009 administration of the TAKS–Reading. Because the summer reading program would not provide students with support, they needed to be physically and mentally able to read without assistance. This determination of ability was made by the homeroom teacher, who did not send home consent forms for any students judged not able to read independently[14]. A total of 3,289 grade 3 students across the four districts—ranging from n=395 to n=1,556—were identified as eligible for this study.

[13] There are several versions of TAKS. This report uses *TAKS* to refer to both the regular TAKS and the TAKS–Accommodated. The TAKS–Accommodated is available to students receiving special education services who meet eligibility criteria for specific accommodations such as larger font or fewer items per page. The content of the two tests is identical, and the specific accommodations do not invalidate interpreting scores from the two versions in the same way; results of these two TAKS versions are combined for state and federal accountability reporting (Texas Education Agency 2008a). To participate in this study, students were required to have a score on the English-language TAKS or TAKS–Accommodated. Students who took a Spanish-language or linguistically accommodated version of the TAKS were not eligible. More information about the different versions of the TAKS is in appendix H.

[14] The district liaison provided verbal guidance to the teachers not to hand out consent forms to students they thought were not physically or mentally able to read independently. However, given the short timeline (5 days)

Consent process

Because of study scheduling constraints, parent consent forms could not be sent home until the last week of the spring semester, leaving no time to follow up with parents to increase consent rates. In addition, some schools chose not to participate and did not return any consent forms. However, recruitment of schools was designed to accommodate a response rate as low as 50 percent. Despite the lack of time for follow-up, 57 percent (1,874) of consent forms were returned, with 54 percent (1,790) of eligible students returning consent forms—more than the target of 1,516 needed for the desired statistical power (table 2-3).

Table 2-3. Parent consent form responses for participating schools, by district and across all districts

	District				
Number of consent forms[a]	*A*	*B*	*C*	*D*	*Total*
Distributed to participating schools	651	1,556	272	395	2,870[b]
Consent refused/not returned	376	408	150	150	1080
Consent granted	275	1,148	122	245	1,790
Consent granted (percent of total forms distributed)	42	74	45	63	62

a. All schools with eligible students were sent consent forms, but not all schools decided to participate. So the consent form numbers in this table are only for participating schools.
b. Only 117 of 152 schools with eligible students agreed to participate in the study.
Source: Consent forms collected May 2009–June 2009.

Districts provided student-level data only for students who received parent consent to participate. (No data were provided for eligible students who did not return consent forms or for students who did not receive parent consent.) Therefore, differences across these groups of students cannot be examined. Return rates varied considerably across the four districts (from 42 percent to 74 percent), but discussions with officials in each district indicated that the consent process was similar, and there appeared to be no systematic reason for the variability.

Random assignment of students

Participating students within each school were randomly assigned to either the treatment or the control group using a random number algorithm in Microsoft® Excel. Random assignment was conducted by a statistician who had no involvement with implementation of the summer reading program. See appendix I for a detailed description of the randomization process. Because some schools had an odd number of students and randomization took place within schools, the

for completing the consent process, no attempt was made to evaluate the consistency or reliability of teacher judgments.

number of students in the treatment and control groups was not precisely equal, though the percentage of students in each group in the overall sample was approximately 50 percent (table 2-4). Although the participating students were randomly assigned to condition, it should be noted that the sample of students was not randomly selected from a larger population of students.

Table 2-4. Randomization rates by district and across all districts

	District				
Number of students	*A*	*B*	*C*	*D*	*Total*
Final sample	275	1,148	117	245	1,785[a]
Treatment group	138	575	60	123	896
Control group	137	573	57	122	889
Students in treatment group as percentage of final sample	50	50	51	50	50

a. Excludes five students who were the only eligible students with parental consent in their school.
Source: Study records collected June 2009.

Baseline equivalence

Each district provided an Excel file with student-level demographic data (school name, student date of birth, sex, race/ethnicity, Individualized Education Program status, [15] and English language learner status) for participating students.[16] Districts also provided scores from the spring 2009 grade 3 TAKS–Reading and the corresponding Lexile measures. Baseline equivalence was tested for these variables using either a *t*-test or a chi-square test. No significant differences were found between students in the treatment and control groups on any of the variables, indicating that the groups were equivalent at baseline (table 2-5 following).

[15] An Individualized Education Program document specifies learning goals and activities for each student receiving special education services.

[16] Note that these demographic data were used to evaluate baseline equivalence, but were not used to determine student eligibility.

Table 2-5. Baseline characteristics for treatment and control groups

Characteristic	*Treatment group mean*[a] *(*n*=896)*	*Control group mean*[a] *(*n*=889)*	*Difference in means*[b]	*Test statistic*[c]	p-*value*
Age in years	9.41 (0.47)	9.43 (0.50)	–0.02	–0.70	.48
Female	0.51 (0.50)	0.50 (0.50)	0.01	0.59	.56
Race/ethnicity[d]				6.39	.17
American Indian	0.00 (0.05)	0.00 (0.05)	0.00		
Asian	0.03 (0.18)	0.05 (0.21)	–0.01		
Black	0.23 (0.42)	0.18 (0.39)	0.05		
Hispanic	0.66 (0.47)	0.69 (0.46)	–0.03		
White	0.07 (0.26)	0.08 (0.27)	–0.01		
With Individualized Education Program (IEP)	0.05 (0.22)	0.04 (0.19)	0.01	1.48	.14
English language learner student	0.47 (0.50)	0.49 (0.50)	–0.02	–0.92	.36
Baseline Lexile measure	400.95 (139.60)	398.19 (140.73)	2.76	0.42	.68

Note: Numbers in parentheses are standard deviations. Components may not sum to 1 because of rounding.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the differences in mean values were calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. χ^2 is the test statistic used for race/ethnicity; t-test is used for all other variables. For categorical variables, the t-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

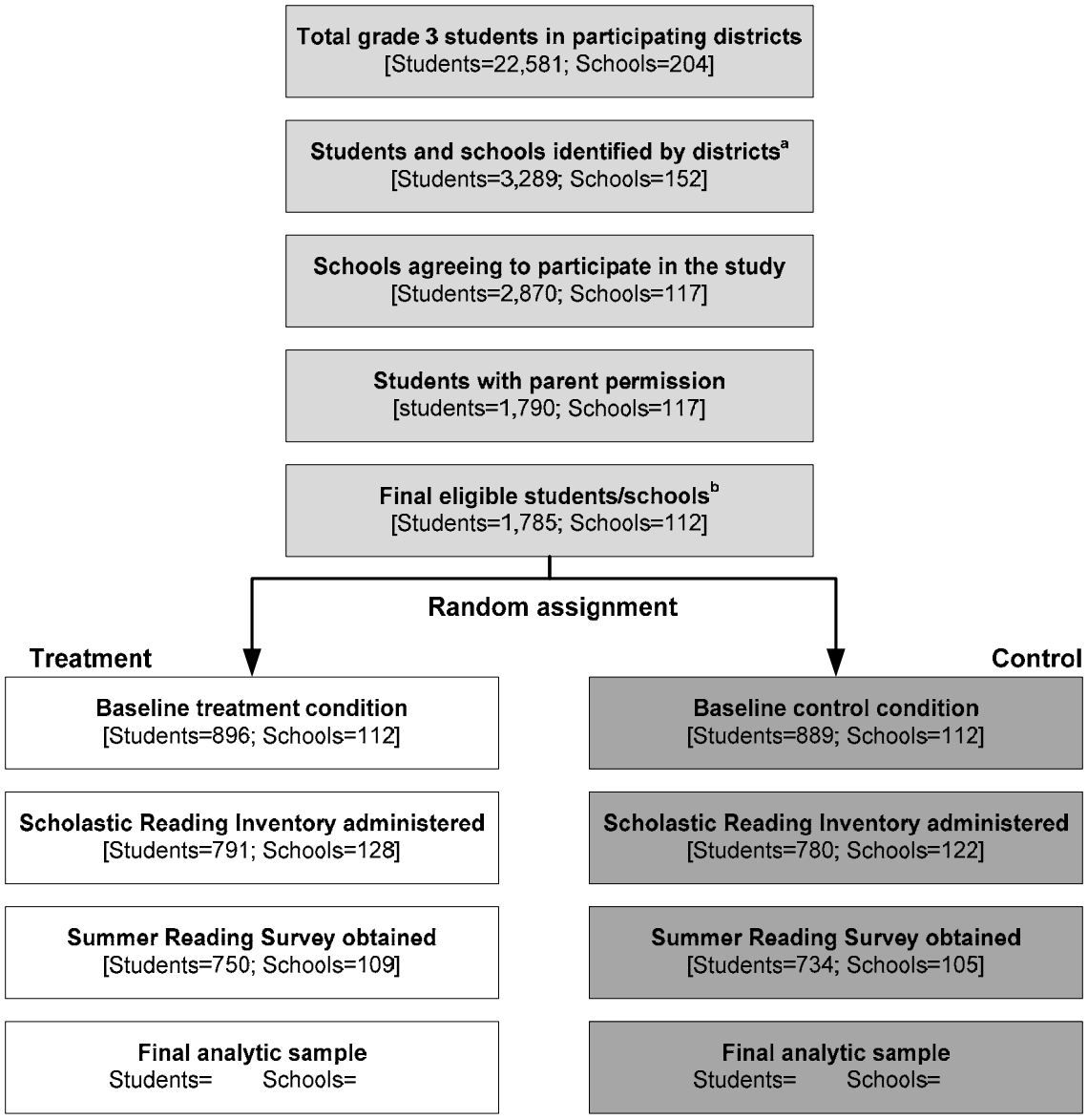
Source: District-provided student data collected May 2009–June 2009.

Study sample at each phase of the study

During the fall semester, the posttest assessment (SRI) and a summer reading survey were administered to students in both the treatment and control groups. Eighty-eight percent of students in both groups (*n*=791 treatment and *n*=780 control) completed the SRI; 84 percent of students in the treatment group (*n*=750) and 83 percent of students in the control group (*n*=734) completed the summer reading survey.[17] The SRI was not completed by 214 students, because they had moved out of the participating districts, were absent on the days of testing, or were enrolled at a school that had technical difficulties installing the SRI. Therefore, the final analytic sample included 1,571 students, 791 in the treatment group and 780 in the control group. A participant flowchart describing the study sample at each stage of the study is in figure 2-1.

[17] The study team did not monitor the administration of the summer reading survey, so we do not know why some students were not given the survey when they took the SRI.

Figure 2-1. Participant flowchart



Note: Randomization was conducted in 112 schools across four districts. The analytic sample included all students who were randomly assigned at baseline and have valid SRI scores. There were 110 schools in the final analytic sample, but not all of the 110 schools had both treatment and control students with valid SRI scores. 109 schools in the analytic sample had valid data for treatment students; 109 schools had valid data for control students; and 110 schools had valid data for either treatment or control students. In addition, 90 students moved to a different school within the same district or to another district in the study between randomization and posttest data collection. As a result, posttest data were collected from 128 schools. However, data analysis was based on students' school assignment at the time of randomization.

a. Districts identified the schools they had selected for possible participation in the study, as well as students within those schools who met eligibility criteria.

b. Five schools were ineligible to participate because only one student in the school had parent consent. Eligible students were those who met all eligibility criteria and had parent permission.

Source: Study records collected May 2009–December 2009.

Data collection

Data for the study came from 44 sources. These sources are described below.

Student interest survey

A student interest survey was administered at the beginning of the study to determine what types of books students liked to read. An initial version of the survey was created by MetaMetrics, Inc., based on a survey used by Kim (2006) and modified for a Durham READS study (MetaMetrics, Inc. n.d.a)[18]. The interest areas included were selected from the Book Industry Systems Advisory Committee (BISAC) (Book Industry Study Group, Inc. 2011) classification scheme for book subjects. Metametrics used cluster analysis to map these to a smaller classification hierarchy aligned with Kim's earlier classification. For the current study, the survey was pilot tested for ease of use and clarity of instructions with eight grade 3 students who met the eligibility criteria and whose parents gave consent. No changes were made to the survey based on this pilot. However, based on district feedback about the appropriateness of interest areas, the final survey was modified to omit the *romance* interest category.

The survey was printed on the back of the consent form sent home to parents. Students were asked to indicate all reading categories of interest to them from among 18 available genres (for example, adventure, art, mystery, sports). Survey results were entered into Excel databases by two study team members working independently. The databases were compared electronically, and any differences were reconciled by a third, independent study team member. The student interest survey is in appendix C.

Grade 3 Texas Assessment of Knowledge and Skills–Reading

TAKS is given to all Texas public school students in grades 3–11; reading is first tested as part of the grade 3 TAKS. The TAKS assessments are linked to the state curriculum standards (Texas Essential Knowledge and Skills; Texas Education Agency 2010d). The untimed grade 3 TAKS–Reading assessment consists of reading passages and 36 multiple choice items assessing students' comprehension, knowledge of literary elements, use of reading strategies, and use of critical-thinking skills (Texas Education Agency 2004).

A recent analysis of the reliability of the grade 3 TAKS–Reading conducted for spring 2009 with data for 316,319 students found an alpha of 0.882 using the Kuder-Richardson 20 test (Texas Education Agency 2010c).[19]

Students participating in the current study took the English language TAKS–Reading in spring 2009. Lexile measures corresponding to students' TAKS scaled scores (Texas Education

[18] Psychometric details about the survey were not available from MetaMetrics.

[19] The Kuder-Richardson 20 is measure of the internal reliability of a test, which reflects how consistently students perform across items within a test.

Agency 2005a) were used at baseline to determine student eligibility and to match reading level of students to books.[20]

Scholastic Reading Inventory

The SRI, used as the outcome measure in this study, is a computer-adaptive assessment for students in grades K–12 that measures reading comprehension. In an adaptive test, items presented to a student are based on the student's performance on earlier items; over time, the difficulty of the questions adjusts to the student's ability (Weiss and Kingsbury 1984). The number of SRI items administered varies by student performance. Items are written in an embedded completion format, which is similar to a fill-in-the-blank format, and reading passages are derived from authentic published material, such as children's fiction and nonfiction literature, magazines, newspapers, and classic literature. The SRI yields Lexile measures. The test-retest reliability of the SRI for grade 4 is 0.929 (Knutson 2008),[21] and the concurrent validity coefficient for the SRI with the Stanford Achievement Test Series, Tenth Edition (Harcourt Assessment, Inc. 2003) is 0.821 (Scholastic, Inc. 2007). According to the SRI *Technical Guide* (Scholastic, Inc. 2007), grade 4 student performance on the SRI is rated as follows: 349L and below = Below basic; 350L to 599L = Basic; 600L to 900L = Proficient; and 901L and above = Advanced.

During the fall 2009 semester, the SRI was administered to students in both treatment and control groups at the students' school by the students' teacher or reading coach. Administration time was approximately 20 minutes. Printouts of the SRI scores were sent by the districts to the study team, and the data were entered into Excel databases by two team members working independently. The resultant databases were compared electronically, and any differences were reconciled by a third, independent team member.

Summer reading survey

The summer reading survey was designed to collect descriptive information on students' summer school participation, their reading activities during the summer, and the amount and types of reading materials in their homes. This paper-and-pencil survey was adapted from surveys used in previous summer reading programs, eliminating survey questions that were irrelevant for this study (Kim 2006, 2007; Kim and White 2008; MetaMetrics, Inc. n.d.a). No attempt was made to collect data to examine the psychometric properties of this modified survey. The summer reading survey was administered by a teacher or reading coach to students in both the treatment and control groups on the same day as the SRI test. Surveys were collected in conjunction with SRI post-testing, and, therefore, were collected from September through

[20] The Lexile measures are included as a result of a linking study conducted at the request of the Texas Education Agency (MetaMetrics, Inc. 2005); see appendix E for additional information.

[21] Grade 4 SRI data are presented because students are in grade 4 at the time of posttest administration.

December, with the majority of survey data being collected in October and November. Because the survey was completed in the fall, student responses might have been affected by their ability to accurately recall their summer activities, and treatment student responses could have been affected by having received books even if they did not read all of them. Student data were entered into Excel databases by two team members working independently. The resultant databases were compared electronically and any differences were reconciled by a third, independent team member. (See appendix C for a copy of the summer reading survey.) Some data from the summer reading survey were used in the analysis of the first exploratory research question and are reported in chapter 5; other survey results are summarized in appendix K.

Response rates

Table 2-6 shows the response rates for each measure in the study by treatment status. All students in the study sample have data for the two measures collected in the spring, the student interest survey, and the Grade 3 Texas Assessment of Knowledge and Skills-Reading. Response rates for the two measures collected in the fall[22], the SRI and Summer Reading Survey, were 88 percent and 83 percent, respectively. The SRI was not administered to 140 students (67 treatment and 73 control) who moved out of the district during the summer, to 30 students (14 treatment and 16 control) absent on the days of testing, nor to 44 students in three schools (24 treatment and 20 control) unable to install the SRI on the school computers. Except for 87 students who took the SRI but did not complete the Summer Reading Survey, data were either obtained for both posttest measures or were missing for both. (Some data from the summer reading survey were used in the analysis of the first exploratory research question and are reported in chapter 5; other survey results are summarized in appendix K.)

Table 2-6. Response rate for each measure by treatment status

	Response rate		
Measures	*Treatment*	*Control*	*Total*
Student interest survey	100	100	100
Grade 3 Texas Assessment of Knowledge and Skills-Reading	100	100	100
Scholastic Reading Inventory	88	88	88
Summer Reading Survey	84	83	83

[22] SRI and Summer Reading Survey data were collected from September through December, with the majority collected in October and November.

Data analysis methods

The initial dataset was examined to identify and correct any out-of-range values for any variables; none were found.[23] An outlier analysis was conducted to identify any extreme values (more than 2.5 standard deviations from the mean) and to examine them for evidence of error.[24] No students or values were dropped as a result of this examination.

Confirmatory analysis

The study was designed to answer the following confirmatory research question:

- For economically disadvantaged students reading below the 50th percentile nationally, does being sent eight free books during the first part of the summer (June/July) matched to reading level and interest area, along with six reminder postcards, result in significantly better reading comprehension scores in the fall?

The confirmatory analysis was conducted using an OLS regression model that compared posttest reading comprehension as measured by the SRI of students in the treatment group with those of students in the control group. A total of 214 cases with missing data for the fall SRI test were omitted from the analysis. Each student's baseline Lexile measure from the spring 2009 grade 3 TAKS–Reading was included as a covariate in the model to obtain more precise parameter estimates (Bloom, Richburg-Hayes, and Black 2007; Raudenbush, Martinez, and Spybrook 2005). In addition, the model needed to account for any baseline differences in the analytic sample. The examination of baseline equivalence (see table J-1 in appendix J) found statistically significant differences for only one variable—Black race/ethnicity (t=2.64, p=0.01). Specifically, the percentage of Black students in the treatment group (22.63 percent) was higher than in the control group (17.31 percent). To adjust for this systematic difference between the treatment and control group, Black race/ethnicity was included as an additional covariate in the confirmatory analysis.

In this study, schools served as a blocking variable; students in each school were randomly assigned to either the treatment or control group. Therefore, the analysis included dummy variables for the schools to account for the random assignment within schools. (Model details are in appendix L; results are in chapter 4.)

Sensitivity analyses

Seven sensitivity analyses were conducted to evaluate the robustness of the results from the primary analysis to analytical decisions made when implementing the primary analysis.

[23] An out-of-range value is one that is invalid for a particular variable. For example, for a variable that represents the number of correct items on a 20-item test, valid values would be 0 to 20. A score of 30 would be an out-of-range value and clearly an error.

[24] An outlier is a value that is extreme but still valid for a particular variable. For example, for a variable with a mean of 10 and a standard deviation of 2, a value of 20 would be an outlier.

These sensitivity analyses are described in the following paragraphs. (Model details are in appendix L; results are in chapter 4.)

Multiple imputation for missing data. An alternative approach for dealing with missing data is multiple imputation. For this sensitivity analysis, a version of multiple imputation called multivariate stochastic sequential regression-based multiple imputation was used to create 10 multiply imputed datasets that included the 214 students with missing SRI scores. These data sets were then used with the primary model to examine the robustness of the results to the method for addressing missing data. (Details regarding multiple imputation are provided in appendix J.)

Analysis without covariates. The baseline covariates are included in the model primarily to provide additional power for testing the confirmatory hypothesis. However, this approach makes assumptions about the relationships between covariates and the outcome variable (for example, a linear relationship between baseline and posttest Lexile measures). Therefore, the results of the analysis were also examined with a model that did not include the covariates (Baseline Lexile measure and Black race/ethnicity) to assess the robustness of the results to a potential misspecification of the relationships.

Alternative covariate for baseline reading. Because the outcome variable in this study is reported on the Lexile scale, a baseline variable also measured on the Lexile scale was used. However, converting the TAKS scaled scores to Lexile measures could have resulted in a loss of information about reading ability. Therefore, the results of the analysis were also examined using an alternative baseline variable—TAKS scaled scores—to assess the robustness of the results to potential error in linking scaled scores to Lexile measures.

Additional covariate for weeks of instruction. Posttest SRI data were collected for 1,571 of the 1,785 participating students; however, because of the logistics involved in purchasing, installing, and administering the SRI in 128 schools,[25] test administration occurred at different times in different schools during the fall semester. Therefore, the number of weeks of instruction before taking the SRI varied for some students. Because random assignment occurred within schools, the average number of weeks of instruction should be approximately the same for treatment and control students; however, 90 students moved to a different school between spring and fall. The average number of weeks of instruction did not differ significantly for treatment and control students ($t=-0.328$, $p=0.743$). To fully examine whether the variation in weeks of instruction before posttest data collection could have affected the results of the confirmatory analysis, a sensitivity analysis that included a covariate for weeks of instruction in the full model was conducted to evaluate the robustness of the findings to this variable.

[25] Randomization was conducted in 112 schools across four districts; however, 90 students moved to a different school within the same district or to another district in the study. The SRI was administered at the school the student attended in the fall; data analysis was based on students' school assignment at the time of randomization.

Interaction of treatment status and school (with equal weighting). In the primary confirmatory model, the impact of the summer reading program was calculated for the average student in the sample, regardless of the school the student attended. Therefore, schools with larger numbers of participating students contributed more to the impact estimate. This approach could be problematic if impacts were different in schools with a large number of participating students because of different school characteristics. As a sensitivity analysis, the treatment effect was estimated for the average school by including interactions between individual schools and the treatment condition and averaging these school-specific treatment effects using equal weighting.

Interaction of treatment status and school (with precision weighting). The confirmatory analysis estimated the treatment effect for the average student by controlling for the school that participating students attended. Another approach to estimating the treatment effect for the average student is to create interaction terms between each school and student treatment status—and to average these impact estimates using precision weighting. With precision weighting, schools with larger samples of participating students are given larger weights than schools with smaller samples, thereby generating an average student impact estimate rather than an average school impact estimate. A sensitivity analysis was conducted using precision weighting based on standard errors to obtain the average student impact estimate.

Random effect. The study used a randomized block design with schools as blocks. In this design, within-block (or school) variation is taken into account by randomizing within each block and then including the blocks as controls in the model. A sensitivity analysis was conducted using an alternative approach to account for possible within-school clustering: a mixed effects model with a random effect for the intercept and slope.

Exploratory analyses

The study also addressed two exploratory research questions. The first exploratory research question was:

- Did students in the summer reading program treatment group report reading more books over the summer than did students in the control group?

The number of books read was a self-reported measure, with the median number of reported books being eight. However, the distribution was positively skewed, with some of the reported values being implausibly high, especially considering that eligibility criteria for the study specified that participating students were reading below the 50th percentile nationally. Therefore, the first exploratory question was analyzed using an ordered probit model, where the number of books read was recoded into five categories (0 books, 1–4 books, 5–8 books, 9–12 books, and 13 or more books). Examining baseline equivalence for the sample of students who reported the number of books read, a statistically significant difference was found between the treatment and control groups in the proportions of Black (t=3.1, p<0.01) and Hispanic (t=-1.93,

p=.0.05) students. Specifically, the percentage of Black students in the treatment group (23.67 percent) was higher than in the control group (16.90 percent), and the percentage of Hispanic students in the treatment group (64.71 percent) was lower than in the control group (69.55 percent) (see table J-2, appendix J). Therefore, these two variables were included as covariates in the model, along with the baseline Lexile measures. The model also included dummy variables for schools. (Details are in appendix L; results are in chapter 5.)

In addition, a sensitivity analysis was run to examine an alternative analytic approach using Ordinary Least Squares (OLS) regression. Because the reported number of books read was extremely positively skewed, the upper 1 percent of the distribution was trimmed; however, the remaining data were still very skewed (skewness=32.88) and, therefore, a normalizing transformation (the square root transformation) was applied to the data before running the analysis.

The second exploratory research question in this study was:

- Did the summer reading program have differential effects on reading comprehension depending on baseline reading proficiency?

The model for the second exploratory research question used final SRI scores as the outcome and included interaction terms between treatment status and baseline reading to predict the differential effect of the summer reading program, depending on baseline reading comprehension. This exploratory analysis included dummy variables for schools to account for random assignment within schools and a covariate for Black race/ethnicity to account for the baseline differences on the SRI. A dummy variable was created for the Lexile measure to represent group membership in the bottom, middle, or top third of the distribution. Scores ranged from 0L to 360L for students in the bottom third (*n*=560), from 380L to 495L for students in the middle third (*n*=580), and from 525L to 565L for students in the top third (*n*=431).[26] The results of each group were compared with those for the other using OLS regression.

[26] The Lexile ranges for the bottom, middle, and top third of the score distributions are not continuous. The study that linked TAKS scaled scores and Lexile measures (MetaMetrics, Inc. 2005) produced a conversion table that includes a Lexile measure for each possible TAKS scaled score. Although each possible TAKS score is linked to a corresponding Lexile measure, not every Lexile measure is linked to a TAKS score. In the grade 3 conversion table, the Lexile measures between 361L and 379L and between 496L and 524L are not linked to a TAKS scaled score.

Chapter 3. Implementation of the summer reading program

The summer reading program evaluated in this study is considered fully implemented if all eight books selected for a student were matched on both reading level and interest area, eight books were sent to each student, and six postcards were sent to each student. Whether students received the books or returned the postcards does not affect whether the program is considered fully implemented for this study. Tracking whether students received the shipments sent to them would require an expensive and resource-intensive follow-up with each student in a timely fashion, which was not feasible in such a large-scale implementation.[27] Even if it were verified that a package was delivered to the correct address, it would not guarantee that the intended student received the package or opened it. Nor would it be feasible to determine how many books (or how much of each a book) an individual student read. Therefore, this study examines whether being sent books and postcards has an effect, regardless of delivery status or the students' actual reading of the books. Within these parameters, this summer reading program was implemented with high fidelity, with over 97 percent of students being sent eight books matched on both reading level and interest area and 100 percent of the students being sent the postcards. This chapter describes the key steps in implementing this summer reading program, fidelity of implementation data, and the costs associated with the materials.

Six steps in implementing the summer reading program

There were six key steps for implementing this summer reading program.

1. *Identifying each participating student's Lexile measure.* District records were used to find student Lexile measures for the spring 2009 grade 3 TAKS–Reading scores.
2. *Identifying each participating student's areas of interest*. Data from the student interest survey were used to identify students' areas of interest.
3. *Developing a database of available books with the corresponding Lexile measure and relevant genres*. Scholastic, Inc. supplied the books and provided a Microsoft® Excel file listing 1,633 age-appropriate books available for use in the study, along with their Lexile measures. The file was sent to MetaMetrics, Inc. to classify each book by genre.
4. *Selecting eight books for each student matched to Lexile measure and interest areas*. A computer program (in Excel Visual Basic for Applications) was written to select eight appropriate books for each student. This program identified all books in the

[27] Although not all students who did not receive all of the books shipped to them could be identified, some cases of failed delivery were known. The results of a sensitivity analysis excluding students in the treatment group to whom the books were not delivered due to a vendor error or an incorrect address are in chapter 4.

database that were in the appropriate Lexile range for the student (from 100L below to 50L above the student's Lexile measure); this excludes books that are either too easy or too difficult for the student to read (MetaMetrics, Inc. 2008). The program then identified the subset of books that matched one or more of the interest areas selected by the student. If there were more than eight matching books, the program randomly selected eight. If there were fewer than eight matching books, all the matching books were selected and the remaining books were selected randomly from those in the appropriate Lexile range and in the most popular interest areas among students of the same sex. For 135 students, one or more of the selected books were out of stock, so other books were selected using the computer program described in this step.

5. *Shipping the eight books*. Staff at Scholastic, Inc. packaged and shipped the eight books, along with an explanatory letter, to all students in the treatment group, via United Parcel Service (UPS) in July 2009. The letter described the summer reading program and encouraged students to return the postcards they would receive during the summer.

6. *Sending the postcards.* In July and August 2009, each student in the treatment group was sent six postcards (one per week for six weeks). The purpose of these postcards was to remind students to read the eight books that had been sent to them, as well as convey interest in their reading activities over the summer.[28] The postcards included several questions asking students about a book they had recently read and requested that students return the completed (pre-addressed, postage-paid) postcards.

Fidelity of implementation

Fidelity was evaluated for selecting eight books matched on both reading level and interest areas, sending the books to each student, and sending six postcards to each student.

1. *Selecting eight matched books for each student.* For all 896 students in the treatment group, the eight selected books were in the appropriate Lexile range (table 3-1), so that step of the summer reading program was conducted with perfect fidelity. For 875 of the students in the treatment group (97.7 percent), eight books in the correct Lexile range and matching at least one selected interest area were identified. For the remaining 21 students, at least one of the books could not be matched with a selected interest area. For 12 of the 21, seven of the eight books matched the students' interest areas.

[28] However, it is important to note that returning the postcards was not part of the intervention, and, therefore, not part of fidelity of implementation. Also, no information from the completed and returned postcards was analyzed in the present study. See appendix C for a copy of the postcard.

2. *Sending the eight books to the student.* Of the 896 students in the treatment group, 891 (99.4 percent) were sent the eight selected books. Five students in the treatment group were not sent the books due to a vendor data error.[29]
3. *Sending the six postcards to the student*[30]. All 896 students in the treatment group were sent all six postcards.

Material costs associated with the summer reading program

The cost of materials for implementing the summer reading program was $25,012 for the treatment group books, postcards, and shipping, or about $28 per student. The cost of implementing the program would vary with the number of participating students and the type of student targeted; for example, books are likely to be more expensive for students in higher grades than for students in grade 3, and discounts to districts/schools would also vary by volume[31], as in this study. This total does not include study-related costs other than for implementation, such as for consent forms/student interest survey printing, books for control group students,[32] posttest software, and personnel.

[29] The book vendor accidently failed to enter five of the treatment students into their system. Due to the intent-to-treat analysis, these students were included in the final analytic treatment group sample.

[30] The intervention in this study included sending the reminder postcards as reading encouragement to the students. However, returning postcards was not part of the intervention and, therefore, not part of the fidelity of implementation. There were 33% of the students who returned one or more postcards.

[31] For example, in this study, over 14,000 books were ordered and a 40% discount was provided.

[32] Control group students were sent eight matched books after posttest data collection was complete.

Chapter 4. Confirmatory analysis results

The primary model for the confirmatory analysis was an OLS regression analysis examining posttest SRI scores for treatment versus control students. The model included covariates for baseline Lexile measure and Black race/ethnicity, as well as dummy variables for schools (see appendix L for the full model). Results from this analysis are in table 4-1. (The full model output is included in appendix M.) The average posttest scores, adjusting for baseline Lexile measure, Black race/ethnicity, and schools, were not statistically significantly different between the treatment and the control groups (estimated impact=4.89; effect size[33]=0.02; p=0.62[34]). The 95 percent confidence interval for the impact is (-14.39, 24.17).

Table 4-1. Summer reading program impact estimate with 95 percent confidence interval, 2009 (*n*=1,571)

Outcome measure	*Treatment group mean (*n=*791)*	*Control group mean (*n=*780)*	*Estimated intent-to-treat impact coefficient*[a]	p-*value*	*95% confidence interval*	*Estimated impact (effect size*[b]*)*
Scholastic Reading Inventory Scores	330.10 (238.94)	321.49 (228.89)	4.89 (9.83)	.62	–14.39, 24.17	0.02

Note: Numbers in parentheses are standard deviations for treatment and control group means and the standard error for estimated intent-to-treat impact. The means and standard deviations are unadjusted.

a. Results are analyzed using covariate-adjusted ordinary least squares regression.

b. Calculated using Hedges' *g*.

Source: Scholastic Reading Inventory data collected September 2009–December 2009.

[33] Effect sizes were calculated using Hedges' *g* (Hedges and Olkin 1985).

[34] The SRI generates Lexile measures which can include scores below 0L. In this study sample, about 8 percent of the scores were negative. Although it is common to report negative Lexiles as 0L, for the main analysis we used the SRI generated Lexile scores regardless of their sign. However, we also estimated the impact based on data in which negative Lexile scores were recoded to zero. Both the impact estimates and statistical significance were qualitatively similar for the main analysis as reported in table 4-1 and the analysis based on data with negative Lexiles recoded to zero (estimated impact = 9.96, p=.226).

Seven sensitivity analyses were conducted (table 4-2). Across all sensitivity analyses, the results were consistent with the primary impact analysis findings reported in Table 4-1. Thus, it can be reasonably concluded that the primary impact results are robust with respect to the analytic decisions for which sensitivity analyses were conducted.

Table 4-2. Sensitivity analyses for summer reading program impact estimate, 2009

Sensitivity analysis	*Analytic model*	*Sample size*[c]		*Estimated intent-to-treat impact coefficient*	p-*value*	*Estimated impact (effect size)*
		Treatment group	*Control group*			
1. Multiple Imputation for missing data[a,b]	Ordinary least squares (OLS)	896	889	9.19 (9.41)	0.33	0.04
2. Analysis without baseline Lexile measure covariate[b,d]	OLS	791	780	6.24 (11.14)	0.58	0.03
3. Alternative covariate for baseline reading[a,d]	OLS	791	780	4.11 (9.86)	0.68	0.02
4. Additional covariate for weeks of instruction[a,b,d]	OLS	791	780	5.25 (9.80)	0.59	0.02
5. Interaction of treatment status and school (with equal weighting)[a,d]	OLS	791	780	4.51 (14.58)	0.76	0.02
6. Interaction of treatment status and school (with precision weighting)[a,d]	OLS	791	780	4.84 (9.75)	0.62	0.02
7. Random effect[a,d]	Multilevel	791	780	8.81 (11.65)	0.45	0.04

Note: Numbers in parentheses are standard errors.

a. Analyses 1 and 3–7 are adjusted for the covariates baseline Lexile measure and Black race/ethnicity.

b. Analyses 1–4 are adjusted for school dummy variables.

c. Data from treatment and control students at 112 schools were included in sensitivity analysis 1; data from 110 schools were included in sensitivity analyses 2–5 and 7; data from 109 schools were included in sensitivity analysis 6.

d. The results for analysis 1 are based on 10 multiply imputed datasets. The other analyses are based on the dataset with casewise deletion.

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009

Chapter 5. Exploratory analyses results

This chapter reports on the results of two exploratory analyses conducted to determine whether intermediate or subgroup-specific effects were observed. The first exploratory analysis examined whether students in the summer reading program reported reading more books over the summer than did control students. The second exploratory analysis examined whether the impact of the summer reading program on reading comprehension differed with baseline reading proficiency.

The primary model for the first exploratory analysis used an ordered probit regression analysis of the reported number of books read for treatment versus control students. The model included covariates for baseline Lexile measure, Black race/ethnicity and Hispanic race/ethnicity, as well as dummy variables for schools (see appendix L for the full model). The number of books was analyzed using the following categories: 0 books, 1–4 books, 5–8 books, 9–12 books, and 13 or more books. The percent of students for each category by treatment status are shown in figure 5-1. Results from the primary exploratory analysis show that students in the treatment group reported reading significantly more books than did students in the control group (effect size=0.11, p=.01). As shown in figure 5-1, this finding is largely accounted for by the 5–8 books category, which contains a higher percent of students in the treatment group (26.81 percent) compared to the control group (18.99 percent).

Figure 5-1. Distribution of student reported number of books by category and treatment status

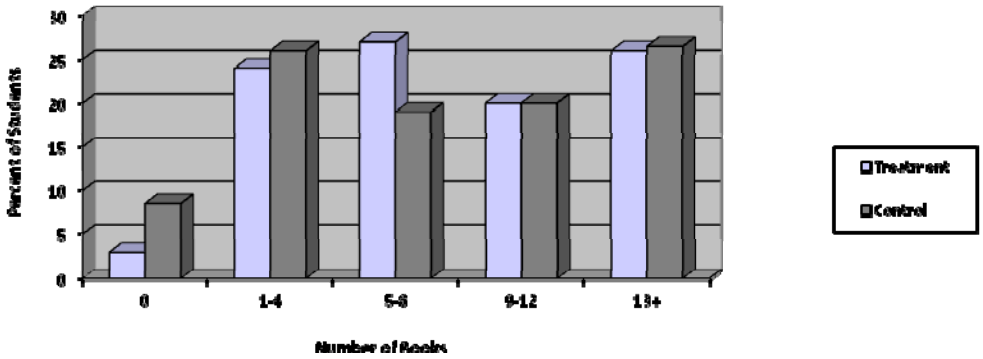


Table 5-1. Summer reading program impact estimate for reported number of books read with confidence interval, 2009 (*n*=1,447)

Outcome measure	*Estimated intent-to-treat impact coefficient*[a]	p-*value*	*95% confidence interval*	*Estimated impact, (effect size)*
Ordered probit regression analysis of student-reported number of books read over the summer	0.14 (0.06)	.01	0.03, 0.25	0.11

Note: The number in parentheses is the standard error for the estimated intent-to-treat impact. The sample size was n=731 treatment and n=716 control students.

a. These results are based on the dataset with casewise deletion. Student reported number of books read over the summer has been recoded into five categories: 0 books, 1-4 books, 5-8 books, 9-12 books, and 13 or more books.

Source: District-provided student data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

As described in chapter 2, the sensitivity analysis used for this first exploratory research question is an ordinary least squares (OLS) regression with 1 percent trimming to address implausible values, and a square-root transformation to address the positively skewed distribution of scores. Casewise deletion was used to address missing data and the model included the same covariates—Black, Hispanic, baseline TAKS Lexile scores—as did the model for primary analysis (see appendix L for model details). Results for this sensitivity analysis show that students in the treatment group reported reading an average of 1.03[35] more books than did students in the control group (table 5-2; see appendix M for full model output).This difference that is statistically significant (effect size=0.12, *p*=0.02).. Students in the control group read, on average, 7.73 books and students in the treatment group read, on average, 8.76 books.

Table 5-2. Summer reading program impact estimate for reported number of books read with confidence interval, 2009 (*n*=1,432)

Outcome measure	*Estimated intent-to-treat impact coefficient*	p-*value*	*95% confidence interval*	*Estimated impact, (effect size)*
Square root of number of books read with 1 percent trimming	0.18 (0.08)	0.02	0.03, 0.34	0.12

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

[35] This result is for number of books (as opposed to square root of number of books), after adjusting the treatment mean based on the estimated impact estimate and rescaling.

Effects for students with different baseline levels

The second exploratory question set out to test whether the effect of the summer reading program differed with baseline reading proficiency for economically disadvantaged students whose baseline reading proficiency scores were all below the 50th percentile nationally.

Results from the OLS regression[36] show that the interaction terms between the summer reading program and baseline Lexile measures were not statistically significant (table 5-3; see appendix M for the full output). In other words, the effect of the summer reading program on student reading comprehension did not significantly differ across the bottom, middle, and top third of the baseline Lexile measures distribution. Although final Scholastic Reading Inventory scores were statistically significantly higher for students with higher baseline Lexile measures (see tables M-13 and M-14 in appendix M), the effect of the summer reading program did not differ as a function of baseline reading comprehension.

Table 5-3. Differential effect of baseline Lexile measure on impact of summer reading program on Scholastic Reading Inventory scores, 2009 (*n*=1,571)

Baseline Lexile measure	*Estimated intent-to-treat impact coefficient*	p-*value*	*95% confidence interval*	*Estimated impact (effect size)*
Upper third versus lower third	0.72 (26.44)	0.98	-51.14, 52.00	0.00
Middle third versus lower third	-11.47 (24.50)	0.64	-59.52, 36.59	-0.05
Upper third versus middle third	12.19 (26.58)	0.65	-39.95, 64.32	0.05

Note: The results are based on ordinary least squares regression that adjusted for the covariate baseline Lexile measure, Black race/ethnicity, and for school dummy variables. Measures are coefficients of interaction terms (upper third versus lower third Lexile measure*treatment status; middle third versus lower third Lexile measure*treatment status; upper third versus middle third Lexile measure*treatment status).

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

[36] Using casewise deletion as the missing data approach for the second exploratory analysis, there was a statistically significant difference between treatment and control groups (see table J-1 in appendix J) on Black race/ethnicity, with a higher proportion of Black students in the treatment group compared to the control group. Therefore, a covariate for Black race/ethnicity was included in this second exploratory analysis.

Chapter 6. Summary of findings, study limitations, and suggestions for future research

This study was a rigorous evaluation of a summer reading program designed as a multidistrict randomized controlled trial. The program consisted of sending each student a single shipment of eight books matched to the student's reading level and interest areas (matched books) followed by six weekly reminder postcards. The study randomly assigned economically disadvantaged grade 3 students reading below the 50th percentile nationally within each participating school to either a treatment or a control group.

Effect of the summer reading program on reading comprehension

The primary finding of the confirmatory analysis is that the summer reading program did not have a statistically significant impact on student reading comprehension. The average Scholastic Reading Inventory score for students in the treatment group was 330.10, compared with 321.49 for students in the control group (effect size=0.02, p=0.62). Seven sensitivity analyses conducted to determine whether the results were consistent with the confirmatory impact analysis showed the finding to be robust to different analytic approaches: all eight analyses yielded results that were not statistically significant. Thus, the overall conclusion is that the summer reading program examined in this study did not impact the posttest reading comprehension scores of economically disadvantaged grade 3 students reading below the 50th percentile nationally.

Two exploratory analyses were conducted to examine the following: (1) whether the summer reading program had an impact on the number of books students reported reading over the summer and (2) whether the impact varied with students' baseline levels of reading proficiency. For the first exploratory research question, a statistically significant effect on the number of books students reported reading over the summer was identified. The covariate-adjusted effect size was 0.11 and the p-value was 0.01. A sensitivity analysis examined the use of OLS regression instead of ordered probit regression and found consistent results. On average, treatment students reported reading 1.03 more books over the summer (mean=8.76 books) than did students in the control group (mean=7.73 books). For the second exploratory research question, the estimated impacts did not differ among groups of students with different baseline reading skills.

Study contribution

The confirmatory finding from this study was not statistically significant. In contrast, five of seven previous studies of summer reading programs, found statistically significant improvements in reading achievement among program participants (Allington et al. 2010; Butler

2010; Crowell and Klein 1981; Kim 2006; Kim and White 2008), including three RCTs (Allington et al. 2010; Kim 2006; Kim and White 2008).

An exploratory analysis found that students who received books over the summer reported reading an average of one more book than did students who did not receive books----a finding that is consistent with previous studies (Kim 2007; Kim and Guryan 2010).

Differences in observed effects across studies could be due to the fact that the summer reading program examined in this study did not include teacher support, instructional components, or parent involvement, which several previous studies—one quasi-experiment (Butler 2010) and four RCTs (Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008)—included to varying degrees. In addition, the program examined in the current study lasted just one summer, whereas that examined in Allington et al. (2010) spanned three summers. Also, the current study sample consisted entirely of economically disadvantaged students, whereas only one of the seven previous studies did (Butler 2010). Further, the current study sample consisted entirely of students reading below the 50th percentile nationally, whereas none of the previous studies targeted this population (Allington et al. 2010; Butler 2010; Crowell and Klein 1981; Kim 2006, 2007; Kim and Guryan 2010; Kim and White 2008).

It also may be that teacher and parent support components are necessary for a summer reading to be effective during a single summer, but that they may be less important if students participate in summer reading programs over a longer time period. For example, Allington (2010) found that when students were provided books over a period of three summers, even without any additional support components, student reading significantly improved.

Limitations and suggestions for future research

Several limitations of this study should be considered. First, the sample was not selected randomly from a larger population but rather from four Texas school districts that met qualifying criteria and agreed to participate. The results would not necessarily apply to students in other districts.

Second, parent consent, required for student participation, ranged from 42 percent to 74 percent across the four participating districts. Because households in which parents are more likely to consent may differ from households in which parents are less likely to consent, the results might not generalize to all students meeting study criteria in participating districts.

Third, because the study only focused on economically disadvantaged students who were below the 50th percentile nationally, the results of this study may not generalize to other groups of students. This exclusion of students reading at or above the 50th percentile could further lead to statistical artifacts due to restriction of range.

Fourth, matched books were sent to students in the treatment group, but there was no follow-up to determine whether students received the books or read them (or what portion they

read). Therefore, the study cannot say whether treatment students who received and read the books were affected differently by the summer reading program than treatment students who may not have received or read the books. It should be noted that postcard return rates were low (33 percent) in the current study. This may indicate that some students read the books but did not bother to return postcards, or that some students may not have read all of the books that were sent to them.

Fifth, scheduling constraints delayed shipment of the books until July, shortening the time students had to read the books.

Sixth, there was a lag between the end of the summer reading program and administration of the posttest and summer reading survey. Both treatment and control students received classroom instruction in the fall semester, and effects of the program may have diminished over time. Thus, the lag might have made any effect harder to detect. In addition, responses to the summer reading survey, completed in the fall, may have been influenced by students' varying ability to accurately recall their summer reading activities.

Seventh, because the random assignment was conducted within schools, it is possible that there was some contamination between treatment and control students, or their parents, and this could have affected behavior of control students or parents.

Eighth, only a single outcome measure, the SRI, was used in this study, which is a different measure than used in some other summer reading studies, potentially making it difficult to compare the results across studies

Finally, the survey instruments in this study were modified from instruments used in previous research. Although those instruments were well-developed, the modifications were minor, and these were not used for evaluating the confirmatory research question, it should be noted that there was no attempt to evaluate the psychometric properties of the modified instruments.

The study's conclusions are constrained by several aspects of the program's design. Allington et al. (2010) found that a summer reading program spanning three summers had a significant effect on reading comprehension; this study's program lasted just one summer. Examining the impact of a summer reading program spanning multiple summers could provide information about dosage effects on summer reading loss. In addition, the current study did not include teacher instruction and parent involvement, components that were part of summer reading programs found to have statistically significant impacts on student reading achievement, either for specific subgroups (Kim 2006) or for the full sample (Butler 2010; Crowell and Klein 1981; Kim and White 2008).

Appendix A. Description of the Lexile Framework® for Reading

The text that follows is excerpted from a study by Regional Educational Laboratory Southwest (Wilkins et al. 2010, pp. 13–17). That study developed a methodology for linking the reading levels of a population of students to the difficulty levels of a set of books. This information is included to help readers understand how students were matched to books based on reading level. Table A1 has been modified to include examples of books that were distributed in this study and that are appropriate for the study's age range.

The Lexile Framework® for Reading is a linguistic theory–based method for measuring the reading difficulty of prose text and the reading capacity of individuals (White and Clement 2001). The Lexile theory operationalizes the semantic and syntactic components of reading by using mean sentence length and $\log_{10}$ word frequency (Stenner et al. 1987). More specifically, the semantic component is gauged using proxies for word frequency—average number of letters or syllables per word. These proxies capitalize on the high negative correlation between length of words and frequency of word usage. Long words and polysyllabic words are used less frequently than short monosyllabic words, making word length a good proxy for the likelihood of an individual being exposed to them. Syntactic complexity is gauged using sentence length.

The framework uses a mathematical formula to assign reading difficulty values to passages of text known as *slices*. As detailed in Stenner et al. (2006), a text file consisting of the entire contents of a selected book is submitted to the Lexile Analyzer. An auto-edit function removes irrelevant and nontext features (such as figures and tables), and the file is divided into 125-word slices.[37] For each slice, two variables are calculated: one using word frequency (the mean $\log_{10}$ word frequency) and one using the mean sentence length. A proprietary regression equation uses the word frequency and sentence length variables to obtain the Lexile measure for that slice of text. This process is repeated for all slices in the text file. The results are combined to obtain the overall Lexile measure for a book.[38]

The difficulty values are reported on a scale called a *Lexile* (L) that ranges from 0L (for emerging readers and beginning texts) to 1700L (for advanced readers and texts). The student Lexile measure indicates the level of text a student can be expected to read with approximately 75 percent comprehension, which is considered "the level at which students can successfully negotiate the material with the use of context clues and other comprehension strategies to fill in

[37] Because the Lexile Analyzer does not end a slice in the middle of a sentence, most slices are longer than 125 words.

[38] A Lexile reader measure is assigned to an individual to reflect his or her reading ability. A Lexile text measure is assigned to a text (such as a book or article) to reflect how difficult it is to comprehend. This study uses the term Lexile measure to refer to both Lexile reader measure and Lexile text measure.

the gaps" (Lennon and Burdick 2004, p. 9). Table A1 shows the Lexile scales for selected books.[39]

In 2001, a panel of reading experts working with the National Center for Education Statistics evaluated the use of the Lexile Framework to compare text difficulty and reader ability (White and Clement 2001). The panel's report emphasized that the Lexile Framework has solid psychometric properties and has been validated across a wide variety of populations. It described the Lexile Framework as a powerful and practical tool for assessing the relationship between text difficulty and reading ability.

The SRI manual (Scholastic, Inc. 2007) provides additional psychometric properties of the Lexile scale. A number of linking studies have been conducted in which the scores on various standardized tests of reading comprehension have been assigned corresponding Lexile measures. These corresponding validity coefficients (correlation between the test score and the Lexile measure) range from .91 to .93 on nationally normed tests, such as the Stanford Achievement Test (ninth edition) and Gates-McGinitie Reading Test (version 4).

The SRI manual also provides results from an analysis in which a variety of basal readers (grade-based textbooks used to teach reading to students) were used to obtain theory-based calibrations. Correlations between the rank order of the basal readers and the Lexile calibrations, after correction for range restriction and measurement error, averaged 0.995.

Table A-1. Examples of published books at various levels of the Lexile scale

Lexile measure	*Example book*	*Publication information*[a]
50	*D.W. all wet*	Brown, M. (1988). *D.W. all wet.* Boston: Little, Brown and Co.
120	*Amelia Bedelia*	Parish, P. (1963). *Amelia Bedelia.* New York: Harper Collins.
260	*Young Cam Jansen and the baseball mystery*	Adler, D. (1999). *Young Cam Jansen and the baseball mystery.* New York: Viking.
370	*Freckle juice*	Blume, J. (1971). *Freckle juice.* New York: Yearling.
430	*The art teacher from the black lagoon*	Thaler, M. (2003). *The art teacher from the black lagoon.* New York: Scholastic, Inc.

[39] For illustration, these samples are somewhat longer than the usual slices. Because Lexile measures consider only word frequency and sentence length, while other dimensions of reading comprehension are not directly part of the Lexile measure calculation, text passages at the same level of the Lexile scale can vary in structure, complexity, contextual cues, and other features.

Lexile measure	*Example book*	*Publication information*[a]
490	*There's a boy in the girls' bathroom*	Sachar, L. (1987). *There's a boy in the girls' bathroom.* New York: Random House.
540	*The babysitter*	Stine, R.L. (1989). *The babysitter*. New York: Scholastic, Inc.
600	*Stellaluna*	Cannon, J. (1993). *Stellaluna.* Orlando, FL: Harcourt.
660	*The face on the milk carton*	Cooney, C. (1990). *The face on the milk carton.* New York: Laural Leaf.
700	*Where the red fern grows*	Rawls, W. (1961). *Where the red fern grows.* New York: Yearling.

a. Because different editions of a book can reflect editorial changes, slight differences in Lexile measures may exist between different publications of the same book. Reference data are provided for the specific publication that corresponds to the Lexile measure listed for each book; as a result, the year of publication provided may not reflect the first year of publication for that book.

Source: MetaMetrics, Inc. n.d.c.

The panel report also identified several concerns about the Lexile Framework:

- Within a particular text, high-frequency words (*a, he*) tend to be common and appear many times; low-frequency words appear rarely; and midfrequency words appear several times. Words that appear several times in the text can range widely in semantic complexity (*ahhh* and *salubrious*); this variability in semantic complexity is overlooked when the measure is a word-frequency count, as it is in the Lexile Framework for Reading.

- It was unclear to the panel whether there were sources of measurement error unaccounted for in the Lexile research conducted to that point.

- The Lexile Framework cannot be used to assess some types of nonliterary or expository text, such as poems, recipes, and lists.

Since the 2001 panel report, MetaMetrics, Inc. (developer of the Lexile Framework) has addressed many of the concerns raised by the panel (White and Clement 2001). For example, the panel noted that estimation of word frequency–related issues could be improved and measurement error reduced by increasing the size of the slices analyzed. At the time of the 200l report, slices were taken from a portion of each textbook. The entire textbook is now sliced and Lexile measures are assigned to each slice (Stenner et al. 2006).

Appendix B. Findings from previous studies of summer reading programs

The seven previous studies of summer reading programs referred to in this study used a variety of methods to report their findings. To ease comparison, the reported findings were recalculated as estimated effect sizes (standardized mean differences). This appendix presents the formulas used to calculate the effect size along with the estimated effect sizes.

Formulas for calculating effect sizes

The formulas used to estimate the effect sizes varied according to the study outcome and sample information:

1. Based on *t*-statistic and sample sizes for each group:

$$ES = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

2. Based on *F* statistic and degrees of freedom:

$$ES = \sqrt{\frac{df_{num} F}{df_{den}}}$$

3. Based on *t*-statistic and degrees of freedom:

$$ES = \frac{2t}{\sqrt{df}}$$

4. Based on χ^2 statistic and sample size when df = 1:

$$ES = 2\sqrt{\frac{\chi^2}{N - \chi^2}}$$

5. Based on χ^2 statistic and sample size when df > 1:

$$ES = 2\sqrt{\frac{\chi^2}{N}}$$

6. Based on means, standard deviations, and sample sizes for each group:

$$ES = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}}$$

7. Based on means, sample sizes for each group, and *p*-value:

$$ES = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\left(\overline{X_1} - \overline{X_2}\right) \Big/ t_{p/2, n_1 + n_2 - 2}}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$

If the study does not provide an exact *p*-value, this formula can yield only an upper or lower bound for the estimated effect size.

8. For dependent *t*-tests, based on mean differences, sample size, and standard error:

$$ES = meandiff \Big/ \frac{se_\Delta \sqrt{N}}{\sqrt{3}}$$

This formula was derived using the assumptions and derivations outlined below. Because this is a dependent sample *t*-test, the standard error of the differences does not accurately reflect the standard deviation of the pretest or posttest scores. The standard error of the differences for a dependent sample *t*-test can be rewritten as:

$$se_\Delta = \sqrt{se_1^2 + se_2^2 + 2r_{12} se_1 se_2}$$

$$= \sqrt{\left(\frac{\sigma_1}{\sqrt{n}}\right)^2 + \left(\frac{\sigma_2}{\sqrt{n}}\right)^2 + 2r_{12}\left(\frac{\sigma_1}{\sqrt{n}}\right)\left(\frac{\sigma_2}{\sqrt{n}}\right)}$$

$$= \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n} + \frac{2r_{12}\sigma_1\sigma_2}{n}}$$

$$= \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + 2r_{12}\sigma_1\sigma_2}{n}}$$

If the pretest and posttest standard deviations are assumed to be approximately equal (they are not reported in the study), then:

$$\approx \sqrt{\frac{2\sigma_1^2 + 2r_{12}\sigma_1^2}{n}} = \sqrt{\frac{\sigma_1^2(2+2r_{12})}{n}}$$

$$= \frac{\sigma_1\sqrt{(2+2r_{12})}}{\sqrt{n}}$$

Solving for $\sigma_{1:}$

$$\sigma_1 = \frac{se_\Delta\sqrt{n}}{\sqrt{2+2r_{12}}}$$

The correlation between pretest and posttest scores was not reported. Bloom et al. (2007) found correlations for pretest and posttest grade 3 reading scores ranged from 0.47 to 0.72. Therefore, it is reasonable to assume that $r \approx 0.5$. In that case,

$$\sigma_1 = \frac{se_\Delta\sqrt{n}}{\sqrt{2+2(.5)}} = \frac{se_\Delta\sqrt{n}}{\sqrt{3}}$$

Previous studies

This section discusses the recalculated estimated effect sizes for previous summer reading program studies.

Allington et al. (2010)

Allington et al. (2010) examined a summer reading program that sent 12 self-selected books to treatment group students over three consecutive summers, beginning the summer after grades 1 and 2. The outcome of interest was reading comprehension. The study found statistically significant differences between treatment and control groups for the full sample and for students in the free or reduced-price lunch program (FRPL; table B-1). Moreover, the effect size was larger for students enrolled in FRPL than for students in the full sample.

Table B-1. Allington et al. (2010 findings with estimated effect sizes and reported *p*-values

Design	*Grades*	*Sample size*		*Outcome measure*	*Analytic sample*	*Predictor*	*Effect size*	*Effect size calculation method*[a]	p-*value*
		Treatment	*Control*						
Randomized controlled trial	1–2	852	478	State reading assessment (reading comprehension)	Full sample	Treatment	0.13	3	0.015
					Students in free or reduced-price lunch program	Treatment	0.20	3	0.001

a. Based on *t*-statistic and degrees of freedom; see previous section of this appendix for details.
Source: Allington et al. 2010.

Butler (2010)

Butler (2010) examined 94 low-income (all FRPL) students in grades 2–4 to determine whether a summer reading program had differential effects on English language learner students and disadvantaged students who speak English as a first language (identified as EL1 students). Books were matched to students using Fountas and Pinnell's (1996) guided reading levels[40] for students who were in guided reading groups; otherwise, books were selected based on teacher input and books that students were reading in class. Books were matched to students' interest areas using a self-selection method that allowed students to choose among books at their reading level.

For the purposes of this report, a conservative approach to classifying studies was used, and Butler's (2010) study was classified as quasi-experimental. This study design could be referred to as a hybrid since random assignment was used for some (but not all) students in the sample. As such, causal inferences cannot be made for all comparisons between groups. In particular, the first group of students (n=27), slated to receive 10 leveled books of interest as well as teacher visits over the summer, were assigned to the first treatment based on proximity to the teacher travel routes, not random assignment, for ease of implementation. The remaining students (n=67) were randomly assigned to either the second treatment group (n=26), slated to receive 10 leveled books of interest over the summer, or the control group (n=41), which received no books. Treatment and control students were encouraged to read over the summer and complete a reading log.

The outcomes of interest were oral reading fluency and oral retelling. Based on a two-way analysis of covariance, there was a statistically significant difference in oral reading fluency

[40] Fountas and Pinnell (1996) provide a list of books by reading level for guided reading. Guided reading is an approach for teaching students reading strategies where the strategies are applied with texts of increasing difficulty to move students towards successful independent reading.

and oral retelling between English language learner students and EL1 students when controlling for their first language (table B-2). A statistically significant increase in oral reading fluency and oral retelling was observed for disadvantaged English language learner students and EL1 students. However, Butler did not present results on baseline equivalence between English language learner students and EL1 students, so results should be interpreted with caution.

Table B-2. Butler (2010) findings with estimated effect sizes and reported *p*-values

		Sample size							
		Treatment[a]							
Design	*Grades*	*1*	*2*	*Control*	*Outcome measure*	*Comparison*	*Effect size*	*Effect size calculation method*[b]	p-*value*
Quasi-experimental	2–4	27	26	41	DIBELS Oral Reading Fluency	Home visits versus books only	–0.11	6	ns
						Home visits versus control	0.59	6	<0.0001
						Books only versus control	0.65	6	<0.0001
						Race versus assignment	0.08	2	0.93
						Reading level versus assignment	0.31	2	0.36
					DIBELS Oral Retelling	Home visits versus books only	–0.02	6	ns
						Home visits versus control	0.65	6	<0.0001
						Books only versus control	0.77	6	<0.0001
						Race versus assignment	0.46	2	0.11
						Reading level versus assignment	0.31	2	0.35

DIBELS is Dynamic Indicators of Basic Early Literacy Skills; ns is not statistically significant.

a. Students in treatment group 1 were visited weekly by a teacher who brought a bin of leveled books, from which students selected one or two books to read; students in treatment group 2 received 10 books to take home over the summer.

b. Based on *F* statistic and degrees of freedom (2); based on means, standard deviations, and sample sizes for each group (6); see previous section of this appendix for details.

Source: Butler 2010.

Crowell and Klein (1981)

Crowell and Klein (1981) examined a summer reading program that sent 10 books, leveled by reading difficulty, to treatment group students in grades 1 and 2 over the summer. The outcomes of interest were reading comprehension and vocabulary. For the purposes of this report, Crowell and Klein's study (1981) is considered quasi-experimental because of the way students were assigned to either the treatment or "comparison" group. Students were rank ordered based on their score on a criterion-referenced reading test and teacher judgment of their reading levels. Students were then systematically assigned in an alternating fashion based on their rank order rather than being randomly assigned. Since this alternate assignment could bias the group students were placed in, this study was not classified as an experiment.

Based on an analysis of covariance, there were statistically significant differences on the vocabulary subtest of the Gates-MacGinitie Reading Test between the treatment and control groups for the full sample and for students in grade 1, but not for students in grade 2 (table B-3). On a separate criterion-referenced sight vocabulary test, there were significant differences between treatment and control group for students in grade 1. Students who were sent books showed slightly more improvement in Gates-MacGinitie reading comprehension scores, but this result was not reported by Crowell and Klein because the improvement was not statistically significant. Moreover, Crowell and Klein did not present results on baseline equivalence between the treatment and control groups, so results should be interpreted with caution.

Table B-3. Crowell and Klein (1981) findings with estimated effect sizes and reported *p*-values

		Sample size							
Design	*Grades*	*Treatment*	*Com-parison*	*Outcome measure*	*Analytic sample*	*Predictor*	*Effect size*	*Effect size calculation method*[a]	p-*value*
Quasi-experimental	1, 2	25	23	GMRT vocabulary subtest[b]	Full sample	Treatment	>0.67	7	<0.05
					Grade 1 (*n*=22)	Treatment	>1.03	7	<0.05
					Grade 2 (*n*=26)	Treatment	<0.94	7	ns
				Criterion-referenced sight vocabulary test[c]	Grade 1 (*n*=22)	Treatment	0.88	1	<0.05

GMRT is Gates-MacGinitie Reading Test; ns is not statistically significant.

a. Based on means, sample sizes for each group, and *p*-value (7); based on *t*-statistic and sample sizes for each group (1); see previous section of this appendix for details.

b. Only means and approximate *p*-value information were provided for this outcome measure. Because the exact *p*-values were not reported, only upper or lower bounds for the estimated effect sizes can be obtained in these cases.

c. Crowell and Klein (1981) do not provide the title of this assessment.

Source: Crowell and Klein 1981.

Kim (2006)

Kim (2006) examined a summer reading program that sent eight books matched on reading level and interest area to treatment group students in grades 1–4 over the summer. The outcome of interest was reading achievement (a comprehension and vocabulary composite). Differences between treatment and control groups were not statistically significant for the full sample, but statistically significant differences were found for Black students, students with baseline reading fluency scores below the median, and students with fewer than 50 children's books in their home (table B-4).

Table B-4. Kim (2006) findings with estimated effect sizes and reported *p*-values

		Sample size							
Design	*Grades*	*Treatment*	*Control*	*Outcome measure*	*Analytic sample*	*Predictor*	*Effect size*	*Effect size calculation method*[a]	p-*value*
Randomized controlled trial	1–4	252	234	ITBS (composite score of reading comprehension and vocabulary)	Full sample	Treatment	0.17	1	0.059
						Treatment *ethnicity	0.28	2	0.035
						Treatment *income	0.14	2	0.133
						Treatment *ethnicity* income	0.18	2	0.33
					Asian	Treatment	0.14	1	0.125
					Black	Treatment	0.24	1	0.011
					Hispanic	Treatment	0.16	1	0.081
					White	Treatment	0.11	1	0.220
					Spring ITBS above median	Treatment	0.11	8	0.327
					Spring ITBS below median	Treatment	0.20	8	0.072
					Spring reading fluency above median	Treatment	0.04	8	0.714

Design	Grades	Sample size		Outcome measure	Analytic sample	Predictor	Effect size	Effect size calculation method[a]	p-value
		Treatment	Control						
					Spring reading fluency below median	Treatment	0.35	8	0.008
					More than 100 books in home	Treatment	0.09	8	0.467
					Fewer than 100 books in home	Treatment	0.19	8	0.060
					More than 50 children's books in home	Treatment	0.04	8	0.735
					Fewer than 50 children's books in home	Treatment	0.29	8	0.020

ITBS is Iowa Test of Basic Skills.

a. Based on *t*-statistic and sample sizes for each group (1); based on *F* statistic and degrees of freedom (2); for dependent *t*-tests, based on mean differences, sample size, and standard error (8); see previous section of this appendix.

Source: Kim 2006.

Kim (2007)

Kim (2007) examined a summer reading program that sent 10 books matched on reading and interest level to treatment group students in grades 1–5 over the summer. The outcomes of interest were number of books read and reading achievement (a comprehension and vocabulary composite). The study found statistically significant differences between the treatment and control groups for number of books read but not for reading comprehension (table B-5).

Table B-5. Kim (2007) findings with estimated effect sizes and reported *p*-values

		Sample size						
Design	*Grades*	*Treatment*	*Control*	*Outcome measure*	*Predictor*	*Effect size*	*Effect size calculation method*[a]	p-*value*
Randomized controlled trial	1–5	138	141	Number of books read	Treatment	0.76	2	0.001
					Treatment *grade	0.30	2	ns
					Treatment *FRPL	0.01	2	ns
				Stanford Achievement Test Series, Tenth Edition (composite score of vocabulary and comprehension)	Treatment	0.10	2	0.36
					Treatment *grade	0.34	2	0.11
					Treatment *FRPL	0.04	2	0.74

FRPL is free or reduced-price lunch program participation; ns is not statistically significant.
a. Based on *F* statistic and degrees of freedom (2); see previous section of this appendix.
Source: Kim 2007.

Kim and Guryan (2010)

Kim and Guryan (2010) examined a summer reading program for low-income grade 4 Hispanic students. Students in all three study groups (two treatment groups and one control group) received three scripted lessons on reading strategies for oral reading fluency, reading comprehension, and postcard completion—the same lessons used in Kim and White (2008). Students in treatment groups 1 and 2 self-selected 14 books of interest at a book fair and were sent the 10 of those selections that were closest to their Lexile measure. In addition, students in treatment group 2 were invited to three 2-hour summer family literacy events. Across both treatment groups, 67 percent of students received books below their Lexile measure. The outcomes of interest were number of books read, reading comprehension, vocabulary, and reading achievement (a comprehension and vocabulary composite). There was a statistically significant difference in the number of books read between treatment groups 1 and 2 but not between the treatment and control groups (table B-6). No statistically significant differences were found for reading comprehension.

Table B-6. Kim and Guryan (2010) findings with estimated effect sizes and reported *p*-values

Design	Grades	Sample size: Treatment 1[a]	Sample size: Treatment 2[b]	Sample size: Control	Outcome measure	Comparison	Effect size	Effect size calculation method[c]	p-value
Randomized controlled trial	4	102	112	110	Number of books read	Three experimental conditions	0.39	5	<0.001
						Treatment versus control	0.42	4	<0.01
						Family literacy versus control	0.42	4	<0.01
						Treatment versus family literacy	0.02	4	ns
					GMRT total reading achievement scores	Three experimental conditions	0.10	2	ns
					GMRT reading comprehension	Three experimental conditions	0.10	2	ns
					GMRT reading vocabulary	Three experimental conditions	0.24	2	ns

GMRT is Gates-MacGinitie Reading Test; ns is not statistically significant.

a. Students received 10 books over the summer.

b. Students received 10 books over the summer as well instructional components for the family.

c. Based on χ^2 statistic and sample size when df > 1 (5); based on χ^2 statistic and sample size when df = 1 (4); based on *F* statistic and degrees of freedom (2); see previous section of this appendix.

Source: Kim and Guryan 2010.

Kim and White (2008)

Kim and White (2008) examined a summer reading program that sent eight books matched on reading level and interest to students in grades 3–5 over the summer. Teachers were randomly assigned to one of four conditions (3 treatment and 1 control), and all participating teachers received two hours of professional development that included three scripted lessons: oral reading scaffolding, reading comprehension scaffolding, and oral reading and comprehension strategy practice including completing a postcard about a book that was read. Students were randomly assigned to treatment or control conditions. Students in all three treatment groups received matched books and postcards over the summer, but the three treatment groups received different reading strategy lessons. Students in treatment group 1 received only the postcard practice scripted lesson; students in treatment group 2 received oral reading

scaffolding and postcard practice scripted lessons; and students in treatment group 3 received all three scripted lessons. The control group received a nonscripted lesson created by the teacher.

The outcome of interest was reading achievement (a comprehension and vocabulary composite). A statistically significant difference was found between the treatment group that included oral reading and comprehension scaffolding and the control group and between the two groups with scaffolding and the two groups without scaffolding (table B-7).

Table B-7. Kim and White (2008) findings with estimated effect sizes and reported *p*-values

		Sample size								
		Treatment[a]								
Design	*Grades*	*1*	*2*	*3*	*Control*	*Outcome measure*	*Predictor*	*Effect size*	*Effect size calculation method*[b]	p-*value*
Randomized controlled trial	3–5	93	100	100	107	ITBS (total reading achievement: composite score of vocabulary and comprehension)	Books with oral reading and comprehension scaffolding versus control	0.24	3	<0.03
							Books with oral reading and comprehension scaffolding versus books only	0.12	6	0.063
							Books with oral reading and comprehension scaffolding versus books with oral reading scaffolding	0.08	6	0.23
							Two groups with scaffolding versus two groups without scaffolding	0.21	3	<0.05

ITBS is Iowa Test of Basic Skills.

a. Treatment group 1 received eight books matched on reading level and interest; treatment group 2 received eight books matched on reading level and interest, along with oral reading scaffolding; treatment group 3 received eight books matched on reading level and interest, along with oral reading and reading comprehension scaffolding.

b. Based on *t*-statistic and degrees of freedom (3); based on means, standard deviations, and sample sizes for each group (6); see previous section of this appendix.

Source: Kim and White 2008.

Appendix C. Student interest survey, explanatory letter, postcard, and summer reading survey

The following exhibits are provided in this appendix, in the following order:

- Sample of student interest survey.
- Explanatory letter sent with books to treatment group.
- Sample postcard.
- Summer reading survey.

Exhibit 1. Sample of student interest survey

REL SOUTHWEST
Regional Educational Laboratory At Edvance Research

Hello! Please help us with a study to learn about reading. We will ask you to read books and answer questions this summer, and take a reading test in the fall. But you do not have to. We will not tell your answers to anyone. Fill out this form and we will send you 8 free books and postcards. Thank you!

Office Use Only:
ID: _____

Student Interest Survey

Please tell us what you like to read.
Check the box next to the types of books you like to read.

- ☐ **ADVENTURE**
- ☐ **ANIMAL STORIES** (Fiction)
- ☐ **ANIMAL STORIES** (Nonfiction)
- ☐ **ART** (crafts, fashion, photography…)
- ☐ **BIOGRAPHIES**
- ☐ **FAIRY TALES, MYTHS & FOLKTALES**
- ☐ **FAMILY & CHILDHOOD**
- ☐ **FANTASY** (wizards, outer space, princesses…)
- ☐ **GRAPHIC NOVELS & COMICS** (super-heroes, true stories)
- ☐ **HISTORY**
- ☐ **HUMOR & GAMES**
- ☐ **ENTERTAINMENT (TV, music, movies…)**
- ☐ **MYSTERY**
- ☐ **NATURE STORIES** (Fiction)
- ☐ **NATURE STORIES** (Nonfiction)
- ☐ **SCIENCE & TECHNOLOGY**
- ☐ **SOCIAL ISSUES** (politics, law …)
- ☐ **SPORTS**

Return this page to your teacher, along with your signed Parental Permission Form.

Thank you!

Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

Exhibit 2. Explanatory letter sent with books to treatment group

REL SOUTHWEST
Regional Educational Laboratory
At Edvance Research

Summer Reading is Fun!

These are books for the Summer Ready Study that are yours to keep. Don't let Mr. Bookworm get lonely this Summer. Read with him! He picked out some great books for you to enjoy. They match the Book Interest Survey that you filled out at school.

You will be getting post cards in the mail. They will remind you about the books. On the back of each post card is a place to tell us what you think about the books you have been reading. We would really like to hear from you, so please remember to send your post cards back to us. Tell your Mom and Dad we have already stamped it for you!

Enjoy your new books!

Edvance Research, Inc. 9901 IH-10 West, Suite 700
San Antonio, Texas 78230
Phone (210) 558-1902 • Fax (210) 558-1075
www.edvanceresearch.com http://edlabs.ed.gov/RELSouthwest

Item # 273291

Exhibit 3. Sample postcard

REL SOUTHWEST
Regional Educational Laboratory At Edvance Research

REL Southwest at Edvance Research
9901 IH-10 West, Suite 700
San Antonio, TX 78230

Stamp

YOU'RE A SUMMER READING STAR!

Answer these questions about the book you just read.

The title of the book: _____

1.) **Did you read this book all by yourself?**
___ Yes ___ No, I read with someone else's help.

2.) **Did you read the whole book?**
___ Yes
___ No

3.) **How hard was the book to read?**
___ Easy ___ Just right ___ Too hard

4.) **How interesting was this book?**
___ Very interesting ___ A little interesting ___ Not interesting

Have a family member mail this soon!

Parent/Guardian signature: _____

top

top

mail info.

Hello!

On the back of the post card is a place to tell us what you think about a book you have read.
We would really like to hear from you, so please remember to send your post card back to us.
Tell your Mom and Dad we have already stamped it for you!

WE HOPE YOU ARE ENJOYING YOUR NEW BOOKS!

Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

Paperwork Burden Statement

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is 1850-0864 (expires 5/31/2012). The time required to complete this information collection is estimated to average 3 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Education, Washington, D.C. 20202-4700. If you have comments or concerns regarding the status of your individual submission of this form, write directly to: Dean Gerdeman, U.S. Department of Education, 555 New Jersey Avenue, N.W., Room 506D, Washington D.C. 20208.

Exhibit 4. Summer reading survey[41]

Your Summer

Office Use Only:

ID: _____

1) **Did you go to summer school?**
□ Yes □ No

2) **How many books did you read this summer? _____**

3) **How many times did you go to a library this summer?**
□ Never
□ 1 or 2 times
□ 3 or more times

4) **Do you get a newspaper at home?**
□ Yes □ No □ I don't know

5) **How many magazines do you get at home?**
□ No magazines
□ 1 or 2 magazines
□ 3 or more magazines
□ I don't know

6) **About how many children's books do you have in your home?**
□ No books
□ 1–15 books
□ 16–30 books
□ 31–50 books
□ More than 50 books

7) **Does anyone at your house read?**
□ Yes □ No □ I don't know

Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

Paperwork Burden Statement

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is **1850-0864 (expires 5/31/2012).** The time required to complete this information collection is estimated to average 3 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. **If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to:** U.S. Department of Education, Washington, D.C. 20202-4700. **If you have comments or concerns regarding the status of your individual submission of this form, write directly to:** Karen Armstrong, U.S. Department of Education, 555 New Jersey Avenue, N.W., Room 506D, Washington D.C. 20208.

[41] Students were not given any additional instruction on how to interpret or respond to the survey items.

Appendix D. Power analysis

The minimum detectable effect size (MDES) represents the smallest *true* program impacts in standard deviation units that can be detected with high probability (Bloom 2005). All else equal, the smaller a detectable effect size, the larger the study sample needs to be. The MDES selected should be large enough that the detectable impact is important but small enough to be feasible given the intervention.

Based on the positive effect sizes for reading gains shown in previous research on summer reading programs (for example, Kim 2006, 2007; Kim and White 2008) and feedback from the current study's technical working group members (a group with considerable randomized controlled trial [RCT] experience) and other experts in the field, it was determined that an effect size of approximately 0.12 was appropriate for the primary confirmatory hypotheses in this study.

Assumptions

The power analysis assumes a design in which students are randomly assigned to the treatment and control groups. The power calculations are based on ordinary least squares (OLS) regression with a baseline covariate, using the following assumptions:

- *Desired statistical power:* 80 percent.
- *Statistical significance level:* α=0.05 for a two-tailed test.
- *Minimum detectable effect size:* desired MDES=0.12 (results are also presented for MDES=0.10 and MDES=0.13).
- *Explanatory power of the baseline covariate for explaining variation in the posttest:* R^2=0.3 (results are also presented for R^2=0.0 and R^2=0.5).
- *Attrition and parent consent return rate:* Results are presented for overall participation rates (including taking the posttest) of 80, 64, and 50 percent. It was assumed that 80 percent of eligible students would return signed parent consent forms and, conservatively, that at least 80 percent of those students would be available for posttesting in the fall, yielding an overall participation rate of 64 percent. Nevertheless, the power was estimated at a more conservative return rate of 50 percent.

Target sample size

Tables D-1–D-3 include findings from the power analyses incorporating the above assumptions. These tables show the required sample size (per group) for MDESs of 0.10 to 0.13, R^2 values of 0.0 to 0.5, and overall participation rates of 50 percent to 80 percent. The targeted sample size is 1,516 students per group, based on MDES=0.12, R^2=0.3, and a 50 percent participation rate, as noted in bold in table D-1.

Table D-1. Sample size needed per group, assuming 50 percent participation rate

Minimum detectable effect size	*Sample size needed per group with R^2 of*		
	0.0	*0.3*	*0.5*
0.10	3,112	2,178	1,558
0.12	2,166	**1,516**	1,084
0.13	1,846	1,294	924

Note: Bold indicates targeted sample size.
Source: Authors' analyses.

Table D-2. Sample size needed per group, assuming 64 percent participation rate

Minimum detectable effect size	*Sample size needed per group with R^2 of*		
	0.0	*0.3*	*0.5*
0.10	2,431	1,702	1,217
0.12	1,692	1,184	847
0.13	1,442	1,011	722

Source: Authors' analyses.

Table D-3. Sample size needed per group, assuming 80 percent participation rate

Minimum detectable effect size	*Sample size needed per group with R^2 of*		
	0.0	*0.3*	*0.5*
0.10	1,945	1,361	974
0.12	1,354	948	678
0.13	1,154	809	578

Source: Authors' analyses.

The power analysis conducted for the confirmatory research question also applies to exploratory analysis 1, which differs only in the outcome variable. But exploratory analysis 2 examines interaction terms, so the MDESs from those power analyses do not apply. Table D-4 provides the MDES for exploratory analysis 2 corresponding to the sample sizes from the above power analysis.

Table D-4. Minimum detectable effect size for exploratory research question 2

Sample size (per group)	*Minimum detectable effect size*
1,516[a]	0.21
1,184[b]	0.24
948[c]	0.27

a. From table D-1.
b. From table D-2.
c. From table D-3.
Source: Authors' analyses.

Appendix E. Texas Assessment of Knowledge and Skills–Lexile linking study

In 2005, the Texas Education Agency asked MetaMetrics, Inc., to link the Texas Assessment of Knowledge and Skills (TAKS) scaled scores to the Lexile scale (MetaMetrics, Inc. 2005). This linking study involved a sample of approximately 500 English-speaking Texas public school students. Students completed both the 2005 TAKS and a MetaMetrics, Inc. reading comprehension test designed to provide Lexile measures for students. While details of the MetaMetrics, Inc. linking procedure are proprietary, in summary, Lexile linking tests were developed to be parallel forms of the TAKS (for example, with similar test content and psychometric properties) and provide a Lexile measure (L). A series of calibration equations were developed using a linear median-anchored approach with the one-parameter logistic model (the Rasch model). These equations were used to calculate grade-specific linking constants, which were then used to develop the TAKS–Lexile conversion tables (MetaMetrics, Inc. 2005).

The key psychometric property in evaluating a linking study is the standard error of the linking. For the 2005 linking study, the standard error varies by grade; for grade 3, it is 3.4L. Because student Lexile measures are rounded to the nearest 5L, the standard error of linking is less than the rounding applied to the Lexile measure.

Because identical TAKS scaled scores are considered equivalent from 2003—the first year the TAKS was administered—to today (Texas Education Agency 2008b), the conversion tables from the 2005 study were applied to the 2009 TAKS data examined in the current study to determine the Lexile measure corresponding to each TAKS scaled score (MetaMetrics, Inc. 2005). The Texas Education Agency has used these data to provide Lexile measures as part of student TAKS score reports since spring 2006 (Texas Education Agency 2005a).

The average scale score for the spring 2009 grade 3 TAKS–Reading assessment was 2318, which converts to an average Lexile measure of 660L (Texas Education Agency 2009d). The baseline performance of the grade 3 students in this study (401L treatment, 398L control) was above the state passing standard (360 Lexiles) but below the state average (660 Lexiles).

Appendix F. Recruitment and study sample details

Recruitment was restricted to Texas school districts with at least 25,000 students to limit the number of districts needed to attain the large student sample required, minimize project management effort, and contain cost. A letter and recruitment flyer were mailed to the superintendents or assistant superintendents in 42 Texas districts that met this criterion. Two weeks later, a follow-up email with an attached recruitment flyer was sent to approximately 300 district-level decisionmakers, including chief academic officers, curriculum directors, directors of research and evaluation, and assistant superintendents. Further follow-up was conducted by phone and email, as needed.

Eight of the 42 school districts expressed interest in participating, 29 districts did not respond, and 5 declined to participate for various reasons (for example, participation in another summer program) or did not have enough students who met the eligibility criteria (for example, enrollment in free or reduced-price lunch [FRPL]). Meetings (face-to-face or conference call) were scheduled with the districts expressing interest, to provide information on the study and to assess district capacity to participate in the study (such as support of key district administrators, willingness to commit time to implementation efforts, and information on any past participation in research efforts). Of the eight interested districts, four were eliminated based on either failure to complete a formal letters of intent—the next stage in the recruitment process—or other factors, including responsiveness to communications and concurrent participation in other research studies. Based on the statistical power estimates, the recruitment target for this study was 3,032 students, and this target was met by the four selected districts. Regional Educational Laboratory (REL) Southwest completed all paperwork required by each district's research and testing departments to conduct the study.

Districts determined which elementary schools would participate in the study using district-generated criteria not disclosed to the study team. Consent forms were provided to all selected schools in the four districts although some schools did not return consent forms, indicating their preference to not participate. Table F-1 provides a description of the study sample by district.

Table F-1. Study sample descriptive statistics, 2009

Item	Mean across districts	Range across districts	Total for all districts
Grade 3 student enrollment (spring 2009)[a]	5,645.3	3,280–7,812	22,581
Number of eligible schools	38.0	25–53	152
Number of eligible students in eligible schools	822.3	395–1,556	3,289
Number of schools agreeing to participate	29.3	21–47	117
Number of consent forms sent to schools agreeing to participate	718.5	272–1,556	2,870
Consent refused	33.6	2–53	84
Forms not returned	399.2	133–364	996
Consent granted	716.0	122–1,148	1,790
Number of schools at random assignment[b]	44.8	20–47	112
Number of treatment students	358.4	60–575	896
Number of treatment students to whom books were sent[c]	356.4	59–573	891

a. To protect district confidentiality, exact grade 3 student enrollment data are not provided

b. Because the study used within-school randomization, each school needed at least two eligible students with parent consent to participate. Five schools had only one student each with consent to participate, so those schools were not included in the final study sample.

c. Five students were not sent books due to vendor error.

Source: Texas Education Agency 2009c; study records.

Appendix G. Participating district profiles

This appendix provides a demographic and student achievement profile for the districts that participated in this study.[42]

The four districts participating in this study are large and densely populated; each has more than 40,000 students (Texas Education Agency 2009b) and is in a major urban or suburban area.[43]

The districts vary by racial/ethnic composition. White and Hispanic students are the two largest racial/ethnic groups in the four participating districts (Texas Education Agency 2009b). Black students represent the third largest group in each district, at 4–16 percent, followed by Asian students at 2–10 percent (Texas Education Agency 2009b).

Economically disadvantaged students (students who enrolled in the free and reduced-price lunch program or other public assistance) in the four participating districts make up 36–69 percent of the total. Participation in bilingual/English as a second language education programs is 13–28 percent. Special education participation rates show much less variability, at 7–9 percent.

Each district met adequate yearly progress standards for 2008 (Texas Education Agency 2009b). The districts were either rated Recognized or Academically acceptable for 2008 (Texas Education Agency 2009b).[44] Reading/English language assessment student proficiency rates on the Texas Assessment of Knowledge and Skills (TAKS), the annual assessment used in Texas, range from 87 percent to 94 percent in participating districts, and TAKS math proficiency rates range from 78 percent to 87 percent (Texas Education Agency 2009b). Most teachers in all four districts have at least six years of teaching experience (Texas Education Agency 2009b).

[42] Data are for 2008/09 unless otherwise noted.

[43] *Major urban* is defined as "the largest school districts in the state that serve the six metropolitan areas of Houston, Dallas, San Antonio, Fort Worth, Austin, and El Paso." *Major suburban* is defined as "other school districts in and around the major urban areas… [that are generally] contiguous to major urban areas" (Texas Education Agency 2008d).

[44] *Rated* refers to the district's classification in the state accountability rating system used by the Texas Education Agency to rate public schools and districts. There are four possible ratings; from lowest to highest they are: *Academically unacceptable*, *Academically acceptable*, *Recognized*, and *Exemplary* (Texas Education Agency 2008d).

Appendix H. Description of the grade 3 Texas Assessment of Knowledge and Skills–Reading

All Texas public high school students are required to take the Texas Assessment of Knowledge and Skills (TAKS; Texas Project First n.d.). In addition to the standard TAKS, there are three English-language versions of the grade 3 TAKS–Reading available for students receiving special education services: TAKS–Accommodated, TAKS–M, and TAKS–Alt. Which version should be administered to a special education student is up to the student's Admission, Review, and Dismissal Committee (Texas Education Agency 2010b).

TAKS–Accommodated is available to students receiving special education services and instruction on or near grade level (Texas Project First n.d.). It uses a larger font, has fewer items per page, and does not include field test questions (Texas Education Agency 2008c). These adjustments do not invalidate interpreting TAKS–Accommodated scores the same way as TAKS scores. The current study does not distinguish between students taking the TAKS and TAKS–Accommodated because data from these tests are combined for state and federal accountability reporting (Texas Education Agency 2008a). In 2009, 90.7 percent of students in grades 3–11 took TAKS and 2.7 percent took TAKS–Accommodated (Texas Education Agency 2009a).

TAKS–M is a modified version of the TAKS available to "students receiving special education services who have a disability that significantly affects academic progress in the grade-level curriculum and precludes the achievement of grade-level proficiency within a school year" (Texas Education Agency n.d.). TAKS–Alt is an alternate version of the TAKS available to "students receiving special education services who have the most significant cognitive disabilities and are unable to participate in the other statewide assessments even with substantial accommodations and/or modifications" (Texas Education Agency 2007). Lexile measures are not provided for students taking the TAKS–M or TAKS–Alt (Texas Education Agency 2009a).

Several Spanish-language and linguistically accommodated versions of the grade 3 TAKS–Reading are also available for recent immigrants who qualify because of limited English proficiency (Texas Education Agency 2010a). Students unable to take an English-language TAKS were not eligible to participate in the study. Which language version should be administered to a student is up to the student's Language Proficiency Assessment Committee (Texas Education Agency 2010b)

The Texas Education Agency (2010d), in collaboration with Pearson Education, Inc. (the company that operates the TAKS testing program for the state of Texas; Pearson Education, Inc. n.d.), conducted an analysis of the reliability and validity of the grade 3 TAKS–Reading assessment administered in spring 2009. The complete reading subtest included 36 items;

316,319 students took the assessment. The Kuder-Richardson 20,[45] used to test the reliability of this measure, produced an alpha of 0.882. The average scale score for grade 3, 2,318 (vertical scale score of approximately 605), was converted into an average Lexile measure of 660L. This scale score exceeded the passing standard for grade 3, set at a scale score of 2,100 (vertical scale score of 483) and 360L.

[45] The Kuder-Richardson 20 is measure of the internal reliability of a test, which reflects how consistently students perform across items on a test.

Appendix I. Random assignment

This study used a design in which eligible students in each participating school were randomly assigned to either the treatment or control group. This appendix provides a detailed account of the random assignment process.

An Excel file was created containing a list of schools in each participating district and the eligible students in each school. Using the Excel rand() function,[46] each student was assigned a random number and then students were sorted by number.

Random assignment was conducted separately for each school. If a school had an even number of students, the half with the smaller random numbers was assigned to the treatment group and the other half to the control group. If a school had an odd number of students, the extra student was randomly assigned a new number. If that number was less than 0.50, the student was assigned to the treatment group; otherwise, the student was assigned to the control group.

This process ensured that, in each school, the number of students in the treatment and control groups differed by at most one student. But because random assignment was conducted separately for each school, the overall balance was not exact, resulting in 896 treatment group students and 889 control group students.

[46] The rand() function randomly assigns a value between 0 and 1.

Appendix J. Missing Data

The only missing data in this study were from the Scholastic Reading Inventory (SRI) scores and summer reading surveys. All other variables (Texas Assessment of Knowledge and Skills [TAKS] scaled scores and demographics) were collected at the beginning of the study and were complete. Of the total baseline sample of 1,785, the numbers of students missing data from one or both of these instruments, overall and by treatment condition, are shown below:

- 186 students (10.4 percent of the total sample) missing both SRI score and summer reading survey data.
    - Treatment: 86 students (9.6 percent of treatment group students)
    - Control: 100 students (11.3 percent of control group students)
- 115 students (6.4 percent) missing only summer reading survey data.
    - Treatment: 60 students (6.7 percent)
    - Control: 55 students (6.2 percent)
- 28 students (1.6 percent) missing only SRI score.
    - Treatment: 19 students (2.1 percent)
    - Control: 9 students (1.0 percent)

Maintenance of baseline equivalence

The primary approach in this study for missing data is casewise deletion. To examine the effect on maintaining baseline equivalence by using casewise deletion, mean comparisons were conducted between treatment and control students not missing posttest SRI scores and between treatment and control students not missing summer reading survey data for number of books read. Baseline equivalence was not maintained for the confirmatory and corresponding sensitivity analyses and for the second exploratory analysis (table J-1) or for the first exploratory analysis with reported number of books read by students as the outcome (table J-2). In particular, race/ethnicity differed significantly by treatment status. Therefore, these models included a covariate for race/ethnicity to account for the differences.

Table J-1. Examination of baseline equivalence using casewise deletion for students with missing Scholastic Reading Inventory scores, 2009

Student characteristics (mean)[a]	*Treatment* (n= *791)*	*Control* (n=*780)*	*Difference in means*[b]	*Test statistic*[c]	p-*value*
Age in years	9.40	9.42			
	(0.46)	(0.49)	–0.02	–0.73	0.47
Sex					
Female	0.52	0.49			
	(0.50)	(0.50)	0.03	1.18	0.24
Race/ethnicity[d]				9.97	0.04
American Indian	0.00	0.00			
	(0.00)	(0.05)	0.00	–1.43	0.15
Asian	0.04	0.05			
	(0.19)	(0.22)	–0.01	–1.30	0.19
Black	0.23	0.17			
	(0.42)	(0.38)	0.05	2.64	0.01
Hispanic	0.66	0.70			
	(0.47)	(0.46)	–0.04	–1.54	0.12
White	0.08	0.08			
	(0.26)	(0.27)	0.00	–0.08	0.94
With Individualized Education Program (IEP)	0.05	0.04			
	(0.21)	(0.19)	0.01	1.07	0.29
English language learner student	0.49	0.52			
	(0.50)	(0.50)	–0.03	–1.24	0.22
Baseline Lexile measure	403.64	401.06			
	(137.57)	(140.65)	2.58	0.37	0.71

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table J-2. Examination of baseline equivalence using casewise deletion for students with missing survey data for number of books read, 2009

Student characteristics (mean)[a]	*Treatment* (n=*732)*	*Control* (n=*715)*	*Difference in means*[b]	*Test statistic*[c]	p-*value*
Age in years (mean)	9.39	9.42	–0.02	–0.99	0.32
	(0.45)	(0.50)			
Female	0.52	0.49	0.03	1.18	0.24
	(0.50)	(0.50)			
Race/ethnicity[d]				12.29	0.02
American Indian	0.00	0.00	0.00	–1.01	0.32
	(0.00)	(0.04)			
Asian	0.04	0.06	–0.02	–1.46	0.15
	(0.20)	(0.23)			
Black	0.24	0.17	0.07	3.17	<0.01
	(0.43)	(0.38)			
Hispanic	0.65	0.70	–0.05	–1.93	0.05
	(0.48)	(0.46)			
White	0.08	0.08	0.00	–0.13	0.90
	(0.27)	(0.27)			
With Individualized	0.05	0.03	0.02	1.46	0.14
Education Program (IEP)	(0.22)	(0.18)			
English language learner	0.48	0.53		–1.61	0.11
student	(0.50)	(0.50)	–0.04		
Baseline Lexile measure	403.24	401.35	1.89	0.26	0.79
	(135.78)	(140.00)			

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

a. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

b. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

c. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

Multiple imputation

Because baseline equivalence was not maintained for the analytic sample using casewise deletion, a sensitivity analysis was conducted to examine an alternative approach to missing data called Multiple Imputation. The results of that sensitivity analysis were similar to the primary analysis showing that the results are robust to the choice of a missing data approach. The details of how multiple imputation was employed for this sensitivity analysis follow.

Multiple imputation is a Monte Carlo technique that replaces missing values with $m>1$ simulated versions. The use of 5 to 10 datasets has been shown to be sufficient for obtaining parameter estimates that are close to fully efficient (Little and Rubin 1987; Rubin 1987; Schafer 1997). In this study, *m* was set at 10.

Assumptions

Multiple imputation requires certain assumptions about the nature of the missingness, including that the data are missing at random (Rubin 1987). This assumption implies that missing scores do not depend on unobserved covariates, after controlling for observed ones. The only missing data being imputed in this study are outcome scores on the SRI for the confirmatory analysis and the reported number of books read for the first exploratory research question. Note that although the assumption of "missing at random" for multiple imputation cannot be directly tested, multiple imputation is still the best available method.

The multiple imputation method for this study—Imputation and Variance Estimation software (IVEware) stochastic sequential regression-based multiple imputation—also assumes that linear extrapolation is appropriate for any imputed data extrapolated beyond the observed data. Mean differences on various observed variables as a function of missing status are provided in tables J-3 through J-6. Significant differences between the groups could raise concerns about the accuracy of the multiple imputation estimates to the degree that linear extrapolation was necessary. Note, however, that for no variable for which significant differences were found did interpolated values lie outside the range of observed values.

Mean differences by missing status on SRI scores for the treatment group are shown in table J-3. There were statistically significant differences between the missing and non-missing groups for boys and girls, American Indian students, English language learner students, and non-English language learner students. In particular, compared with the group without any missing data for SRI scores, the group with missing data had greater proportions of males, American Indian students, and non-English language learner students.

Table J-3. Treatment student characteristics, by missing Scholastic Reading Inventory score status (2009)

	Scholastic Reading Inventory score status				
Student characteristics (mean)[a]	*Missing (*n=*105)*	*Nonmissing (*n=*791)*	*Difference in means*[b]	*Test-statistic*[c]	p-*value*
Age in years	9.45 (0.53)	9.40 (0.46)	0.05	0.98	0.33

Student characteristics (mean)[a]	*Scholastic Reading Inventory score status*		*Difference in means*[b]	*Test-statistic*[c]	p-*value*
	Missing (n=*105)*	*Nonmissing* (n=*791)*			
Female	0.40 (0.49)	0.52 (0.50)	–0.12	–2.40	0.02
Race/ethnicity[d]				17.89	<0.01
American Indian	*	0.00 (0.00)	*	3.89	<0.01
Asian	*	0.04 (0.19)	*	–0.93	0.35
Black	0.24 (0.43)	0.23 (0.42)	0.01	0.27	0.79
Hispanic	0.69 (0.47)	0.66 (0.47)	0.02	0.50	0.62
White	0.04 (0.19)	0.08 (0.26)	–0.04	1.41	0.16
With Individualized Education Program (IEP)	0.07 (0.25)	0.05 (0.21)	0.02	0.82	0.41
English language learner student	0.34 (0.48)	0.49 (0.50)	–0.14	–2.75	0.01
Baseline Lexile measure	380.71 (153.21)	403.64 (135.57)	–22.93	–1.58	0.11

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding.

* To avoid a potential disclosure risk, in cases where a small proportion of students had missing data, the proportion and corresponding difference in proportions are not provided.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Mean differences by missing status on SRI scores for the control group are shown in table J-4. There were statistically significant differences between the missing and non-missing groups for race/ethnicity as a whole, English language learner students, and non-English language learner students. In particular, compared with the group without any missing data for SRI scores, the group with missing data had greater proportions of both Black and non-English language learner students.

Table J-4. Control student characteristics, by missing Scholastic Reading Inventory score status (2009)

	Scholastic Reading Inventory score status				
Student characteristics (mean)[a]	*Missing (*n=*109)*	*Nonmissing (*n=*780)*	*Difference in means*[b]	*Test-statistic*[c]	p-*value*
Age in years	9.46	9.42	0.04	0.69	0.49
	(0.54)	(0.49)			
Female	0.50	0.49	0.01	0.19	0.85
	(0.50)	(0.50)			
Race/ethnicity[d]				6.33	0.18
American Indian	0.00	0.00	0.00	–0.53	0.60
	(0.00)	(0.05)			
Asian	*	0.05	*	–1.48	0.14
		(0.22)			
Black	0.26	0.17	0.08	–2.12	0.03
	(0.44)	(0.38)			
Hispanic	0.65	0.70	–0.05	–0.98	0.33
	(0.48)	(0.46)			
White	0.07	0.08	0.00	–0.13	0.90
	(0.26)	(0.27)			
With Individualized Education Program (IEP)	*	0.04 (0.19)	*	–0.51	0.61
English language learner student	0.30	0.52	–0.21	–4.18	0.00
	(0.46)	(0.50)			
Baseline Lexile measure	377.61	401.06	–4.77	–1.63	0.10
	(140.26)	(140.65)			

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding.

* To avoid a potential disclosure risk, in cases where a small proportion of students had missing data, the proportion and corresponding difference in proportions are not provided.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Mean differences by missing status on summer reading survey data for reported number of books read for the treatment group are shown in table J-5. There were statistically significant differences between the missing and non-missing group for age, race/ethnicity overall, Asian students, American Indian students, Hispanic students, English language learner students, and non-English language learner students. In particular, compared with the group without any missing data for number of books read, the group with missing data had greater proportions of both American Indian and Hispanic students.

Table J-5. Treatment student characteristics, by missing summer reading survey data status on number of books read (2009)

	Summer reading survey data status				
Student characteristics (mean)[a]	*Missing* (n=*164)*	*Nonmissing (*n=*732)*	*Difference in means*[b]	*Test-statistic*[c]	p-*value*
Age in years	9.47	9.39	0.08	1.98	0.05
	(0.52)	(0.45)			
Female	0.47	0.52	–0.05	–1.15	0.25
	(0.50)	(0.50)			
Race/ethnicity[d]				16.24	<0.01
American Indian	*	0.00	*	2.99	<0.01
		(0.00)			
Asian	*	0.04	*	–1.74	0.08
		(0.20)			
Black	0.19	0.24	–0.05	–1.31	0.19
	(0.39)	(0.43)			
Hispanic	0.74	0.65	0.09	2.21	0.03
	(0.44)	(0.48)			
White	0.05	0.08	–0.03	–1.25	0.21
	(0.22)	(0.27)			
With Individualized Education Program	0.05	0.05	0.00	–0.09	0.93

Student characteristics (mean)[a]	*Summer reading survey data status*		*Difference in means*[b]	*Test-statistic*[c]	p-*value*
	Missing (n=*164)*	*Nonmissing (*n=*732)*			
(IEP)	(0.22)	(0.22)			
English language learner student	0.40	0.48	–0.08	–1.88	0.06
	(0.49)	(0.50)			
Baseline Lexile measure	390.76	403.24	–12.48	–1.03	0.30
	(155.52)	(135.78)			

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding.

* To avoid a potential disclosure risk, in cases where a small proportion of students had missing data, the proportion and corresponding difference in proportions are not provided.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

Mean differences by missing status on summer reading survey data for reported number of books read for the control group are shown in table J-6. There were statistically significant differences between the missing and non-missing group for Asian students, Black students, English language learner students, and non-English language learner students. In particular, compared with the group without any missing data for number of books read, the group with missing data had a smaller proportion of Asian students and greater proportions of both Black and non-English language learner students.

Table J-6. Control student characteristics, by missing summer reading survey data status on number of books read (2009)

Student characteristics (mean)[a]	*Summer reading survey data status*		*Difference in means*[b]	*Test-statistic*[c]	p-*value*
	Missing (n=*174)*	*Nonmissing (*n=*715)*			
Age in years	9.45	9.42	0.03	0.72	0.47
	(0.52)	(0.50)			
Sex					
Female	0.53	0.49	0.04	–0.96	0.34
	(0.50)	(0.50)			
Race/ethnicity[d]				13.01[a]	0.01
American Indian	*	0.00	*	1.09	0.28
		(0.04)			
Asian	*	0.06	*	–2.83	<0.01
		(0.23)			
Black	0.24	0.17	0.07	2.21	0.03
	(0.43)	(0.38)			
Hispanic	0.68	0.70	–0.02	–0.43	0.66
	(0.47)	(0.46)			
White	0.07	0.08	–0.01	0.42	0.68
	(0.25)	(0.27)			
With Individualized Education Program (IEP)	0.04	0.03	0.01	0.33	0.74
	(0.20)	(0.18)			
English language learner student	0.34	0.53	–0.18	–4.28	<0.01
	(0.48)	(0.50)			
Baseline Lexile measure	385.20	401.35	–16.15	–1.36	0.17
	(143.37)	(140.00)			

Note: Numbers in parentheses are standard deviations. Proportions may not sum to 1 because of rounding.

* To avoid a potential disclosure risk, in cases where a small proportion of students had missing data, the proportion and corresponding difference in proportions are not provided.

a. Except for *Age* and *Baseline Lexile measure*, which are reported as means, all data are reported as proportions.

b. The reported means were rounded but the difference in mean values was calculated based on the unrounded means, so the values shown in this column may not equal the difference in the reported means.

c. Chi-square for race/ethnicity; *t*-test for all other variables. For categorical variables, the *t*-tests were tests of proportions.

d. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Source: District-provided student data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

Imputation model

Multiple imputation for this study was implemented separately for the treatment and control groups using the IVEware multivariate stochastic sequential regression-based imputation (Raghunathan et al. 2001). The "impute" procedure in IVEware produces imputed values for each individual in the dataset conditional on all values observed for that individual. The basic strategy is to create imputations through a sequence of multiple regressions, varying the type of regression model by the type of variable imputed (continuous, binary, categorical, counts). Covariates include all other variables observed or imputed for that individual (that is, all other variables included in the dataset used for imputation). The imputations are defined as selections from the posterior predictive distribution specified by the regression model with a flat or noninformative prior distribution for the parameters in the regression model.

The imputation model included the following variables: baseline Lexile measure, race/ethnicity, sex, Individualized Education Program status, district, school, age at baseline, weeks of fall instruction prior to post-testing, SRI scores, and summer reading survey responses.

Appendix K. Summer reading survey results

The summer reading survey was administered concurrently with the Scholastic Reading Inventory (SRI) at posttest and included a total of seven questions. Table K-1 describes the survey results by question, with results for three questions about the presence of print material collapsed. (The summer reading survey also included an open-ended question on how many books the students had read over the summer; this question is examined in chapter 5 as an exploratory analysis.) No differences in responses between the treatment and control groups were statistically significant.

Table K-1. Summer reading survey results by study group, 2009

Summer activities and student characteristics	*Treatment* (n=*750)*		*Control* (n=*734)*				
	n	*Mean*	n	*Mean*	*Difference*	z-*score*[b]	p-*value*
Went to summer school	748	0.39	733	0.36	0.03	1.39	0.183
					(0.02)		
Went to the library	748	0.61	734	0.62	–0.01	–0.13	0.898
					(0.02)		
Reported having some print material[a] in the home	750	0.95	734	0.93	0.02	1.85	0.067
					(0.01)		
Reported that someone in the house reads	721	0.90	702	0.90	0.00	–0.51	0.608
					(0.01)		

Note: Numbers in parentheses are standard errors. Data are reported as proportions. Note that sample size varies by question as not all students addressed every question within the summer reading survey; analyses were conducted for nonmissing data only.

a. Newspapers, magazines, and children's books.

b. Calculated on unrounded values and may differ from results using values reported in this table.

Source: Summer reading survey.

Appendix L. Models used for primary, sensitivity, and exploratory analyses

Chapter 2 of the report briefly describes the statistical methods used in the analyses. All regression models were run using Stata/SE 10.0. Mim, a prefix command in Stata for analyzing multiply imputed data, was used to combine the results across the 10 imputed datasets (Royston 2004, 2005a,b). This appendix presents the estimation models used in the confirmatory, exploratory, and sensitivity analyses. The baseline Lexile measure was centered on the average baseline score of study participants in all models that included it.

Primary confirmatory impact analysis

This section presents the estimation model used in the confirmatory and corresponding sensitivity analyses.

Primary analysis

The model for assessing the confirmatory hypothesis is:

$$Y_i = \pi_0 + \pi_1 * (Baseline)_i + \pi_2 * (SRP)_i + \pi_3 * (Black)_i + \sum_{g=2}^{G} \pi_g (school\_g)_i + e_i$$

where:

- Y_i is the final Scholastic Reading Inventory (SRI) score of student *i*.
- $Baseline_i$ is the baseline Lexile measure of student *i*.
- SRP_i is a dummy variable representing whether the student was assigned to the treatment group.
- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- π_0 is the expected outcome of students, controlling for other covariates in the model.
- π_1 is the influence of the baseline Lexile measure covariate on the outcome of student *i*, controlling for other covariates in the model.
- π_2 is the mean difference in the outcome between students in the treatment and control groups, controlling for other covariates in the model.

- π_3 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- *school_g*$_i$, *g* = 2, 3, …, *G*, are (*G*–1) dummy variables representing the schools, with *school_1* as the omitted reference school.
- π_g = 2, 3, …, *G*, represents the (*G*–1) fixed school effects for the *G* schools.
- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

Chapter 4 reports the results of the primary analysis of whether students in the treatment group had scores that were significantly different, on average, than students in the control group. Confidence intervals and effect size of the impact estimate are also reported. Specifically, the effect size was computed as a standardized mean difference (Hedges' *g*) by dividing the adjusted mean difference (π_2) by the unadjusted pooled within-group standard deviation of the outcome.

Sensitivity analyses

1. *Multiple imputation for missing data.* The confirmatory impact model was estimated using multiple imputation instead of casewise deletion for missing data.
2. *Analysis without covariates.* The confirmatory impact model was estimated without the Lexile measure score or the Black race/ethnicity binary variable included as baseline covariates.
3. *Alternative covariate for baseline reading.* The confirmatory impact model was estimated using an alternative covariate for baseline reading comprehension. In particular, the Texas Assessment of Knowledge and Skills (TAKS) scaled, rather than the Lexile measure, was used as a covariate for baseline reading comprehension.

$$Y_i = \pi_0 + \pi_1 * (AlternativeBaseline)_i + \pi_2 * (SRP)_i + \pi_3 * (Black)_i + \sum_{g=2}^{G} \pi_g (school\_g)_i + e_i$$

where:

- Y_i is the final SRI score of student *i*.
- *AlternativeBaseline*$_i$ is the alternative baseline covariate of student *i*, coded using the TAKS scaled score.
- *SRP*$_i$ is a dummy variable representing whether the student *i* was assigned to the treatment group.

- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- π_0 is the average outcome of students, controlling for other covariates in the model.
- π_1 is the influence of the alternative baseline covariate on the outcome of student *i*, controlling for other covariates in the model.
- π_2 is the mean difference in the outcome between students in the treatment and control groups, controlling for other covariates in the model.
- π_3 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- $school\_g_i$, g = 2, 3, …, G, are (G–1) dummy variables representing the G schools, with *school_1* as the omitted reference school.
- π_g = 2, 3, …, G, represents the (G–1) fixed school effects for the G schools.
- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

4. *Additional covariate for weeks of instruction.* The confirmatory impact model was estimated using an additional covariate for weeks of instruction in the fall prior to final testing.

$$Y_i = \pi_0 + \pi_1 * (Baseline)_i + \pi_2 * (WeeksPrior)_i + \pi_3 * (SRP)_i + \pi_4 * (Black)_i + \sum_{g=2}^{G} \pi_g (school\_g)_i + e_i$$

where:

- Y_i is the final SRI score of student *i*.
- $Baseline_i$ is the baseline Lexile measure of student *i*.
- $WeeksPrior_i$ is the weeks prior to posttesting for student *i*.
- SRP_i is a dummy variable representing whether the student was assigned to the treatment group.
- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity). π_0 is the average outcome of students, controlling for other covariates in the model.

- π_1 is the influence of baseline Lexile measure on the outcome of student *i*, controlling for other covariates in the model.
- π_2 is the effect of weeks prior to posttesting on the outcome of student *i*, controlling for other covariates in the model.
- π_3 is the mean difference in the outcome between students in the treatment and control groups, controlling for other covariates in the model.
- π_4 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- $school\_g_i$, g = 2, 3, …, G, are (G–1) dummy variables representing the G schools, with *school_1* as the omitted reference school.
- π_g = 2, 3, …, G, represents the (G–1) fixed school effects for the G schools.
- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

5. *Including interaction of treatment status and school (with equal weighting).* The impact of the summer reading program was also estimated by including interaction terms between treatment status and each school to provide school-specific impact estimates, which were then aggregated to obtain the overall impact. The model is:

$$Y_i = \pi_0 * (Baseline)_i + \pi_2 * (Black)_i + \sum_{g=1}^{G} \pi_g (school\_g)_i + \sum_{g=1}^{G} \lambda_g \left[(school\_g) * (\mathrm{SRP}) \right]_i + e_i$$

where:

- Y_i is the outcome for student *i*.
- $Baseline_i$ is the baseline Lexile measure of student *i*.
- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- $school\_g_i$, g = 1, 2, …, G, are G dummy variables representing the G schools.
- $(school\_g*SRP)_i$, g = 1, 2, …, G are G interaction terms between the G schools and treatment status.
- π_0 is the influence of the baseline Lexile measure on the outcome of student *i*.
- π_2 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.

- π_g = 1, 2, …, *G*, represents the *G* fixed school effects for the *G* schools.
- λ_g = 1, 2,…, G, represents the influence of the interaction between each of the *G* schools and treatment status.
- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

The school-specific effects were weighted as follows:

$$\text{Combined estimate} = \sum_{g=1}^{G} w_g \lambda_g$$

where:

- w_i is the weight for equal weighting, and $w_g = 1/G$ for all *g*.
- λ_g is the school specific estimate of the treatment effect

$$\sum_{g=1}^{G} w_g = 1.$$

The standard error of the combined estimate is calculated as:

$$SE = \sqrt{\sum_{g=1}^{G} w_g^2 * \text{var}(\lambda_g) + 2 * \sum_{g<j} \sum w_g * w_j * \text{cov}(\lambda_g, \lambda_j)}\ .$$

6. *Including interaction of treatment status and school (with precision weighting).* The impact of the summer reading program was also estimated as described for sensitivity analysis 6, but with precision weighting to aggregate the school-specific impact estimates. In particular, for school *g*, $w_g = \dfrac{\left(\dfrac{1}{se_g^2}\right)}{\left(\sum_{j=1}^{G} \dfrac{1}{se_j^2}\right)}$.

7. *Random effect.* To account for within school clustering, the following model of schools as random effects was run:

Level 1 (student-level)

$$Y_{ij} = \pi_{0j} + \pi_{1j}*(Baseline)_{ij} + \pi_{2j}*(SRP)_{ij} + \pi_{3j}*(Black)_{ij} + e_{ij}$$

where:

- Y_{ij} is the final SRI score for student *i* in school *j*.

- $Baseline_{ij}$ is the baseline Lexile measure of student *i* in school *j*.
- SRP_{ij} is a dummy variable representing whether the student *i* was assigned to the treatment group in school *j*.
- $Black_{ij}$ is a dummy variable representing whether the student *i* in school *j* was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- π_{0j} is the outcome of students in school *j* for the control students when the covariate for students with average baseline Lexile measures.
- π_{1j} is the influence of the baseline Lexile measure on the outcome in school *j* for control students.
- π_{2j} is the mean difference in the outcome between students in the treatment and control groups for students with average baseline Lexile measures in school *j*.
- π_{3j} is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- e_{ij} is a random error associated with student *i* in school *j*; $e_{ij} \sim N(0, \sigma^2)$.

The student intercept (π_0) and treatment slope (π_2) estimated from the above model will be modeled as varying randomly across schools, as shown in the following level-2 specification:

Level 2 (school-level)

(1) $\pi_{0j} = \beta_{00} + r_{0j}$

(2) $\pi_{1j} = \beta_{10}$

(3) $\pi_{2j} = \beta_{20} + r_{2j}$

(4) $\pi_{3j} = \beta_{30}$ where:

- β_{00} is the expected outcome for control students with average baseline Lexile measures.
- r_{0j} is a random error associated with school *j* on student outcomes at intercept.
- β_{10} is the average influence of the baseline Lexile measures across all schools.
- β_{20} is the average difference between treatment and control students across all schools, adjusted for student baseline Lexile measures.
- r_{2j} is a random error associated with the treatment effect.

$$\begin{pmatrix} r_{0j} \\ r_{2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right].$$

Exploratory research question 1

This section presents the estimation model used in the exploratory research and corresponding sensitivity analyses.

Primary analysis

Given the positively skewed outcome variable for the number of books students reported reading over the summer, the variable was recoded into five categories. An ordered probit model was used to estimate how much treatment status affected the probability of students belonging to a particular category of number of books read.

If Y^*_i is the number of books student *i* reported reading during the summer, categorical variable Y_i is defined such that:

$$Y_i = \begin{cases} 1 \text{ if } Y_i^* = 0 \\ 2 \text{ if } 1 \le Y_i^* < 5 \\ 3 \text{ if } 5 \le Y_i^* < 9 \\ 4 \text{ if } 9 \le Y_i^* < 13 \\ 5 \text{ if } Y_i^* \ge 13 \end{cases}$$

Then the ordered probit regression is:

$$\Pr(Y_i \le k) = \Phi\left[\alpha_k + \pi_1 * (Baseline)_i + \pi_2 * (SRP)_i + \pi_3 * (Black)_i + \pi_4 * (Hispanic)_i + \sum_{g=2}^{G} \gamma_g * (school\_g)_i\right]$$

for k = 1,2, or 3

where:

- Φ is the cumulative distribution function for the standard normal, for example, $N(0,1)$.
- Y_i is the number of books read (on a 1–5 scale) for student i.
- $Baseline_i$ is the baseline Lexile measure of student i.
- SRP_i is a dummy variable representing whether the student was assigned to the treatment group.
- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- $Hispanic_i$ is a dummy variable representing whether the student was Hispanic (1 = Hispanic race/ethnicity; 0= other race/ethnicity).
- α_k is the average outcome of students, controlling for other covariates in the model.
- π_1 is the influence of the baseline Lexile measure on the outcome of student i, controlling for other covariates in the model.
- π_2 is the difference in the outcome between students in the treatment and control groups, controlling for other covariates in the model.
- π_3 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- π_4 is the mean difference in the outcome between students who are Hispanic and students who are not Hispanic, controlling for other covariates in the model.
- school_g_i, g = 2, 3, …, G, are (G–1) dummy variables representing the G schools, with *school_1* as the omitted reference school.
- γ_g = 2, 3, …, G, represents the (G–1) fixed school effects for the G schools.
- e_i is a random error associated with student i; $e_i \sim N(0, \sigma^2)$.

Sensitivity analysis

The impact of the summer reading program on the number of books read over the summer was tested using ordinary least squares (OLS) regression[47] and the following model:

$$Y_i = \pi_0 + \pi_1 * (Baseline)_i + \pi_2 * (SRP)_i + \pi_3 * (Black)_i + \pi_4 * (Hispanic)_i + \sum_{g=2}^{G} \pi_g (school\_g)_i + e_i$$

where:

- Y_i is the square root of the number of books read for student *i*.
- $Baseline_i$ is the baseline Lexile measure of student *i*.
- SRP_i is a dummy variable representing whether the student was assigned to the treatment group.
- $Black_i$ is a dummy variable representing whether the student was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- $Hispanic_i$ is a dummy variable representing whether the student was Hispanic (1 = Hispanic race/ethnicity; 0= other race/ethnicity).π_0 is the average outcome of students, controlling for other covariates in the model.
- π_1 is the influence of the baseline Lexile measure on the outcome of student *i*, controlling for other covariates in the model.
- π_2 is the difference in the outcome between students in the treatment and control groups, controlling for other covariates in the model.
- π_3 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.
- π_4 is the mean difference in the outcome between students who are Hispanic and students who are not Hispanic, controlling for other covariates in the model.
- $school\_g_i$, g = 2, 3, …, G, are (G–1) dummy variables representing the G schools, with *school_1* as the omitted reference school.
- π_g = 2, 3, …, G, represents the (G–1) fixed school effects for the G schools.

[47] Because the distribution of books was positively skewed, a square-root transformation was used on the outcome score. Due to implausibly high estimates, 1 percent trimming was used followed by multiple imputation to obtain an intent-to-treat analysis.

- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

Of primary interest is the main effect of SRP_i, which represents the summer reading program's effect on the square root of the number of books students reported reading over the summer. Baseline reading comprehension was controlled for by including baseline Lexile measures as a covariate in the model. A statistically significant value of the main effect for SRP_i would indicate that the summer reading program had a statistically significant impact on the square root of the number of books read over the summer, regardless of baseline reading comprehension, Black race/ethnicity, and Hispanic race/ethnicity.

Exploratory research question 2

The potential differential impact of the study's summer reading program on a student's reading comprehension, depending on baseline reading comprehension group membership as measured by the baseline Lexile measure, was tested using multiply imputed OLS regression. Two equations were used, changing the reference group to obtain all relevant comparisons between thirds of the distribution. In the first equation, the middle group was used as the reference group; in the second equation, the bottom group was used. The first is shown below:

$$Y_i = \pi_0 + \pi_1*(BaselineBottomThird)_i + \pi_2*(BaselineTopThird)_i + \pi_3*(SRP)_i + \pi_4*(BaselineBottomThird*SRP)_i + \pi_5*(BaselineTopThird*SRP)_i + \pi_6*(Black)_i + \sum_{g=2}^{G} \pi_g (school\_g)_i + e_i$$

where:

- Y_i is the final SRI score for student *i*.
- $BaselineBottomThird_i$ is a dummy variable indicating that student *i* is in the bottom third of the baseline Lexile measure distribution.
- $BaselineTopThird_i$ is a dummy variable indicating that student *i* is in the top third of the baseline Lexile measure distribution.
- SRP_i is a dummy variable representing whether the student was assigned to the treatment group.
- $Black_i$ is a dummy variable representing whether the student *i* was Black (1 = Black race/ethnicity; 0= other race/ethnicity).
- $(BaselineBottomThird*SRP)_i$ is the interaction term between the dummy variable indicating that student *i* was in the bottom third of the baseline Lexile measure distribution and the dummy variable representing whether the student was assigned to the treatment group.

- $(BaselineTopThird*SRP)_i$ is the interaction term between the dummy variable indicating that student *i* was in the top third of the baseline Lexile measure distribution and the dummy variable representing whether the student was assigned to the treatment group.

- π_0 is the average outcome of students when all other covariates in the model are set to zero.

- π_1 is the difference between being in the bottom third and being in the middle third of the baseline Lexile measure distribution on the outcome of student *i* when all other covariates in the model are set to zero.

- π_2 is the difference between being in the top third and being in the middle third of the baseline Lexile measure distribution on the outcome of student *i* when all other covariates in the model are set to zero.

- π_3 is the difference in the outcome between students in the treatment and control groups when all other covariates in the model are set to zero.

- π_4 is the differential effect of the summer reading program depending on whether students were in the bottom third rather than the middle third of the baseline Lexile measure distribution when all other covariates in the model are set to zero.

- π_5 is the differential effect of the summer reading program depending on whether students were in the top third rather than the middle third of the baseline Lexile measure distribution when all other covariates in the model are set to zero.

- π_6 is the mean difference in the outcome between students who are Black and students who are not Black, controlling for other covariates in the model.

- $school\_g_i$, $g = 2, 3, \ldots, G$, are $(G–1)$ dummy variables representing the G schools, with *school_1* as the omitted reference school.

- $\pi_g = 2, 3, \ldots, G$, represents the $(G–1)$ fixed school effects for the G schools.

- e_i is a random error associated with student *i*; $e_i \sim N(0, \sigma^2)$.

Of primary interest are the coefficients on the $(BaselineBottomThird*SRP)_i$ and $(BaselineTopThird*SRP)_i$ interaction terms. These interaction terms examine whether the treatment effects themselves (treatment vs. control) are different based on which third of the Lexile distribution the student is in. In particular, a statistically significant value of the coefficient on the interaction term, π_4 (shown in table M-15 as "Lower third X treatment"). indicates that the effect of the summer reading program on the outcome score for reading

comprehension was significantly different for students in the bottom third rather than the middle third of the distribution for baseline reading comprehension scores. A statistically significant value of the coefficient on the interaction term, π_5 (shown in table M-15 as "Upper third X treatment"), indicates that the effect of the summer reading program on the outcome score for reading comprehension was significantly different for students in the top third rather than the middle third of the distribution for baseline reading comprehension scores. The comparison between the lower and upper thirds is given in the second equation (not shown) in which the lower third was the reference group. These results are shown in table M-16.

Appendix M. Tables of analytic output[48]

Table M-1. Primary confirmatory impact model using casewise deletion for missing data (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	4.89	9.83	0.62	9.72	0.62
Baseline Lexile measure	0.75	0.04	<0.01	0.04	<0.01
Black	5.73	13.37	0.67	14.13	0.69
Constant	305.39	73.27	<0.01	64.46	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black race/ethnicity, and schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-2. Primary confirmatory question sensitivity analysis 1, using multiple imputation for missing data (*n*=1,785)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	9.19	9.41	0.33	9.37	0.327
Baseline Lexile measure	0.76	0.04	<0.01	0.04	<0.01
Black	7.54	13.30	0.57	13.97	0.590
Constant	302.76	73.73	<0.01	64.46	<0.01

Note: The results in this table are based on 10 multiply imputed datasets analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black race/ethnicity, and schools. The estimates for school fixed effects are excluded.

[48] An examination of the model assumptions found that the residuals were heteroscedastic. In particular, the Breusch-Pagan/Cook-Weisberg test for heteroscedasticity was run for models using OLS regression in the study and these were found to be statistically significant (i.e., rejecting the hypothesis that the residuals were homoscedastic). Therefore, robust standard errors have been calculated for each analysis and added to each table.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-3. Primary confirmatory question sensitivity analysis 2, without covariates (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	6.24	11.14	0.58	11.10	0.57
Constant	295.01	83.20	<0.01	61.06	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-4. Primary confirmatory question sensitivity analysis 3, with alternative covariate for baseline reading (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	4.11	9.86	0.68	9.75	0.67
Baseline Texas Assessment of Knowledge and Skills scaled score (recode)	1.03	0.05	<0.01	0.06	<0.01
Black	5.62	13.42	0.68	14.13	0.69
Constant	305.42	73.51	<0.01	63.74	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black race/ethnicity, and schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-5. Primary confirmatory question sensitivity analysis 4, with additional covariate for weeks of instruction (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	5.25	9.80	0.59	9.71	0.59
Baseline Lexile measure	0.74	0.04	0.00	0.04	<0.01
Black	0.89	13.43	0.95	14.16	0.95
Weeks of instruction	6.62	2.19	<0.01	2.15	<0.01
Constant	235.77	76.61	<0.01	67.65	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black race/ethnicity, weeks of instruction, and schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-6. Primary confirmatory question sensitivity analyses 5 and 6, including interaction of treatment status and school (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Baseline Lexile measure	0.76	0.04	<0.01	0.04	<0.01
Black	6.66	13.87	0.63	15.14	0.66
Treatment X School 1	4.07	146.72	0.98	125.91	0.97
Treatment X School 2	52.92	271.53	0.85	2.88	<0.01
Treatment X School 3	187.52	58.69	<0.01	62.14	<0.01
Treatment X School 4	-341.62	271.93	0.21	16.12	<0.01
Treatment X School 5	-34.25	59.32	0.56	43.27	0.43
Treatment X School 6	30.24	88.22	0.73	132.64	0.82
Treatment X School 7	56.96	146.67	0.70	135.46	0.67
Treatment X School 8	-2.50	128.79	0.99	113.66	0.98
Treatment X School 9	230.00	192.05	0.23	68.27	<0.01
Treatment X School 10	1.20	78.72	0.99	81.55	0.99
Treatment X School 11	-116.22	156.80	0.46	93.22	0.21
Treatment X School 12	273.36	271.53	0.31	2.47	<0.01
Treatment X School 13	-177.34	116.31	0.13	83.31	0.03
Treatment X School 14	70.91	157.06	0.65	119.71	0.55

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 15	62.16	271.57	0.82	5.76	<0.01
Treatment X School 16	-126.77	175.55	0.47	66.21	0.06
Treatment X School 17	-105.37	112.44	0.35	111.19	0.34
Treatment X School 18	-129.96	175.31	0.46	82.39	0.11
Treatment X School 19	103.83	121.48	0.39	102.98	0.31
Treatment X School 20	7.44	175.41	0.97	128.60	0.95
Treatment X School 21	215.87	192.02	0.26	101.17	0.03
Treatment X School 22	19.41	99.43	0.85	102.56	0.85
Treatment X School 23	-46.53	175.52	0.79	225.03	0.84
Treatment X School 24	68.73	76.87	0.37	72.66	0.34
Treatment X School 25	-18.67	49.76	0.71	58.01	0.75
Treatment X School 26	107.63	88.22	0.22	80.28	0.18
Treatment X School 27	-323.84	166.44	0.05	106.22	<0.01
Treatment X School 28	-243.76	235.50	0.30	203.27	0.23
Treatment X School 29	-73.47	88.39	0.41	113.01	0.52
Treatment X School 30	-174.69	175.28	0.32	62.46	0.01
Treatment X School 31	-33.23	96.76	0.73	98.24	0.74

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 32	-46.50	78.72	0.56	68.50	0.50
Treatment X School 33	40.78	51.87	0.43	39.46	0.30
Treatment X School 34	-173.06	146.72	0.24	254.27	0.50
Treatment X School 35	109.81	106.83	0.30	80.72	0.17
Treatment X School 36	55.71	235.37	0.81	86.63	0.52
Treatment X School 37	-117.78	110.90	0.29	102.45	0.25
Treatment X School 38	-126.56	55.43	0.02	44.94	<0.01
Treatment X School 39	-85.00	140.65	0.55	48.32	0.08
Treatment X School 40	15.97	78.40	0.84	76.90	0.84
Treatment X School 41	-44.30	77.41	0.57	109.28	0.69
Treatment X School 42	106.06	75.31	0.16	98.35	0.28
Treatment X School 43	40.83	77.39	0.60	77.59	0.60
Treatment X School 44	150.06	235.20	0.52	74.66	0.04
Treatment X School 45	118.89	235.19	0.61	20.49	<0.01
Treatment X School 46	-96.02	175.32	0.58	290.66	0.74
Treatment X School 47	-42.66	271.66	0.88	9.67	<0.01
Treatment X School 48	4.25	80.18	0.96	81.28	0.96

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 49	-50.71	112.56	0.65	182.07	0.78
Treatment X School 50	-9.03	235.25	0.97	338.91	0.98
Treatment X School 51	121.86	83.92	0.15	82.70	0.14
Treatment X School 52	111.41	71.43	0.12	78.04	0.15
Treatment X School 53	80.55	235.29	0.73	12.17	<0.01
Treatment X School 54	-61.04	156.94	0.70	129.98	0.64
Treatment X School 55	-25.23	146.68	0.86	126.83	0.84
Treatment X School 56	-19.48	156.87	0.90	199.45	0.92
Treatment X School 57	201.72	271.56	0.46	5.35	<0.01
Treatment X School 58	-1.07	116.26	0.99	133.94	0.99
Treatment X School 59	-38.34	65.97	0.56	56.33	0.50
Treatment X School 60	-2.98	63.18	0.96	55.14	0.96
Treatment X School 61	174.87	128.99	0.18	122.36	0.15
Treatment X School 62	64.65	146.64	0.66	100.21	0.52
Treatment X School 64	67.35	82.02	0.41	69.52	0.33
Treatment X School 65	104.77	72.68	0.15	58.12	0.07
Treatment X School 66	-84.95	59.25	0.15	56.28	0.13

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 67	93.40	88.26	0.29	134.70	0.49
Treatment X School 68	-253.05	175.30	0.15	155.21	0.10
Treatment X School 69	-244.91	140.55	0.08	78.32	<0.01
Treatment X School 70	300.52	91.11	<0.01	79.34	<0.01
Treatment X School 71	-51.38	73.99	0.49	73.34	0.48
Treatment X School 72	-330.55	146.66	0.02	161.41	0.04
Treatment X School 73	113.62	146.64	0.44	83.25	0.17
Treatment X School 74	165.00	88.40	0.06	93.60	0.08
Treatment X School 75	78.38	156.84	0.62	142.76	0.58
Treatment X School 76	-115.53	70.27	0.10	69.72	0.10
Treatment X School 77	-84.48	106.86	0.43	137.20	0.54
Treatment X School 78	337.22	156.85	0.03	73.25	<0.01
Treatment X School 79	-67.52	156.83	0.67	54.17	0.21
Treatment X School 80	-20.86	88.22	0.81	84.24	0.80
Treatment X School 81	-284.86	116.26	0.01	81.78	<0.01
Treatment X School 82	-316.92	107.12	<0.01	191.44	0.10
Treatment X School 83	46.95	272.15	0.86	20.46	0.02

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 84	10.88	157.01	0.94	175.56	0.95
Treatment X School 85	86.40	116.26	0.46	137.96	0.53
Treatment X School 86	-79.24	59.27	0.18	46.40	0.09
Treatment X School 87	131.39	192.05	0.49	175.01	0.45
Treatment X School 88	233.19	235.15	0.32	47.42	<0.01
Treatment X School 89	14.54	74.06	0.84	72.29	0.84
Treatment X School 90	-120.05	192.12	0.53	43.25	0.01
Treatment X School 91	-121.34	192.51	0.53	262.70	0.64
Treatment X School 92	63.84	235.20	0.79	75.83	0.40
Treatment X School 93	95.06	70.13	0.18	56.43	0.09
Treatment X School 94	-7.33	96.17	0.94	98.81	0.94
Treatment X School 95	12.81	67.89	0.85	53.62	0.81
Treatment X School 96	55.15	146.93	0.71	121.71	0.65
Treatment X School 97	284.68	271.65	0.29	9.06	<0.01
Treatment X School 98	-34.72	71.47	0.63	76.16	0.65
Treatment X School 99	-94.77	157.10	0.55	149.85	0.53
Treatment X School 100	12.45	76.90	0.87	62.54	0.84

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment X School 101	105.71	86.33	0.22	72.91	0.15
Treatment X School 102	74.27	86.33	0.39	109.41	0.50
Treatment X School 103	-82.32	82.24	0.32	94.26	0.38
Treatment X School 104	99.69	103.69	0.34	95.83	0.30
Treatment X School 105	160.84	116.29	0.17	81.37	0.05
Treatment X School 106	338.41	192.00	0.08	55.32	<0.01
Treatment X School 107	-139.21	235.15	0.55	22.50	<0.01
Treatment X School 108	-36.05	97.05	0.71	134.30	0.79
Treatment X School 109	-31.62	192.22	0.87	173.45	0.86

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black race/ethnicity, and schools. The data displayed in this table represent the school-specific impact estimates for each school. The estimates for school fixed effects are excluded.

Source. District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-7. Primary confirmatory question sensitivity analysis 7, mixed effects model with random effects for intercept and treatment (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*
Fixed effects			
Intercept	328.23	11.15	<0.01
Treatment	8.81	11.65	0.45
Baseline TAKS scores	0.77	0.04	<0.01
Black	-4.28	12.98	0.74
Random effects			
Intercept 1 (R0)	77.74	10.02	
Level-1 E	191.93	3.64	
Treatment (R2)	53.76	14.50	

TAKS is Texas Assessment of Knowledge and Skills.

Note: The results in this table are based on a dataset with casewise deletion analyzed using a mixed effects model with students nested in schools, using random effects for the intercept and treatment effect and adjusting for centered baseline Lexile measure and Black race/ethnicity. The data in this table represent the impact estimate of the summer reading program on student reading comprehension using a random effects model to take into account within school clustering of student performance on the SRI, as well as within school clustering of the treatment effect.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-8. Exploratory research question 1ordered probit, casewise deletion (*n*=1,447)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	0.14	0.06	0.01	0.06	0.01
Baseline Lexile measure	<0.01	<0.01	0.16	<0.01	0.17
Black	-0.35	0.11	<0.01	0.12	<0.01
Hispanic	-0.26	0.10	0.01	0.10	0.01
μ_1	-0.36	0.48		0.72	
μ_2	0.79	0.48		0.72	
μ_3	1.44	0.48		0.72	
μ_4	2.03	0.48		0.72	

Note: The results in this table are based on a dataset with casewise deletion analyzed using an ordered probit model that adjusted for centered baseline Lexile measure, Black and Hispanic race/ethnicity, and schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

Table M-9. Exploratory research question 1 sensitivity analysis, OLS regression, square root transformation, 1 percent trimmed mean, casewise deletion (*n*=1,432)

Varible	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Treatment	0.18	0.08	0.02	0.08	0.02
Baseline Lexile measure	<0.01	<0.01	0.13	<0.01	0.13
Black	-0.47	0.15	<0.01	0.16	<0.01
Hispanic	-0.35	0.13	0.01	0.14	0.01
Constant	1.24	0.61	0.04	0.61	0.04

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measure, Black and Hispanic race/ethnicity, and schools. The estimates for school fixed effects are excluded.

Source: District-provided data collected May 2009–June 2009; summer reading survey data collected September 2009–December 2009.

Table M-10. Exploratory research question 2, with middle third baseline Lexile measure group as reference category (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Lower third	-152.44	17.53	<0.01	17.49	<0.01
Upper third	71.29	18.6	<0.01	17.18	<0.01
Treatment	-1.65	17.12	0.92	16.48	0.92
Lower third X treatment	11.47	24.5	0.64	24.88	0.64
Upper third X treatment	12.19	26.58	0.65	25.44	0.63
Black	1.58	13.83	0.91	14.61	0.91
Constant	351.16	76.54	<0.01	58.65	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for centered baseline Lexile measures, Black race/ethnicity, and schools. The reference group was the middle third on baseline Lexile measures. The estimates for school fixed effects are excluded.

[a]A statistically significant result for these coefficients would mean that the difference between the treatment and control group (that is, the impact estimate) differs by baseline reading comprehension.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

Table M-11. Exploratory research question 2, with bottom third baseline Lexile measure group as reference category (*n*=1,571)

Variable	*Coefficient*	*SE*	p-*value*	*Robust SE*	*Robust* p-*value*
Middle third	152.44	17.53	<0.01	17.49	<0.01
Upper third	223.73	18.51	<0.01	17.99	<0.01
Treatment	9.82	17.22	0.57	18.28	0.59
Middle third X treatment	-11.47	24.50	0.64	24.88	0.65
Upper third X treatment	0.72	26.44	0.98	26.35	0.98
Black	1.58	13.83	0.91	14.61	0.91
Constant	198.72	76.52	0.01	59.00	<0.01

Note: The results in this table are based on a dataset with casewise deletion analyzed using ordinary least squares regression that adjusted for Black race/ethnicity and schools. The reference group was the bottom third on baseline Lexile measures. The estimates for school fixed effects are excluded.

[a]A statistically significant result for these coefficients would mean that the difference between the treatment and control group (that is, the impact estimate) differs by baseline reading comprehension.

Source: District-provided data collected May 2009–June 2009; Scholastic Reading Inventory data collected September 2009–December 2009.

References

Albuquerque/Bernalillo (NM) County Library System. (2010). *Be creative at your library*. Retrieved January 5, 2010, from http://www.cabq.gov/library/src2.html

Alexander, K. L., Entwisle, D. R., and Olson, L.S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72,* 167–80.

Alllington, R.L., Guice, S., Baker, K., Michelson, N., and Li, S. (1995). Access to books: Variations in schools and classrooms. *Language and Literacy Spectrum*, 5 (spring), 23–25.

Allington, R. L. and McGill-Franzen, A. (2003). The impact of summer setback on the reading achievement gap. *The Phi Delta Kappan, 85,* 68–75.

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., et al. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411–27.

Anderson, A. and S. Stokes. (1984). Social and institutional influences on the development and practice of literacy. In H. Goelman, A. Oberg, and F. Smith, (Eds.) *Awakening to Literacy* (pp. 24–37). Portsmouth, NJ: Heinemann.

Anonymous. (1910). San Antonio Carnegie Library. *Library Journal, 35*(8), 379.

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage.

Bloom, H. S., Richburg-Hayes, L., and Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Book Industry Study Group. (2011). *BISAC committees*. http://www.bisg.org/committee-cat-2-bisac-committees.php (accessed April 20, 2011).

Bracey, G. W. (2002). Summer loss: The phenomenon no one wants to deal with. *Phi Delta Kappan, 84,* 12–14.

Burns, M. S., Griffin, P., and Snow, C. (Eds.). (1999). *Starting out right: A guide to promoting children's reading success.* Washington, DC: National Academy Press.

Butler, T. L. (2010). *The effects of summer reading programs on low-income English language learners' and low-income English L1s' reading achievement in grades 2–4*. Digital dissertation, University of Arkansas. Ann Arbor, MI: ProQuest LLC.

Carlson, C. D. (1998). Socioeconomic status and reading achievement: The mediating role of home processes and pre-reading skills (No. AAM9835726). *Dissertation Abstracts International, 59(5-A),* 1509.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Conger, R. D. and Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology, 58,* 175–99.

Constantino, R. (2005). Print environments between high and low socioeconomic status communities. *Teacher Librarian, 32,* 22–25.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., and Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66,* 227–68.

Crowell, D. and Klein, T. (1981). Preventing summer loss of reading skills among primary children. *The Reading Teacher, 34,* 561–64.

Dallas Independent School District. (n.d.). *Summer reading.* Retrieved June 1, 2010, from http://teacherweb.com/PA/DallasSchoolDistrict/DistrictHomePage/faq3.aspx

Dallas Public Library. (2010). Mayor's summer reading program. Retrieved January 5, 2010, from http://dallaslibrary.org/summerReading/Program

David, J. (1979). *A descriptive study of Title I summer programs.* Menlo Park, CA: Stanford Research Institute.

Elley, W. B., and Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly, 19,* 53–67.

Fiore, C. D. (1998). *Running summer library reading programs: A how to do it manual*. New York: Neal-Schuman Publishers.

Fountas, I. C., and Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Fountas, I. C., and Pinnell, G. S. (2001). *Guiding readers and writers grades 3–6: Teaching comprehension, genre, and content literacy*. Portsmouth, NH: Heinemann.

Fry, E. (2002). Readability versus leveling. *The Reading Teacher, 56,* 286–91.

Fryer, R.G., and Levitt, S.D. (2002). *Understanding the black-white test score gap in the first two years of school.* Cambridge, MA: National Bureau of Economic Research.

Halle, B., Kurtz-Costes, B., and Mahoney, J. L. (1997). Family influences on school achievement in low-income, African American children. *Journal of Educational Psychology, 89,* 527–37.

Harcourt Assessment, Inc. (2003). *Stanford Achievement Test Series, Tenth Edition.* San Antonio, TX: Pearson Education.

Harlwell, M. R., Maeda, Y., and Lee, K. (2004, April). *Replicating and extending White's (1982) meta-analysis of the relationship between SES and student achievement.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Hart, B., and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Company.

Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

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

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development, 58,* 1151–61.

Hiebert, E.H., and Reutzel, D.R. (Eds.). (2010). *Revisiting silent reading: new directions for teachers and researchers.* Newark, DE: International Reading Association.

Hoff E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1368–78.

Houston Independent School District. (2010). *Houston ISD millionaire club: Summer reading program 2010.* Retrieved June 1, 2010, from http://www.hisdlibraryservices.org/millionaire/index.htm

Kim, J. S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized trial. *Educational Evaluation and Policy Analysis, 28,* 335–55.

Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology, 99,* 505–15.

Kim, J. S., and 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.

Kim, J., and 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,* 20–31.

Knutson, K. A. (2008). *Using the SRI to predict reading performance levels in California: An alignment guide to the California standards test.* New York: Scholastic, Inc. Retrieved June 3, 2010, from http://teacher.scholastic.com/products/sri_reading_assessment/pdfs/SRI_ProfPaper_PredictPerformance.pdf

Krashen, S. (1995). School libraries, public libraries, and the NAEP reading scores. *School Library Media Quarterly, 23,* 235–37.

Kruger, G. L., and Gilroy, A. A. (2009). Broadband internet access and the digital divide: Federal assistance programs. *Congressional Research Service.* Retrieved June 2, 2010, from http://www.nationalaglawcenter.org/assets/crs/RL30719.pdf

Lee, J., Grigg, W., and Donahue, P. (2007). *The Nations Report Card: Reading 2007* (NCES 2007-496). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Lennon, C., and Burdick, H. (2004). The Lexile Framework as an approach for reading measurement and success. *The Lexile Framework for Reading.* Retrieved August 17, 2009, from http://www.lexile.com/PDF/Lexile-Reading-Measurement-andSuccess-0504.pdf

Leslie, L., and Caldwell, J. (2006). *Qualitative reading inventory–4*. Upper Saddle River, NJ: Pearson Education.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Louisiana Department of Education. (n.d.). *Educational improvement and assistance: Summer reading list and links.* Retrieved August 3, 2010, from http://www.doe.state.la.us/lde/eia/2689.html

Luftig, R. L. (2003). When a little means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers. *Reading Research and Instruction, 42,* 1–13.

Malach, D. A., and Rutter, R. A. (2003). For nine months kids go to school, but in summer this school goes to kids. *The Reading Teacher, 57,* 5–10.

Manning, M., Lewis, M., and Lewis, M. (2010). Sustained Silent Reading: An Update of the Research. In E.H. Hiebert, & D. Reutzel (Eds.), *Revisiting Silent Reading* (pp. 112–28). Newark, DE: International Reading Association.

McGill-Franzen, A.M., Lanford, C., and Adams, E. (2002). Learning to be literate: A comparison of five urban early childhood programs. *Journal of Educational Psychology,* 94(3), 443–64.

McGill-Franzen, A., and Allington, R. (2003). Bridging the summer reading gap. *Instructor Magazine, 58,* 17–19.

McGill-Franzen, A., Allington, R. L., Solic, K., Williams, L., Zmach, C., Graff, J., et al. (2005). The rich/poor achievement gap: Contributions of summer reading loss. *Tennessee Reading Teacher, 34*(1), 28–30.

McQuillan, J., and Au, J. (2001). The effect of print access on reading frequency. *Reading Psychology, 22,* 225–48.

MetaMetrics, Inc. (n.d.a). *Durham reads.* Retrieved June 3, 2010, from http://www.lexile.com/using-lexile/summer-reading/current-summer-reading-programs-with-lexile/durham-reads/

MetaMetrics, Inc. (n.d.b). *How to get a Lexile measure.* Retrieved August 3, 2010, from http://www.lexile.com/about-lexile/How-to-get-lexile-measures/

MetaMetrics, Inc. (n.d.c). *The Lexile Framework for Reading: Find the right book for you.* Retrieved February 4, 2010, from http://www.lexile.com/findabook

MetaMetrics, Inc. (n.d.d). *What is a Lexile measure?* Retrieved January 20, 2011, from http://lexile.com/about-lexile/lexile-overview/

MetaMetrics, Inc. (2005). *Linking the TAKS with the Lexile Framework.* Durham, NC: MetaMetrics, Inc.

MetaMetrics, Inc. (2008). *What does the Lexile measure mean?* Retrieved March 5, 2010, from http://www.lexile.com/m/uploads/downloadablepdfs/WhatDoestheLexileMeasureMean.pdf

Mraz, M., and Rasinski, T. V. (2007). Summer reading loss. *The Reading Teacher, 60,* 784–89.

National Center for Education Statistics. (n.d.). *Common Core of Data*. Washington, DC: U.S. Department of Education. Retrieved August 25, 2009, from http://nces.ed.gov/ccd/

National Research Council (Ed.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Neuman, S. (1986). The home environment and fifth-grade students' leisure reading. *Elementary School Journal, 86,* 335–43.

Neuman, S. B., Celano, D. C., Greco, A. N., and Shue, P. (2001). *Access for all: Closing the book gap for children in early education.* Newark, DE: International Reading Association.

New Orleans Public Library. (2010). *Be creative.* Retrieved January 5, 2010, from http://www.neworleanspubliclibrary.org/~nopl/programming/05_09/srp2009schedule.pdf

Pearson Education, Inc. (n.d.). *Texas assessment: Pearson access.* Retrieved March 16, 2011, from https://www.pearsonaccess.com/cs/Satellite?pagename=Pearson/QuickLink/tx

Perie, M., Grigg, W., and Donahue, P. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Puma, M. J., Karweit, N., Price, C., Ricciuti, A. E., Thompson, W., and Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes.* Bethesda, MD: Abt Associates.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models, *Survey Methodology, 27*(1), 85–95.

Raudenbush, S. W., Martinez, A., and Spybrook, J. (2005). *Strategies for improving precision in group randomized experiments.* New York: W.T. Grant Foundation.

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal, 4,* 227–41.

Royston, P. (2005a). Multiple imputation of missing values: Update. *Stata Journal, 5,* 188–201.

Royston, P. (2005b). Multiple imputation of missing values: Update of ice. *Stata Journal, 5,* 527–36.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman and Hall.

Schiefele, U. (1999). Interest and learning from text. *Scientific studies of reading, 3*(3), 257–79.

Scholastic, Inc. (1999). *Scholastic Reading Inventory.* New York: Scholastic, Inc.

Scholastic, Inc. (2007). Scholastic Reading Inventory technical guide working with student performance data. New York: Scholastic, Inc.

Scholastic Inc. (2007). Scholastic Reading Inventory Technical Guide. New York: Scholastic Inc. Education Group.

Scholastic, Inc. (2008). *2008 Kids and family reading report*. New York: Scholastic, Inc.

Snow, C. E., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stenner, A., Smith,D., Horabin, I., and Smith, M. (1987). The Lexile Test of Reading Comprehension. Durham, N.C.: MetaMetrics.

Stenner, A. J., Burdick, H., Sanford, E. E., and Burdick, D. S. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement*, *7*(3), 307–22.

Storch, S. A. and Whithurst, G. J. (2001). The role of family and home in the literacy development of children from low-income backgrounds. In P. Britto and J. Brooks-Gunn (Eds.), *The role of family literacy environments in promoting young children's emerging literacy skills* (pp. 39–52). San Francisco: Jossey-Bass.

Texas Education Agency. (n.d.). *TAKS–M Participation Requirements.* Retrieved July 17, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/taksm/TAKS_M_Participation_requirements.pdf

Texas Education Agency. (2004). *Texas Assessment of Knowledge and Skills: Information booklet, reading, grade 3 (revised).* Retrieved June 3, 2010, from http://ritter.tea.state.tx.us/student.assessment/taks/booklets/reading/g3e.pdf

Texas Education Agency. (2005a). *Scoring and reporting information.* Retrieved May 26, 2010, from http://ritter.tea.state.tx.us/student.assessment/resources/conferences/tac/2005/reporting_update.pdf

Texas Education Agency. (2005b). *The Lexile Framework for Reading.* Retrieved August 6, 2010, from http://ritter.tea.state.tx.us/curriculum/elar/lexile.html

Texas Education Agency. (2007). *TAKS–Alt participation requirements.* Retrieved July 17, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/taksalt/training/participation_requirements_july_2007.pdf

Texas Education Agency. (2008a). *2008 accountability manual.* Retrieved July 20, 2009, from http://ritter.tea.state.tx.us/perfreport/account/2008/manual/

Texas Education Agency. (2008b). *Student assessment division technical digest 2006–2007 (Chapter 18: Equating).* Retrieved July 31, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/techdig07/Chapters/Chapter18-Equating.pdf

Texas Education Agency. (2008c). *Texas Assessment of Knowledge and Skills (TAKS) resources.* Retrieved July 17, 2009, from http://www.tea.state.tx.us/index3.aspx?id = 948&menu_id3 = 793

Texas Education Agency. (2008d). *State of Texas education: At a glance.* Retrieved March 15, 2010, from http://loving1.tea.state.tx.us/lonestar/Menu_state09.aspx?group = summary

Texas Education Agency. (2009a). *Interpreting assessment reports: Spring 2009*. Retrieved May 25, 2009, from http://www.tea.state.tx.us/index3.aspx?id = 3282&menu_id = 793

Texas Education Agency. (2009b). *2008–09 Academic Excellence Indicator System.* Retrieved August 2, 2010, from http://ritter.tea.state.tx.us/perfreport/aeis/2009/index.html

Texas Education Agency. (2009c). *2009–2009 student enrollment reports.* Retrieved August 10, 2010, from http://ritter.tea.state.tx.us/adhocrpt/adste09.html

Texas Education Agency. (2009d). *Technical digest 2008–2009.* Retrieved October 11, 2010, from http://www.tea.state.tx.us/index3.aspx?id = 2147484418&menu_id = 793

Texas Education Agency. (2010a). *2009–2010 LPAC Decision-Making Process for the Texas Assessment Program*. Retrieved May 26, 2010, from http://ritter.tea.state.tx.us/student.assessment/ELL/lpac/LPAC2-Manual2010.pdf

Texas Education Agency. (2010b). *Frequently asked questions and answers about the Texas assessment system.* Retrieved October 11, 2010, from http://ritter.tea.state.tx.us/student.assessment/FAQ.pdf

Texas Education Agency. (2010c). *Raw score conversion tables.* Retrieved October 11, 2010, from http://www.tea.state.tx.us/index3.aspx?id = 3270&menu_id3 = 793

Texas Education Agency. (2010d). *Texas Essential Knowledge and Skills: February 2010 update*. Retrieved June 3, 2010, from http://ritter.tea.state.tx.us/rules/tac/chapter110/index.html

Texas Project First. (n.d.). *Texas Assessment of Knowledge and Skills (TAKS)*. Retrieved July 31, 2009, from http://www.texasprojectfirst.org/TAKS.html

Topping, K. J. (1987). Paired reading: A powerful technique for parent use. *The Reading Teacher, 40,* 608–14.

United States Department of Agriculture. (2008). *Eligibility manual for school meals: Federal policy for determining and verifying eligibility*. Retrieved December 10, 2010, from http://www.fns.usda.gov/cnd/Governance/notices/iegs/EligibilityManual.pdf

Vernon-Feagans, L., Hammer, C. S., Miccio, A., and Manlove, E. (2001). Early language and literacy skills in low-income African American and Hispanic children. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 192–210). New York: Guilford Press.

Weiss, D. J., and Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21,* 361–75.

White, S., and Clement, J. (2001). *Assessing the Lexile Framework: Results of a panel meeting* (NCES 2001-08). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Wilkins, C., Hartman, J., Howland, N., and Sharma, N. (2010). *How prepared are students for college-level reading? Applying a Lexile®-based approach*. (Issues & Answers Report, REL 2010–No. 094). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved December 10, 2010, from http://ies.ed.gov/ncee/edlabs

Woodward, M. (1901). Pittsburgh Carnegie Library. *Library Journal, 26*(11), 828.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA