The Effects of Summer Programs on K-12 Students'

Reading and Mathematics Achievement:

A Meta-Analysis

Chen Xie

East China Normal University

Amanda Neitzel

Johns Hopkins University

Alan Cheung

Chinese University of Hong Kong

Robert E. Slavin

Johns Hopkins University

The Effects of Voluntary Summer Programs on K-12 Students' Reading and Mathematics Achievement: A Meta-Analysis

Abstract

There has long been interest in using summertime to provide supplemental education to students who need it. But are summer programs effective? This review includes 19 randomized studies on the effects of summer intervention programs on reading and mathematics, based on rigorous quality criteria. In reading, there were two types of summer programs: summer school and summer book reading approaches. In mathematics, there was only summer school. The mean effect of summer school programs on reading achievement were positive (mean ES = +0.23), but there were no positive effects, on average, of summer book reading programs (ES=0.00). In mathematics, positive mean effects were also found for summer programs (ES=+0.17). However, the positive-appearing means for summer schools were not statistically significant in a meta-regression, and depended on just two reading and one math study with very large impacts. These successful interventions focused on well-defined objectives with intensive teaching.

Keywords: summer school, book reading, reading achievement, mathematics achievement, low-income students, low-achieving students.

Introduction

The summer months have long been of interest to educators and policy makers intent on improving student achievement. Traditionally, students throughout the world are on vacation in the summer. However, there has long been a concern that students out of school are forgetting what they learned during school time. Couldn't this time be used in a way to enhance the academic achievement of students, especially those performing below expectations? This question has taken on increased importance in recent years, as summer programs have been frequently proposed as means of improving outcomes for students whose learning progress has been interrupted by Covid-19 school closures (e.g., Sawchuk, 2020).

In the 1970's, Barbara Heyns (1978) studied summer learning, and her findings transformed thinking about summer learning ever since. What she found out was that from fall to spring, sixth and seventh graders from disadvantaged homes and those from middle class homes made similar gains in learning. Over the summer, however, middle class students maintained or built on their gains, while disadvantaged children declined in learning levels. Middle class students were engaging in more school-like activities over the summer, such as reading and organized activities, while disadvantaged students had little opportunity to use the skills they had gained in school. As a result, disadvantaged students arrived each fall performing less well than their middle class peers. Later, Alexander, Entwisle, & Olson (2007), in a nine-year longitudinal study of children in Baltimore, replicated and extended Heyns' findings, and showed how the cumulative and lasting effects of summer losses among disadvantaged students were undermining the gains they were making each school year. These and several other studies of summer loss led Allington & McGill-Franzen (2018) to conclude that summer loss is essentially all that matters in creating inequality.

Not surprisingly, these findings led many researchers and reformers to wonder whether summer losses among disadvantaged students could be mitigated by offering students instruction or other activities likely to help these students continue to grow academically during the summer. These may be mandatory programs, as in (mostly) high school credit recovery, in which students attend summer school to avoid a failing grade. However, our interest is in voluntary summer school programs, which students may be encouraged but not required to attend. Voluntary summer schools are designed to get students ahead in reading, math, or both during the summer months. Another common approach, summer book reading, does not involve direct teaching during the summer, but rather provides students with books and incentives to read and respond to them.

Extent of Summer Loss

Previous studies have estimated "summer slide," the effects of summer vacation on student achievement. Several of these have demonstrated that summer break widened the achievement gap across social lines, and low socioeconomic status children suffered greater loss than did middle class students during summer vacation, especially in reading achievement (Alexander, Entwisle, & Olson, 2001, 2007; Burkam, Ready, Lee, & LoGerfo, 2004; Downey, von Hippel, & Broh, 2004).

However, recent very large studies of scores on MAT tests, given in fall, winter, and spring, have called into question the whole idea of summer slide, and specifically the idea that disadvantaged students lose ground during the summer but make gains equal to those of non-disadvantaged students during the school year (Atteberry & McEachin, 2020; Kuhfeld, 2019).

Whether or not “summer slide” exists, summer clearly provides an opportunity to provide struggling students more time for instruction. If summer programs are targeted to disadvantaged or low-achieving students, and if they improve student achievement, then these programs may reduce learning gaps. Struggling students often lack vital foundational skills, so they may need several weeks of intensive instruction that delivers a supplemental curriculum (Kidron & Lindsay 2014; Lauer et al., 2006). School districts may offer struggling students extra help over the summer to enable them to start the new year with stronger basic skills. However, stakeholders have to know which summer programs can help them achieve their goals.

The present meta-analysis systematically reviews rigorous studies evaluating the effects of summer intervention programs to identify which programs are effective in accelerating student progress in reading and mathematics. This review also aims to explore the types of summer programs that are most effective for low-income and low-achieving students.

Definitions and Categories of Summer Programs

In this review, summer programs may solely involve academic instruction, or may blend academic activities with recreational and enrichment activities. However, the review only focuses on effects on reading and mathematics achievement. Two types of summer programs are identified from the studies included in this review: summer school (for reading and/or mathematics), and summer book reading.

Summer School

Summer school refers to programs that resemble the reading and/or mathematics instruction teachers might use when school is in session, although summer school class sizes may be smaller and small-group tutoring may be included. Many programs also provide enrichment activities, such as sports or trips (e.g., Borman & Dowling, 2006; Schacter & Jo, 2005; Zvoch & Stevens, 2013).

Summer Book Reading.

Summer book reading programs ask participating students to read books at home or at the library during summer vacation (e.g., Dynia, Piasta, Justice, & Columbus Metropolitan Library, 2015; Kim, 2006; Kim & Guryan, 2010). Some programs send participating parents messages suggesting reading and summer learning activities (e.g., Kraft & Monti-Nussbaum, 2017).

Literature Review

Several meta-analyses have been conducted since 2005 to review the effects of summer programs on reading and mathematics achievement (Kidron & Lindsay, 2014; Kim & Quinn, 2013; Lauer et al., 2006; Quinn et al., 2014). The reviews concluded that voluntary summer programs had positive effects on reading and mathematics achievement, although their overall reported effect sizes have ranged from +0.05 to +0.30.

Lauer et al. (2006) reviewed the effects of out-of-school-time programs for K-12 at-risk students. The search period was from 1985 to 2003. Fourteen independent effect sizes of summer schools on reading achievement met inclusion criteria. Their overall mean effect size was +0.05. In terms of mathematics achievement, 12 effect sizes of summer schools were included, and the mean effect size was +0.09.

Kim and Quinn (2013) included 41 studies of summer reading programs conducted from 1998 to 2011, involving children from kindergarten to Grade 8. Their mean effect size on total reading achievement was +0.09 for summer school (26 studies) and +0.12 for book reading programs (14 studies).

Quinn et al. (2014) reported a meta-analysis of mathematics outcomes of summer school. Overall, the 12 included studies averaged an effect size of +0.08 on mathematics achievement.

Kidron and Lindsay (2014) evaluated the effects of increased learning time programs on academic and nonacademic outcomes for K-12 students. They retrieved studies published from1998 to 2011 and found that six summer school studies met What Works Clearinghouse (2010) standards. Their mean effect size for literacy achievement was +0.16. They did not find any qualifying studies of mathematics summer schools.

Need for the Current Synthesis

One major contribution of the present meta-analysis is that it considerably updates the most recent reviews of research on summer programs for reading and mathematics. The three most recent reviews, by Kim & Quinn (2013, reading), Quinn, Lynch, & Kim (2014, mathematics), and Kidron & Lindsay (2014, reading and math) synthesized findings of summer school studies through 2011. Given the rapid pace of rigorous research over the past decade, 2011 is a long time ago. In the present synthesis, 73% of the included studies appeared in 2012 or later, and therefore could not have been included in the earlier reviews.

The inclusion criteria for the Kidron & Lindsay (2014) synthesis were based on What Works Clearinghouse (2010) standards, which are similar to those used in the present meta-analysis. Kidron & Lindsay (2014) identified six summer school reading programs (ES=+0.16) and no math programs. They did not examine summer book reading programs. In contrast, Kim & Quinn (2013) and Quinn et al. (2014) used much less stringent methodological inclusion standards. They reported effect sizes of +0.09 (k=26) for reading summer schools, +0.12 (k=14) for summer book reading, and +0.08 (k=12) for mathematics. Although these reviews did require that studies use randomized or matched experimental and control groups, the reading studies reported by Kim & Quinn (2014) included eight studies with no well-matched control group, eleven studies that were very small (n<60), and four matched studies in which the control groups consisted of students who chose not to attend (e.g., Haymon, 2009). All of these factors tend to inflate reported effect sizes (Cheung & Slavin, 2016).

The present synthesis uses strict inclusion criterion like those of Kidron & Lindsay (2014), but with an additional decade of rigorous studies, the present review is able to report findings for rigorous studies of reading summer school, home book reading, and math summer school. The present review therefore provides a stronger basis for evidence-based policy and practice.

Research Questions

1. What are the overall effects of included studies on reading and mathematics achievement?
2. How do effects of summer programs differ for a) low-income students and b) for low achievers?

Methods

The meta-analysis method employed by this article follows procedures from Lipsey and Wilson (2001) and Borenstein, Hedges, Higgins, and Rothstein (2009). The method comprises five key steps: (a) retrieve all potential studies; (b) screen studies by pre-set criteria; (c) code data and features of qualified studies; (d) compute effect sizes; and (e) implement statistical analyses.

Literature Search Procedure

We attempted to locate all studies published in English that examined the effectiveness of summer programs. First, we searched databases, including EBSCO (Academic Search Complete, Education Full Text, Education Source, ERIC, PsycINFO, and Teacher Reference Center), JSTOR, and ProQuest Dissertation & Theses Global. We used Boolean operators, parentheses, and wildcards to create the query. Specifics of our code appear in Figure A1 in the online appendix.

Previous meta-analyses and systematic reviews on summer programs were examined to see whether any studies were missed. We also followed up the references of all qualified studies to be sure we do not miss any qualifying studies.

Issues Unique to Studies of Summer Programs

There are several unusual features of studies of outcomes of summer school programs that are important to note. These influenced our inclusion criteria and statistical methods.

1. Because participation in summer school programs is voluntary, summer school studies suffer from serious problems of attrition, as students who initially volunteered for a summer program and were randomly assigned or matched into the experimental group either did not show up at all, or dropped out. Modern experimental research methods insist that all subjects initially assigned to experimental or control groups be retained in their original conditions, even if (for example) some experimental students do not actually receive the treatment. This is necessary because students who drop out are likely to be less highly motivated or conscientious, lower achieving, or have less supportive parents than those who do attend. If these students are dropped from the analysis, this tends to inflate the apparent program outcomes for the experimental group. For this reason, we only included intent-to-treat analyses (ITT), in which all subjects were kept in the analysis.
2. In most experimental studies in education, “business-as-usual” control groups are still receiving ordinary instruction in the subject at hand. However, in research on summer programs, the control group is likely to be receiving no academic interventions. For this reason, summer programs may show positive effects on fall testing, which reflect nothing more than recent involvement in

anything academic, or may indicate learning of content that control students have not been taught, but will be taught early in the regular school year. For these reasons, follow-up outcomes are of particular importance.

3. Ordinarily, in school-year research, the additional costs of an intervention are modest. The school was going to be open anyway. In contrast, summer school programs require teaching staff, and in order to get students to attend, they may provide sports, trips, and other non-academic services. Buildings must be kept open, cleaned, air conditioned, and kept secure. Lunch may be provided, and sometimes transportation. The point is, the costs of summer school can be very high, even in comparison to the most expensive of school-year interventions.

   Of course, none of these considerations apply to summer book reading programs, which tend to be relatively inexpensive.

Inclusion Criteria

The following inclusion criteria were used to identify possible qualifying studies. When studies did not show the necessary data and details, we sent emails to ask for authors' help.

1. The study evaluated the effect of a voluntary summer program on reading or/and mathematics achievement in K-12 settings.
2. The study had to take place in North America, Europe, Israel, Australia, or New Zealand. In practice, all included studies took place in the U.S. or U.K.
3. The study was published or released in English in 2000 or later.

4. At least 50% of participants in the study were general education students.
5. The study had to employ a control group design, either assigned at random (RCT) or in a matched quasi-experiment (QED), with assignments to conditions specified in advance.
6. The study had to include a minimum of 30 students and/or teachers in each condition.
7. Pre-intervention mean differences could not be greater than 0.25 standard deviation for literacy or mathematics measures.
8. From pretest to posttest, attrition differences between condition groups had to be no more than 15%.
9. The study had to provide data to allow for an intent-to-treat (ITT) analysis, meaning that all participants were identified in advance and remained in the analysis whether or not they actually participated.
10. Study duration had to be at least 4 weeks during the summer term.
11. The dependent measures had to be quantitative, usually standardized assessments. Experimenter-made measures were not accepted.
12. The study had to report achievement data sufficient to calculate effect sizes.

Coding

Each study was independently coded by at least two researchers. We coded the included studies independently. When facing disagreements, all authors discussed together and came to a final agreement.

The fully coded data are available on GitHub (Neitzel, Xie, Cheung, & Slavin, 2020).

Effect Size Computation

An effect size refers to the standardized difference between experimental and control group posttests after adjustment for pretests and other covariates. The effect size statistic used in this review is based on Cohen's *d* (Cohen, 1987). If a study did not report adjusted effect sizes, we subtracted effect sizes for pretest from effect sizes for posttest (a difference in differences). If a study reported two or more outcome variables in reading or math, we computed the mean effect sizes.

Some studies had several posttests at different time points in order to observe near-term and long-term effects (e.g., Augustine et al., 2016; Kraft & Monti-Nussbaum, 2017; Schacter & Jo, 2005). In this case, we used the measure taken at the end of the program. If participation in summer programs was offered over consecutive summers (e.g., Augustine et al., 2016; Borman & Dowling, 2006), results are considered those from the final year, although other findings are reported in the text. Any later measures were considered follow-up, and these effect sizes are reported separately.

Statistical Analyses

In meta-analysis models, studies were weighted to give more weight to studies with the greatest precision (Hedges, Tipton, & Johnson, 2010). In practice, this primarily involves weighting for sample size. Weights for each study were calculated according to the following formula:

$$W_j = \frac{1}{k_j(\bar{v}_j + \tau^2)}$$

where W_j is the weight for study j, k_j is the number of findings in study j, $\bar{v}_j$ is the average finding-level variance for study j, and τ^2 is the between-study variance in the study-average effect sizes (Hedges et al., 2010). We used a multivariate meta-regression model with robust variance estimation (RVE) to conduct the meta-analysis. We estimated two meta-regression models. First, we estimated a null model to produce the average effect size without adjusting for any covariates. Second, we estimated a meta-regression model with the identified moderators of interest and covariates. The packages metafor (Viechtbauer, 2010) and clubSandwich (Pustejovsky, 2020) were used to estimate all random-effects models with RVE in the R statistical software (R Core Team, 2020).

Results

Characteristics of Included Studies

Over 4,000 studies were screened (see Figure 1). Nineteen studies met the inclusion criteria, including a total sample size of 21,460 students (see Tables 1 to 3). Three studies were intended to improve both reading achievement and mathematics achievement, fourteen were intended to improve only reading, and two were intended to improve only mathematics. Fifteen studies took place in elementary schools, two in secondary schools, and two studies included participants from both levels. All of the studies used student-level random assignment, with intent-to-treat procedures. Five of the studies were published from 2000 to 2011, and could therefore have been included in the most recent prior reviews (Kidron & Lindsay, 2014; Kim & Quinn, 2013;

Quinn et al., 2014). Fourteen were published from 2012 to 2020.

Results by Programs and Categories

As noted earlier, included reading studies were classified into two categories: summer school reading and summer book reading. We briefly describe each program in the following sections.

Summer school reading. The category of summer school reading consisted of six qualifying studies of five programs (Table 1).

Summer Reading Camp. Schacter and Jo (2005) evaluated the effect of a reading summer day-camp intervention for first graders. The program was implemented for 7 weeks, five days a week, for nine hours a day. Two hours a day were devoted to reading, and the rest of the time was for summer camp activities (e.g., arts and crafts, swimming). The reading time included whole-class phonics instruction and Open Court decodable books taught to students in groups of 15. This was followed by small-group reading and writing activities. Teachers taught direct, explicit, and systematic decoding, comprehension, vocabulary, and writing skills. On Gates measures at the end of the program, the effect size was +1.16. On December Gates follow-up measures, the mean effect size was +0.92, but on May Stanford-9 measures, the effect size diminished to +0.16.

Teach Baltimore was an ambitious project intended to provide disadvantaged students initially completing kindergarten and first grade with high-quality, reading-focused instruction for three consecutive summers, to counteract “summer slide.” Volunteer college students were randomly assigned to receive seven weeks of summer school for each of three years, or to serve in a control group and receive no special services over the summer. The daily schedule included 3 hours of reading and writing instruction, as well as math, science, physical activity, and breakfast and lunch. The reading instruction used *Open Court Phonics*, as well as read-alouds, silent reading, literature, and creative writing activities. The college student instructors received extensive training, as well as mentoring from certified teachers.

An evaluation of the first year of Teach Baltimore was reported by Borman, Rachuba, Hewes, Boulay, & Kaplan (2001), and a report on the full 3-year project was reported by Borman & Dowling (2006). In the first year, all kindergarten and first grade students in 10 high-poverty schools were invited to participate. Students were pre-tested on CTBS reading tests. Parallel forms of the same tests were given in the fall, as posttests. Overall intent-to-treat treatment effects were ES=+0.01, or essentially zero, by fall of the third year.

In the first two summers, students in Teach Baltimore gained (non-significantly) more than did control students, but on the following spring tests, these gains had diminished to near zero. In other words, while the summer school did show a small benefit each summer, there was no evidence of a *cumulative* effect.

Academically Intense Literacy Programs. Zvoch and Stevens (2013) examined a district-sponsored academically intensive literacy program for students completing kindergarten and first grade. Students received 3.5 hours of literacy instruction every day for five weeks. They received teacher-directed instruction in the critical beginning reading skills of phonemic awareness, alphabetic understanding, and fluency/automaticity. Literacy skills were then modeled and practiced, primarily in small groups of 3 to 5 students, much like small-group tutoring, with students grouped based on skill level. Effect sizes on DIBELS Nonsense Word (phonics) and Fluency tests averaged +0.60 ($p < .05$) in kindergarten and +0.78 ($p < .05$) in first grade.

BELL. Somers et al. (2015) examined a widespread program called BELL, a five-week academic summer program. Low-achieving middle school students received instruction in groups of 20 for 12 hours a week in reading and 12 hours in math, for a total of 60 hours in each subject, plus two hours a week in each subject on projects. A teacher and an aide worked with each class, using standard textbooks. Students also received 8 hours a week of enrichment activities. In a randomized experiment, reading effects averaged near zero (ES=+0.01) on GRADE reading tests.

Gorard, Siddiqui, and See (2015) evaluated BELL in summer schools in England. The four-week program used textbooks and regular literacy and numeracy lessons taught by trained teachers and teaching assistants. Each afternoon, students participated in a variety of sports and enrichment activities. In a randomized evaluation in Years 5-6 (U.S. grades 4-5), the mean reading effect size was +0.17 (n.s.) on Progress in Reading tests.

The mean effect size for reading outcomes across the two studies of BELL was +0.03 (n.s.).

Voluntary District-Run Programs. Augustine et al. (2016) conducted a study of district-led, voluntary summer programs in five urban school districts: Boston, Dallas, Jacksonville, Pittsburgh, and Rochester. Districts incorporated a common set of five elements: voluntary, full-day programming combining academics and enrichment; at least three hours of instruction (language arts and mathematics) per day; small class sizes of no more than 15 students per adult; and free participation, transportation, and meals. Students randomly assigned to the experimental group were invited to participate in summer school for two summers. Reading outcomes on GRADE reading tests averaged an effect size of 0.00 at the end of the first summer, and +0.04 (n.s.) at the end of the second summer.

Effect sizes across all summer reading programs averaged +0.23. This was not significant in the meta-regression, due mainly to small numbers of studies.

Summer book reading. The category of summer book reading consisted of eleven qualifying studies of five programs (Table 2). It is worth noting that many of these studies evaluated minor program variations, but for the purposes of this review, only overall means were analyzed.

READS. Seven studies evaluated forms of Reading Enhances Achievement During Summer (READS), created by James Kim. There were variations among the studies in terms of the interventions, but outcomes were very similar. In all studies, students were given books and asked to fill out worksheets on them. Postcard reminders were sent to parents and students. Participating students were encouraged by their teachers to practice oral reading at home with a family member. The books were within students' independent reading levels and matched to their reading preferences. Both treatment and control group students attended special reading lessons during the last month of school, on comprehension strategies and oral fluency.

Across the seven studies of READS, the mean effect size was +0.00 (n.s.).

Books and Reminders. Wilkins et al. (2012), Maxwell et al. (2014), and Dynia et al. (2015) used similar summer reading programs to READS. Wilkins et al. (2012) provided participants with a single shipment of eight books matched to their reading levels and interest, and followed up by sending them reminder cards each week for six weeks. The effect size was +0.02. Maxwell et al. gave four book packs to participating students in England and invited them to attend two summer events that included one-to-one reading activities. This study had a mean effect size of +0.13. Dynia et al. (2015) instructed participants to read books of their own choosing. Participants visited the library throughout eight weeks to pick up incentives for reading. The effect size of this study was -0.08.

Kraft and Monti-Nussbaum (2017) explored the potential of a summer text-messaging program to encourage reading. Parents received a total of 18 text messages emphasizing the importance of reading and the role of parents in encouraging reading at home during the summer months. The texts also provided information on resources and ideas for summer learning activities. The effect size of this study was +0.05.

The overall mean effect size for all book reading programs was 0.00 (n.s.).

Summer school mathematics. The category of summer school mathematics consisted of five qualifying studies of four programs (Table 3).

Elevate Math. The Elevate Math summer program was evaluated by Snipes, Huang, Jaquet, and Finkelstein (2015). The program was designed for seventh graders who did not qualify for eighth grade algebra, but came close to qualifying. Elevate

Math included a Common Core-based curriculum that provided four hours each day of blended learning. Classroom instruction was led by a teacher and an aide. One hour was spent on Khan Academy (a free online learning system with optional video lessons). Teachers received 40 hours of professional development on the content and pedagogy, and in-class coaching. Students received a total of 75 hours of instruction over four weeks.

Snipes et al. (2015) carried out an evaluation of Elevate Math in six suburban California school districts in Silicon Valley. Students were randomly assigned to receive the four-week session or to serve in a control group. Students were pre- and post-tested on the Mathematics Diagnostic Testing Project (MDTP) algebra readiness test. On posttests, controlling for students' sixth grade California Standards Test scores, outcomes significantly favored the experimental group (ES=+0.64, $p<.05$). In addition, 29% of experimental and only 12% of control students met requirements for algebra readiness (ES=+0.53, $p<.05$).

Tenmarks. Lynch and Kim (2017) evaluated an online summer mathematics program called Tenmarks, in comparison to an untreated control group. Key components of Tenmarks are curriculum materials that adjusted content to children's individual skill levels, embedded text, video "hints" that students can click on for assistance, and digital games that children can play as rewards for completing worksheets. Students receive weekly text messages encouraging them to log into Tenmarks. A randomized evaluation of Tenmarks found a mean effect size of -0.01.

Three studies evaluated mathematics outcomes of summer school programs that included both reading and mathematics. The reading outcomes of these studies have been described previously.

BELL. Two of the summer school studies with reading and mathematics outcomes evaluated the BELL program. A U.S. study of BELL (Somers et al., 2015) involved low-achieving students in grades 5-7. As noted previously, BELL students received 60 hours of mathematics instruction, plus 70 hours of reading and 40 hours of enrichment, over a 5-week period. Mathematics was taught by a teacher and an aide in class groups of 20 students, using a Common Core-aligned textbook, plus two hours a week on math projects. A randomized experiment found small and non-significant effects on GMADE math tests (ES=+0.07, n.s.).

Gorard, Siddiqui, & See (2015) carried out an evaluation of BELL in England. Low-achieving students in Years 5-6 (equivalent to U.S. grades 4-5) were randomly assigned to attend a 4-week summer school session, or to serve in a control group. Students in BELL received daily class instruction in math as well as reading, with enrichment activities. On the Progress in Maths test, the mean effect size averaged 0.00. The mean effect size across both studies of BELL for mathematics outcomes was +0.06 (n.s.).

Voluntary District-Run Programs. A study by Augustine et al. (2016), described previously, evaluated summer school programs involving mathematics and language arts in five U.S. urban districts. Third grade volunteers were randomly assigned to participate in summer school for two years. The students who received summer school received at least three hours of instruction per day for at least 5 weeks, in class sizes of no more than 15. The evaluation found a small but significant positive effect on GMADE Mathematics scores at the end of the first year (ES=+0.08, $p<.05$). However, at the end of the second year, there were no significant cumulative differences (ES=+0.03, n.s.).

The mean effect size for summer mathematics programs was +0.17. This was not statistically significant in the meta-regression, due mainly to small numbers of studies.

Overall Effects

The results for the null model and full meta-regression models are shown in Table 4. The meta-regression model controlled for program category, study poverty level and baseline achievement level. The overall effect sizes of summer programs on reading and mathematics achievement were +0.23 and +0.17, respectively (see Table 4). The mean effect size was 0.00 for book reading programs. As noted earlier, none of these mean effect sizes was statistically significant in the meta-regression.

Moderator Analyses

We analyzed two key study features for all summer programs: (a) baseline achievement; and (b) student poverty. The results are shown in Table 5.

Effects for low-income students. In this review, low-income students refer to studies in which 75% or more of the participating students were eligible for free or reduced-price lunches in the U.S., or at least 25% eligible for free school meals in the U. K. (These criteria represent the 75^{th} percentile of free lunch in each country.) Student poverty was not a significant moderator for summer school reading programs, summer book programs, or summer school math programs.

Effects for low achievers. This review defines low achievers as students whose academic achievement was below grade level before summer programs began. The other possible groups were for average and above achievers, and mixed achievers (all students). Student baseline achievement was not a significant moderator for summer school reading programs, summer book reading programs, or summer school math programs.

Program Cost

Seven included studies provided details on program cost. These costs need to be considered carefully, because they included different program components (see Table A6 in the online appendix). The cost of summer school programs is consistently higher than that of summer book reading programs. Translating reported costs to 2020 U.S. dollars, three studies of summer school programs reported that their costs were, respectively, $560 per student (Snipes et al., 2015), $1,507 per student (Augustine et al., 2016), and $1,995 per student (Gorard et al., 2015). The cost of summer school programs mainly included the costs of academic classroom staff salaries, administration, enrichment activities, transportation, food, curriculum development, and training.

The reported costs of four summer book reading programs were, respectively, $31 per student (Wilkins et al., 2012), $79 per student (Kim, 2006), $274 per student (Maxwell et al., 2014), and $364 per student (Guryan et al., 2014). These costs mainly consisted of the costs of books, postage, and labor.

Discussion

Interpretation and Comparison of Overall Effects

The present meta-analysis found that included studies of summer programs produced an overall effect size of +0.23 for six studies of summer school reading, 0.00 for 11 summer book reading studies, and +0.17 for five studies of summer school mathematics. None of these was statistically significant in the meta-regression. Our effect size estimates are larger than those in previous reviews, but our estimate for summer book programs (ES=0.00) was much lower than the mean (ES=+0.12) reported by Kim & Quinn (2013). However, that review included many studies that did not meet the more selective inclusion criteria of the present synthesis.

The positive outcomes of summer school programs for reading and mathematics seem to offer promise, but the means, in this case, were not significant in the meta-regression, and in any case they are misleading. In reading, there were two programs with extraordinarily large effect sizes, and all others had non-significant outcomes, mostly with effect sizes near zero. The two studies with large impacts were the studies of Summer Reading Camp (Schacter & Jo, 2005) and an academically intensive literacy program (Zvoch & Stevens, 2013). The large effect sizes of these programs (+1.16 and +0.69, respectively) greatly affected the category mean. In summer school mathematics, Elevate Math (Snipes et al., 2015) had a positive effect size of +0.64, but no other mathematics approach had positive outcomes. To understand “what works” in summer programming, a mean is not appropriate in these cases. Instead, we need to understand these three programs, and contrast them with all others.

In summer school reading, the two very successful programs were remarkably similar to each other. Summer Reading Camp (Schachter & Jo, 2005) was focused on first graders, and the academically intensive reading program (Zvoch & Stevens, 2013) involved K-1. Both had a very strong focus on phonics, as well as comprehension and writing. Both started reading lessons with whole-class instruction, followed by one-to-small group tutoring.

It is important to note that a third program, Teach Baltimore (Borman & Dowling, 2006), also provided a summer school with a strong emphasis on phonics to students initially completing kindergarten and first grade. Students randomly assigned to the summer school condition were invited to participate in summer for as much as three years. The three-year outcomes of Teach Baltimore were near zero in intent-to-treat analyses, due in large part to a high rate of no-shows and poor attenders, which got worse over the years. Also, Teach Baltimore used college students, not teachers, to deliver instruction. Analyses using Complier Average Causal Analyses (CACE), in which students in the experimental group who actually participated were compared to students in the control group who were statistically determined to be most likely to have participated, did show significant positive outcomes (ES=+0.30, $p<.05$).

All other summer reading programs failed to show significantly positive outcomes. Two studies evaluated BELL, whose reading program emphasizes traditional teaching using traditional textbooks with middle school students. The weighted mean effect size was +0.03. Augustine evaluated traditional whole-class teaching in summer schools in five U.S. cities, and the average effect size was zero.

This evidence suggests that using summer time to teach small classes using traditional teaching methods and materials is not likely to produce large reading impacts.

Among five studies of summer school mathematics, the outcomes were as uneven as those for reading. The overall positive-appearing effect size for mathematics was almost entirely due to an effect size of +0.64 for one study, the Snipes et al. (2015) evaluation of Elevate Math. This intervention identified seventh graders who barely missed the test score needed for entry to eighth grade algebra on their spring state tests. The students then received 75 hours of daily instruction in the skills assessed on an algebra diagnostic test, and then they were found to have improved quite a lot on that test. Control students who also barely failed the test received no instruction. It would have been surprising if the outcomes had been any different. Elevate Math may be a good solution for the specific problem faced by these students and schools qualifying borderline seventh-graders for eighth grade algebra, but may not be applicable to most of mathematics instruction.

Other summer school mathematics outcomes were very small. Tenmarks, a computer-assisted instruction approach used over the summer, made no difference in outcomes (Lynch & Kim, 2017). Similar outcomes were seen in studies of BELL (two studies, with outcomes of 0.07 and 0.00), and Augustine et al. (2016) (ES=+0.03 over two summers of participation).

One important conclusion from the findings of this review is that there is nothing magic about summer school. Despite the concern and debate about “summer slide,” and the possibility that summers out of school are particularly damaging to the learning of disadvantaged students, there is no evidence that the fact of receiving 4 to 7 weeks of summer school, or reading assignments all summer, are successfully combatting summer losses. Instead, when summer programs work, it is because they provide intensive instruction on specific objectives (i.e., phonics for K-1 students or pre-algebra for seventh graders who just missed qualifying for eighth grade algebra), believed by school leaders to be major hurdles to student progress. All of the approaches that did appear to have positive effects could have been implemented during the regular school year with (perhaps) equal impacts. What summer provides is additional time. Compared to a 40-week ordinary school year, 4 to 7 additional weeks is not a large increase. What matters, clearly, is what happens over that time period. If it is essentially more of the same kind of instruction students received previously, then it is not likely to make much of a difference. If it is very focused on essential skills, using small-group instructional methods similar to small group tutoring (Authors, in press), then summer school is more likely to be effective.

A key question in the study of summer school is how long any benefits of summer school may last. The studies of Summer Reading Camp and of Teach Baltimore found that gains students made from spring to fall (in comparison to control groups) diminished substantially by the following spring.

In sum, the research on summer programs of all kinds does not justify enthusiasm about summer as a uniquely appropriate time for intervention. Comparable interventions during the school year, such as small-group tutoring (Baye, Lake, Inns, & Slavin, 2019; Neitzel, Lake, Pellegrini, & Slavin, 2020) may be more effective, less costly, and more integrated with the core instructional functions of the school.

Implications for Future Research and Practice

While the numbers of studies in each category do not allow for definitive conclusions, the findings of the current review may have important implications for future research and practice related to the development, evaluation, and scale-up of summer programs.

1. When programs work, it is probably because of the nature of the programs, not the fact that they took place in the summer.

2. Future evaluation studies must pay much more attention to the quality of the research design and implementation. There were some common problems in the evaluation studies not included in the present review: no control group, no pretest, insufficient sample size, baseline inequality, including control groups that consisted of students who were invited but did not join summer school, and large attrition differences between conditions.

3. The present review found that summer book reading programs did not produce positive outcomes, on average. Most evaluation studies reported no significant results, and even the significant effect sizes were very small. These programs were much less expensive than were summer schools, but their costs were not inconsiderable. School leaders interested in such approaches should focus on the most effective, but moderate their expectations.

References

References marked with an asterisk indicate studies included in this review.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational evaluation and policy analysis, 23* (2), 171-191. doi:10.3102/01623737023002171

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180. doi:10.1177/000312240707200202

Allington, R. L., & McGill-Franzen, A. (Eds.). (2018). *Summer reading: Closing the rich/poor reading achievement gap*. New York, NY: Teachers College Press.

Atteberry, A., & McEachin, A. (2020). School's out: The role of summers in understanding achievement disparities. *American Educational Research Journal* https://doi.org/10.3102/0002831220937285

*Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., et al. (2016). *Learning from summer: Effects of voluntary summer learning programs on low-income urban youth*. Retrieved from Rand website: https://www.rand.org/pubs/research_reports/RR1557.html

Baye, A., Lake, C., Inns, A., & Slavin, R. (2019). Effective reading programs for secondary students. Reading Research Quarterly, 54 (2), 133-166.

Borenstein, M., Hedges.L.V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.

*Borman, G. D., Rachuba, L., Hewes, G., Boulay, M., & Kaplan, J. (2001, Spring).

Can a summer intervention program using trained volunteers narrow the achievement gap? First-year results from a multi-year study. *ERS Spectrum,* 19-30.

*Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis, 28,* 25-48. doi:10.3102/01623737028001025

Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004). Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education, 77*(1), 1-31. doi:10.1177/003804070407700101

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences*. Hillside, NJ:Lawrence Erlbaum Associates.

Cheung, A., & Slavin, R. (2016). How methodological features affect effect sizes in education. Educational Researcher, 45 (5), 283-292.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*(5), 613-635. doi:10.1177/000312240406900501

*Dynia, J. M., Piasta, S. B., Justice, L. M., & Columbus Metropolitan Library. (2015). Impact of library-based summer reading clubs on primary-grade children's literacy activities and achievement. *The Library Quarterly, 85*(4), 386-405.

doi:10.1086/682733

*Gorard, S., Siddiqui, N., & See, B. H. (2015). How effective is a summer school for catch-up attainment in English and maths? *International Journal of Educational Research, 73*, 1-11. doi:10.1016/j.ijer.2015.07.003

*Guryan, J., Kim, J. S., & Park, K. H. (2016). Motivation and incentives in education: Evidence from a summer reading experiment. *Economics of Education Review, 55*, 1-20. doi:10.1016/j.econedurev.2016.08.002

*Guryan, J., Kim, J. S., & Quinn, D. M. (2014). *Does reading during the summer build reading skills? Evidence from a randomized experiment in 463 classrooms* (No. w20689). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w20689

Haymon, G. W. (2009). *The impact of summer school on student academic achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3354741)

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

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

Kidron, Y., & Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta‐analytic review* (REL 2014-015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia.

*Kim, J. S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis, 28*, 335-355. doi:10.3102/01623737028004

*Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology, 99*, 505-515. doi: 10.1037/0022-0663.99.3.505

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

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research, 83*(3), 386-431. doi:10.3102/0034654313483906

*Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in Grades 3 to 5: An experimental study. *Scientific Studies of Reading, 12*, 1-23. doi: 10.1080/10888430701746849

*Kraft, M. A., & Monti-Nussbaum, M. (2017). Can schools enable parents to prevent summer learning loss? A text-messaging field experiment to promote literacy skills. *The ANNALS of the American Academy of Political and Social Science, 674*(1), 85-112. doi:10.1177/0002716217732009

Kuhfeld, M (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan*, *101* (11), 25-29.

Lauer, P. A., Akia, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275-313. doi:10.3102/00346543076002275

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London, UK: Sage.

*Lynch, K., & Kim, J. S. (2017). Effects of a summer mathematics intervention for low-income children: A randomized experiment. *Educational Evaluation and Policy Analysis, 39*(1), 31–53. https://doi.org/10.3102/0162373716662339

*Maxwell, B., Connolly, P., Demack, S., O'Hare, L., Stevens, A., Clague, L., et al. (2014). *Summer Active Reading Programme: Evaluation report and executive summary*. Retrieved from ERIC website: https://eric.ed.gov/?id=ED581463

Neitzel, A., Lake, C., Pellegrini, M., & Slavin, R. (in press). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly.*

Neitzel, A.J., Xie, C., Cheung, A., & Slavin, R.E. (2020). Data archive for *The effects of voluntary summer programs on K-12 students' reading and mathematics achievement: Systematic review and meta-analysis*. Baltimore, MD: Center for Research and Reform in Education (CRRE), Johns Hopkins University. Retrieved from https://github.com/aj-neitzel/The-Effects-of-Voluntary-

Summer-Programs-on-K-12-Students-Reading-and-Mathematics-Achievement.

Pustejovsky, J. (2020). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (Version R package version 0.4.1) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

Quinn, D. M., Lynch, D., & Kim, J. S. (2014, March). *Replicating the moderating role of income status on summer school effects across subject areas: A meta-analysis*. Paper presented at the Society for Research on Educational Effectiveness Spring 2014 Conference, Washington, D.C.

R Core Team. (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sawchuk, S. (2020, August 26). Overcoming Covid-19 learning loss. *Education Week*, *40* (2), 6.

*Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, *28*, 158-169. doi:10.1111/j.1467-9817.2005.00260.x

*Snipes, J., Huang, C-W., Jaquet, K., & Finkelstein, N. (2015). *The effects of the Elevate Math summer program on math achievement and algebra readiness* (REL 2015-096). Retrieved from Eric website: https://eric.ed.gov/?id=ED558157

*Somers, M. A., Welbeck, R., Grossman, J. B., & Gooden, S. (2015). *An analysis of* *the effects of an academic summer program for middle school students*. Retrieved from ERIC website: https://files.eric.ed.gov/fulltext/ED558507.pdf

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the **metafor** package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

What Works Clearinghouse. (2010). *The What Works Clearinghouse procedures and standards handbook* (Version 2.1). Washington, DC: Author.

*White, T. G., Kim, J. S., Kingston, H. C., & Foster, L. (2014). Replicating the effects of a teacher- scaffolded voluntary summer reading program: The role of poverty. *Reading Research Quarterly, 49*(1), 5-30. doi:10.1002/rrq.62

*Wilkins, C., Gersten, R., Decker, L., Grunden, L., Brasiel, S., Brunnert, K., et al. (2012). *Does a summer reading program based on Lexiles affect reading comprehension?* (NCEE 2012-4006). Retrieved from ERIC website: https://files.eric.ed.gov/fulltext/ED530029.pdf

*Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized field trial. *Early Childhood Research Quarterly, 28*(1), 24-32. doi:10.1016/j.ecresq.2012.05.002

Table 1
Summer School Reading Programs

Study	Design	Duration	N	Grades	Sample Characteristics	Posttest	Effect sizes
Summer Reading Camp						Category Mean: +0.23	
Schacter & Jo (2005)	SR	7 weeks	3 schools, 118 students (54E, 64C)	1	Los Angeles. 63%AA, 37% H, 100% FRL.	Gates Total	+1.16
						Follow-up: 3m, +0.92; 6m, +0.16	
Teach Baltimore							
Borman & Dowling (2006)	SR	7 weeks in each of 3 summers	10 schools, 686 students, (438E, 248C)	K-1	Baltimore. 97% AA, 85% FRL.	CTBS/4 Total Reading	+0.01
District-sponsored academically intensive literacy program							
Zvoch & Stevens (2013)	SR	5 weeks	93 students (47E, 46C)	K-1	Small city in the Pacific Northwest. 15% EL, 63% FRL.	DIBELS Nonsense Word and Fluency	+0.69
Building Educated Leaders for Life (BELL)							
Somers et al. (2015)	SR	5 weeks	5 schools 919 students (585E, 334C)	5-7	Low-performing students, 45% AA, 34% H, 9% EL, 89% FRL.	GRADE Total	+0.01

Gorard et al. (2015)	SR	4 weeks	43 schools, 310 students (167E, 143C)	Years 5-6	Low-performing students in the U.K.	Progress in English	+0.17
Voluntary district-run programs							
Augustine et al. (2016)	SR	5+ weeks for 2 consecutive summers	5 school districts, 4484 students (2544E, 1940C)	3-4	Boston, Dallas, Jacksonville, Pittsburgh, Rochester. 47% AA, 40% H, 31% EL, 87% FRL.	GRADE, (effect of 2 summers)	+0.04
						Follow-up: 6m, +0.01; 18 m, +0.03	

Abbreviations for Tables 1-3: FRL: Free and Reduced Price Lunch (U.S.); FSM: Free School Meals (U.K.); AA: African-American; H: Hispanic; A: Asian-American; W: White; EL: English learner; CTBS: Comprehensive Test of Basic Skills; DIBELS: Dynamic Indicators of Basic Early Literacy System; GMADE: Group Math and Diagnostic Evaluation; GMRT: The Gates-MacGinitie Reading Test; GRADE: Group Reading; Assessment and Diagnostic Evaluation; ITBS: Iowa Test of Basic Skills; MDTP: Mathematics Diagnostic Testing Project; STAR: Standardized Test for the Assessment of Reading; SR: Student Randomized; STEP: State Test of Educational Progress; *: $p < .05$

Note: More detailed versions of all tables appear in the online appendix

Table 2
Summer Book Reading Programs

Study	Design	Duration	N	Grades	Sample Characteristics	Posttest	Effect sizes
Reading Enhances Achievement During Summer (READS)							Category Mean: 0.00 Program Mean: 0.00
Kim (2006)	SR	NA	10 schools, 486 students (252E, 234C)	4	19% AA, 26% H, 18% A, 56% EL, 40% FRL.	ITBS, DIBELS	+0.02
Kim (2007)	SR	NA	1 school 279 students (138E, 141C)	1-5	Suburban district. 42% W, 23% EL, 23% FRL.	Stanford 10 Reading: Grade 1	+0.04
Kim & White (2008)	SR	NA	400 students (93 Book, 100 Book+ ORS, 100 Book +ORC, 107C)	3-5	Suburban Mid-Atlantic 25% AA, 29% H, 29% EL, 38% FRL.	ITBS, DIBELS	+0.03
Kim & Guryan (2010)	SR	NA	317 students	4	California. 99% H, 73% EL, 96% FRL.	GMRT Total	+0.03
White et al. (2014)	SR	NA	19 schools, 1188 students	3	Urban North Carolina. 51% AA, 30% H, 72% FRL.	ITBS Reading Comprehension	-0.03

Guryan et al. (2014)	SR	NA	59 schools, 5319 students (2659E, 2660C)	2-3	Below-average readers in North Carolina school districts. 38% AA, 22% H, 17% EL, 77% FRL.	ITBS	+0.01
Guryan et al. (2016)	SR	NA	9 schools, 397 students	3-4	Northeastern U.S. 43% AA, 11% H, 50% FRL.	GMRT Total	-0.10
Providing books and reminder postcards							
Wilkins et al. (2012)	SR	NA	112 schools, 1571 students (791E, 780C)	3	Low-performing students in Texas. 20% AA, 68% H 100% FRL.	Scholastic Reading Inventory	+0.02
Summer Active Reading							
Maxwell et al. (2014)	SR	NA	48 schools, 182 students (93E, 89C)	6	Students at risk in the north of England. 34% FSM.	New Group Reading Test	+0.13
Summer reading clubs							
Dynia et al. (2015)	SR	36 days, over 8 weeks	76 students, (41E, 35C)	2-3	Students at risk in Columbus, OH. 29% FRL.	Test of Word Reading Efficiency, Gates	-0.08
Text-messaging							
Kraft & Monti-Nussbaum (2017)	SR	8 weeks	2 schools, 224 students (114E, 110C)	K-3	Struggling readers in Rhode Island. 12% AA, 32% H, 53% FRL.	STAR, STEP	+0.07
						Follow-up: 2m, +0.11; 5m, +0.11; 9m, +0.14	

Table 3

Summer School Mathematics Programs

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes
Elevate Math							Category Mean: +0.17
Snipes et al. (2015)	SR	4 weeks	8 schools 349 students (165E, 184C)	7	Six districts in California's Silicon Valley. High basic level or low proficient level. 34% A, 52% H, 57% FRL.	MDTP Algebra Readiness	+0.64
Tenmarks							
Lynch & Kim (2017)	SR	10 weeks	4 schools, 196 students	3-9	Urban Northeast. 38% AA,31% H, 59% FRL.	NAEP items, district curriculum-based assessment	-0.01
Building Educated Leaders for Life (BELL)							
Somers et al. (2015)	SR	5 weeks	5 schools 919 students (585E, 334C)	5-7	Low-performing students in the West and the Southeast. 45% AA, 34% H, 9% EL, 89% FRL.	GMADE Total	+0.07
Gorard et al. (2015)	SR	4 weeks	43 schools, 306 students (164E, 142C)	5-6	Low-performing students in the U.K.	Progress in Maths	0.00
Voluntary district-run programs							
Augustine et al. (2016)	SR	5+ weeks for 2 summers	5 school districts, 4505 students (2553E, 1952C)	3-4	Boston, Dallas, Jacksonville, Pittsburgh, Rochester. 47% AA, 40% H, 31% EL, 87% FRL.	GMADE (effect of 2 consecutive years of invitation to summer school)	+0.03

Table 4

Meta-regression results

Coefficient	β	SE	t	df	p
Null Model					
Intercept	0.10	0.05	2.03	19.79	0.056
Meta-Regression					
Summer school reading	0.23	0.19	1.21	3.61	0.298
Summer book reading	0.00	0.02	-0.20	9.47	0.847
Summer school mathematics	0.17	0.14	1.24	2.64	0.314
Summer school reading X high poverty	-0.08	0.18	-0.45	1.06	0.726
Summer book reading X high poverty	0.07	0.05	1.39	2.57	0.273
Summer school mathematics X high poverty	-0.23	0.01	-22.69	1.58	*n.s.*
Summer school reading X high achievers	-0.29	0.54	-0.53	2.04	0.649
Summer book reading X high achievers	0.10	0.02	5.48	1.46	0.060
Summer school mathematics X high achievers	-0.05	0.32	-0.16	2.04	0.888
Summer school reading X low achievers	-0.31	0.54	-0.58	2.05	0.621
Summer book reading X low achievers	0.13	0.04	2.92	1.39	0.150
Summer school mathematics X low achievers	-0.14	0.33	-0.43	2.05	0.706

Note. SE=standard error; df=degrees of freedom; Model was estimated without an intercept, with interaction terms for the differential effects of socioeconomic status and baseline achievement for each category.

Table 5

Overall effect sizes

	Level	k	n	ES	SE	t	df	p
Summer school reading (all)		6	12	+0.23	0.19	1.21	3.61	0.298
Summer book reading (all)		11	26	0.00	0.02	-0.20	9.47	0.847
Summer school mathematics (all)		5	12	+0.17	0.14	1.24	2.64	0.314
			Poverty as Moderator					
Summer school reading	high poverty	5	9	+0.21	0.19	1.11	3.22	0.345
	lower poverty	2	3	+0.29	0.24	1.18	1.85	0.367
Summer book reading	high poverty	6	8	+0.04	0.04	1.13	6.12	0.299
	lower poverty	7	18	-0.03	0.03	-0.86	7.23	0.416
Summer school mathematics	high poverty	3	6	+0.06	0.14	0.41	2.56	0.716
	lower poverty	3	6	+0.28	0.14	2.07	2.69	0.141
			Baseline Achievement as Moderator					
Summer school reading	high achievers	1	1	+0.15	0.08	1.85	2.45	0.182
	low achievers	4	7	+0.12	0.08	1.49	2.54	0.249
	mixed achievers	2	4	+0.43	0.54	0.81	1.06	0.561
Summer book reading	high achievers	1	1	+0.07	0.03	2.54	6.68	0.040
	low achievers	3	4	+0.10	0.05	2.12	2.64	0.136
	mixed achievers	9	21	-0.03	0.02	-1.18	8.49	0.271
Summer school mathematics	high achievers	1	1	+0.19	0.05	4.22	1.93	0.055
	low achievers	3	6	+0.10	0.05	2.16	1.97	0.165
	mixed achievers	2	5	+0.24	0.32	0.76	1.01	0.586

Note. SE=standard error; df=degrees of freedom; k=number of studies; n=number of effect sizes; Mean effect sizes for each moderator category were calculated by estimated a model including the same covariates as those shown in Table 1 without an intercept, with interaction terms for the differential effects of socioeconomic status and baseline achievement for each category.

Figure 1

Study Selection Flow Chart

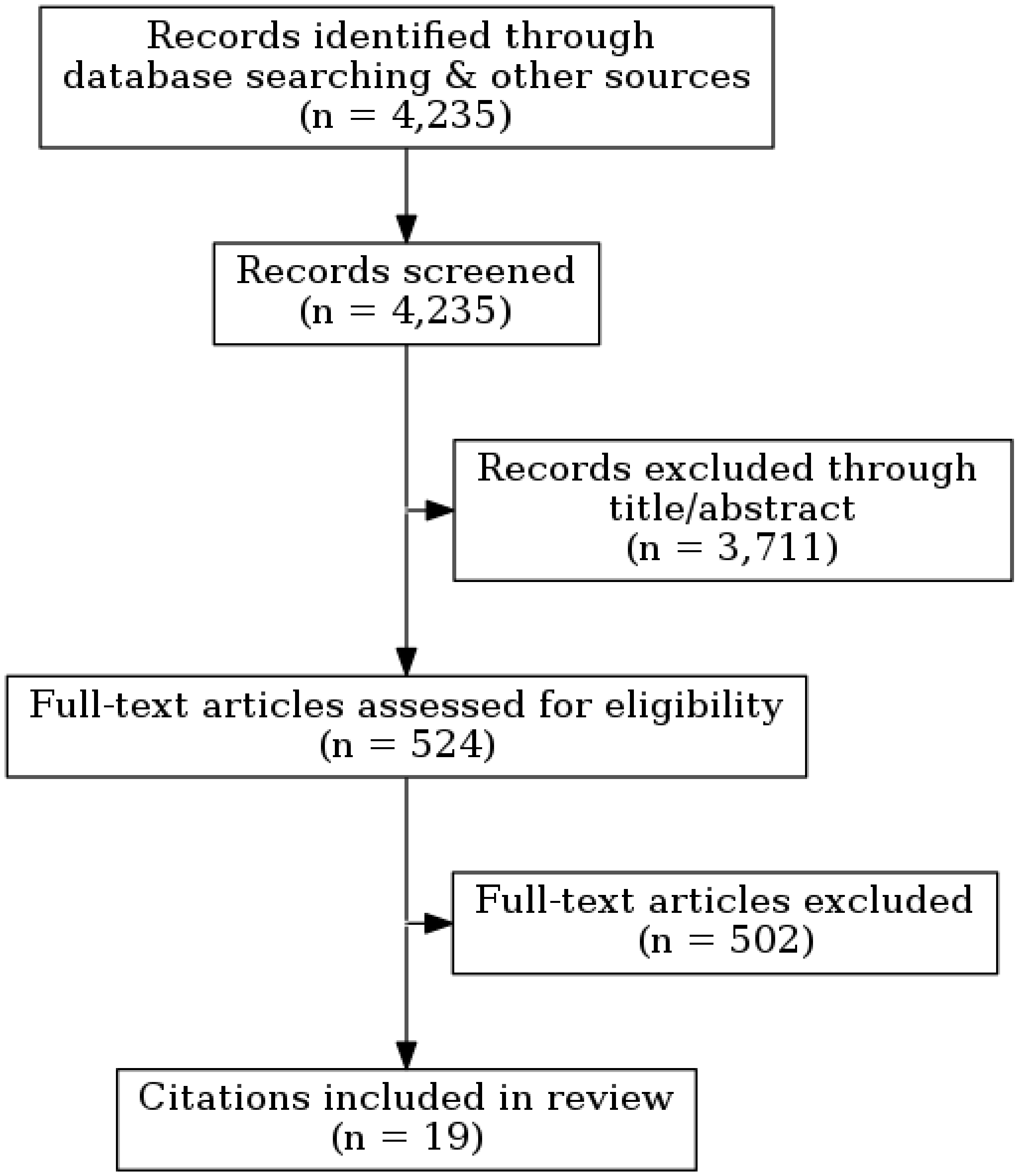

Table A1

Summer School Reading Programs

Study	Design	Duration	N	Grades	Sample characteristics	Posttest/ subgroup/timepoint	Effect sizes	Overall effect size [a]
Summer Reading Camp								Category Mean: +0.23
Schacter & Jo (2005)	Student randomized	7 weeks	3 schools, 118 students (54E, 64C)	1	3 schools in south Los Angeles. 63%AA, 37% H, 100% FRL.	Gates Total		+1.16
						Follow-up		
						Gates Total (December 2001, long-term effect)	+0.92	
						Stanford 9, Total (May 2002, long-term effect)	+0.16	
Teach Baltimore								
Borman & Dowling (2006)	Student randomized	7 weeks in each of 3 summers	10 schools, 686 students, (438E, 248C)	K-1	10 high-poverty urban schools in Baltimore City. 96% AA, 85% FRL.	CTBS/4 Total Reading		+0.01

District-sponsored academically intensive literacy program								
Zvoch & Stevens (2013)	Student randomized	5 weeks	93 students (47E, 46C)	K-1	Struggling readers in a district in a moderately sized city in the Pacific Northwest. 15% EL, 63% FRL.	DIBELS Nonsense Word and Fluency Kindergarten	+0.60	+0.69
						First grade	+0.78	
Building Educated Leaders for Life (BELL)								
Somers et al. (2015)	Student randomized	5 weeks	5 schools 919 students (585E, 334C)	5-7	Students performing below grade level in three districts, one in the West, and the Southeast. 45% AA, 34% H, 9% EL, 89% FRL.	GRADE Total		+0.01
Gorard et al. (2015)	Student randomized	4 weeks	43 schools, 310 students (167E, 143C)	5-6	Low-performing students in Brighton, Enfield, and Islington in the U.K.	Progress in English: Year 5	+0.12	+0.17
						Year 6	+0.31	

Voluntary district-run programs								
Augustine et al. (2016)	Student randomized	No less than 5 weeks in each of 2 consecutive summers	5 school districts, 4484 students (2544E, 1940C)	3-4	Low-income students in 5 urban school districts, Boston, Dallas, Jacksonville, Pittsburgh, Rochester. 47% AA, 40% H, 31% EL, 87% FRL.	GRADE : Year 1	0.00	
						GRADE: Year 2 (effect of 2 consecutive summers)	+0.04	+0.04
						Follow-up		
						Statewide standardized tests, Spring 2014 (long-term effect)	+0.01	
						Statewide standardized tests, Spring 2015 (long-term effect)	+0.03	

Abbreviations for Tables A1-A3: FRL: Free and Reduced Price Lunch (U.S.); FSM: Free School Meals (U.K.); AA: African-American; H: Hispanic; A: Asian-American; W: White; EL: English learner; CTBS: Comprehensive Test of Basic Skills; DIBELS: Dynamic Indicators of Basic Early Lteracy System; GMADE: Group Math and Diagnostic Evaluation; GMRT: The Gates-MacGinitie Reading Test; GRADE: Group Reading; Assessment and Diagnostic Evaluation; ITBS: Iowa Test of Basic Skills; MDTP: Mathematics Diagnostic Testing Project; STAR: Standardized Test for the Assessment of Reading; STEP: State Test of Educational Progress

Table A2

Summer Book Reading Programs

Study	Design	Duration	N	Grades	Sample characteristics	Posttest/ subgroup/ timepoint	Effect sizes	Overall effect size [a]
Reading Enhances Achievement During Summer (READS)								Category Mean: 0.00 Program Mean: 0.00
Kim (2006)	Student randomized	NA	10 schools, 486 students (252E, 234C)	4	4 Title I schools and 6 non-Title I schools. 19% AA, 26% H, 18% A, 56% EL, 40% FRL.	ITBS Total Reading	+0.08	+0.02
						DIBELS Oral Reading Fluency	-0.05	
Kim (2007)	Student randomized	NA	1 school 279 students (138E, 141C)	1-5	Suburban district. 42% W, 23% Spanish EL, 23% FRL.	Stanford 10 Reading: Grade 1	-0.01	+0.04
						Grade 2	+0.07	
						Grade 3	+0.13	
						Grade 4	-0.34	
						Grade 5	+0.31	

Kim & White (2008)	Student randomized	NA	2 schools, 400 students	3-5	2 public schools in a large suburban district in the mid-Atlantic region. 25% AA, 29% H, 29% EL, 38% FRL.	ITBS Total Reading: Book vs.C	+0.07	+0.03
						DIBELS Oral Reading Fluency	-0.03	
Kim & Guryan (2010)	Student randomized	NA	4 schools, 317 students	4	A public school district in California. 99% H, 73% EL, 96% FRL.	GMRT Total		+0.03
White et al. (2014)	Student randomized	NA	19 schools, 1188 students	3	Below-average readers in a midsized urban school district in North Carolina. 51% AA, 30% H, 72% FRL.	ITBS Reading Comprehension		-0.03
Guryan et al. (2014)	Student randomized	NA	59 schools, 5319 students (2659E, 2660C)	2-3	Below-average readers in 7 North Carolina school districts. 38% AA, 22% H, 17% EL, 77% FRL.	ITBS: Grade 2	+0.01	+0.01
						Grade 3	+0.02	

Guryan et al. (2016)	Student randomized	NA	9 schools, 397 students	3-4	Urban school district in the northeastern U.S. 43% AA, 11% H, 50% FRL.	GMRT Total	-0.10
Providing books and reminder postcards							
Wilkins et al. (2012)	Student randomized	NA	112 schools, 1571 students (791E, 780C)	3	Reading scores were below the 50th percentile in 4 Texas districts. 20% AA, 68% H 100% FRL.	Scholastic Reading Inventory	+0.02
Summer Active Reading							
Maxwell et al. (2014)	Student randomized	NA	48 schools, 182 students (93E, 89C)	6	Students at risk in the north of England. 34% FSM.	New Group Reading Test: FSM	+0.13

Summer reading clubs								
Dynia et al. (2015)	Student randomized	36 days, over 8 weeks	76 students, (41E, 35C)	2-3	Students at risk in Columbus, OH. 29% FRL.	Test of Word Reading Efficiency, Sight Words	-0.13	-0.08
						Test of Word Reading Efficiency, Phonemic Decoding	+0.04	
						Gates-MacGinitie Reading Comprehension	-0.14	
Text-messaging								
Kraft & Monti-Nussbaum (2017)	Household randomized	8 weeks	2 schools, 224 students (114E, 110C)	K-3	Struggling readers in a public charter school network in Rhode Island. 12% AA, 32% H, 53% FRL.	STAR, K & Grade 1	-0.08	+0.07
						STAR, Grade 2 &3	+0.15	
						STAR total	+0.03	
						STEP, K & Grade 1	-0.08	
						STEP, Grade 2 &3	+0.23	
						STEP total	+0.10	
						Follow-up		
						STAR (Nov. 2015, long-term effect)	+0.10	
						STEP (Nov. 2015, long-term effect)	+0.12	
						STAR	+0.07	

(Feb. 2016, long-term effect)	
STEP (Feb. 2016, long-term effect)	+0.13
STAR (June 2016, long-term effect)	+0.09
STEP (June 2016, long-term effect)	+0.19

[a] This column shows the near-term effect for the full sample. If more than one outcome measure were used, this column reports the mean of different measures.

Table A3

Summer School Mathematics Programs

Study	Design	Duration	N	Grades	Sample characteristics	Posttest/ Subgroup/ timepoint	Effect sizes	Overall effect size [a]
Elevate Math								Category Mean: +0.17
Snipes et al. (2015)	Student randomized	4 weeks	8 schools 349 students (165E, 184C)	7	Six districts in California's Silicon Valley. High basic level or the low proficient level. 34% A, 52% H, 57% FRL.	MDTP Algebra Readiness		+0.64
Tenmarks								
Lynch & Kim (2017)	Student randomized	10 weeks	4 schools, 196 students	3-9	A large, urban school district in the northeast region of the U.S. 38% AA,31% H, 59% FRL.	District curriculum-based assessment	-0.01	-0.01
						NAEP items	-0.02	

Building Educated Leaders for Life (BELL)								
Somers et al. (2015)	Student randomized	5 weeks	5 schools 919 students (585E, 334C)	5-7	Students performing below grade level in 3 districts in the West and the Southeast. 45% AA, 34% H, 9% EL, 89% FRL.	GMADE Total		+0.07
Gorard et al. (2015)	Student randomized	4 weeks	43 schools, 306 students (164E, 142C)	5-6	Low-performing students in Brighton, Enfield, and Islington in the U.K.	Progress in Math: Year 5	+0.12	0.00
						Year 6	-0.25	
Voluntary district-run programs								
Augustine et al. (2016)	Student randomized	No less than 5 weeks in each of 2 consecutive summers	5 school districts, 4505 students (2553E, 1952C)	3-4	Five urban school districts, Boston, Dallas, Jacksonville, Pittsburgh, Rochester. 47% AA, 40% H, 31% EL, 87% FRL.	GMADE in Fall 2013: first summer	+0.08	+0.03
						GMADE in Fall 2014: (the effect of 2 consecutive years of invitation to summer school)	+0.03	

Note: [a] This column shows the near-term effect for the full sample. If more than one outcome measure were used, this column reports the mean of different measures.

Table A4

Program Costs (in 2020 U.S. dollars)

Study	Cost (per student)	Cost structure	Overall Effect on Reading Achievement	Overall Effect on Mathematics Achievement
Summer school programs				
Snipes et al. (2015) (Elevate)	$560	Include the costs of a credentialed teacher, a college-level teaching assistant, a college field trip, and Common Core State Standards–based curriculum and professional development for teachers and the college-level teacher assistants. Laptops were provided by the classroom and a site principal.	NA	+0.64
Gorard et al. (2015) (BELL)	$1,995	Include administration, resources and activities ($546), salary costs and training ($1,302), and food and transportation ($289).	+0.17	0.00
Augustine et al. (2016) [a] (District Programs)	$1,340	Range from $1,198 to $1,904 with the average of $1,507. Include academic classroom staff salaries (35%), district and site administration (25%), enrichment costs (20%), transportation (7%), curriculum category (4%), professional development costs (4%), and foods (4%).	0.00	+0.08 (first year) 0.00 (two years)

Summer book reading programs				
Guryan et al. (2014)	$364	Ranges from $280 to $448. Include the costs of books, mailings, stipends to teachers, salaries of staff who managed the implementation, and testing.	+0.01	NA
Kim (2006)	$79	Includes the price for books, postage, and labor.	+0.02	NA
Maxwell et al. (2014)	$274	Ranges from $248 to $303. Includes resources, activity days, training, and salary and overheads.	+0.13	NA
Wilkins et al. (2012)	$31	Includes treatment group books, postcards and shipping.	+0.02	NA

Note: [a] The program evaluated by Augustine et al. (2016) lasted two consecutive summers, but the study only showed the average expenditures for the 2014 summer programs in three districts. Borman and Dowling (2006) is not listed in the table, as its program cost excluded in-kind support, evaluation costs, and the Teacher Readiness in Urban Education Master's program costs.

Figure A1

Boolean Operators for Initial Selection

	(comparison group OR control OR effect* OR efficacy OR evaluation OR impact OR matched group OR posttest OR pretest OR QED OR quasi-experiment* OR random* OR RCT OR treatment)
AND	(approach OR curricul* OR improvement OR intervention OR instruct* OR program OR remedial OR school OR strategy OR success OR teach OR tutor*)
AND	(elementary OR kindergarten OR first grade OR second grade OR third grade OR fourth grade OR fifth grade OR Grade 1 OR Grade 2 OR Grade 3 OR Grade 4 OR Grade 5 OR K-5 OR primary school OR secondary OR sixth grade OR seventh grade OR eighth grade OR ninth grade OR tenth grade OR eleventh grade OR twelfth grade OR Grade 6 OR Grade 7 OR Grade 8 OR Grade 9 OR Grade 10 OR Grade 11 OR Grade 12 OR middle school OR high school)
AND	(read* OR achievement OR comprehension OR decoding OR phonics OR phonemic awareness OR fluency OR vocabulary OR letter identification OR literacy OR woodcock OR word recognition OR word identification OR word analysis OR oral reading)
AND	Summer

Note: When retrieving studies whose outcome was mathematics achievement, the fourth parenthesis in the last query was replaced by (achievement OR measurement OR algebra OR number OR fractions OR numer OR geometry OR problem solving OR math* OR reasoning). The retrieval field for the index words was limited to "Title, Keyword, and Abstract".*