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

James S. Kim and David M. Quinn

Harvard University

Abstract

This meta-analysis reviewed research on summer reading interventions conducted in the

United States and Canada from 1998 to 2011. The synthesis included 41 classroom- and home-

based summer reading interventions, involving children from kindergarten to Grade 8.

Compared to control group children, children who participated in classroom interventions,

involving teacher-directed literacy lessons, or home interventions, involving child-initiated book

reading activities, enjoyed significant improvement on multiple reading outcomes. The

magnitude of the treatment effect was positive for summer reading interventions that employed

research-based reading instruction and included a majority of low-income children. Sensitivity

analyses based on within-study comparisons indicated that summer reading interventions had

significantly larger benefits for children from low-income backgrounds than for children from a

mix of income backgrounds. The findings highlight the potentially positive impact of classroom-

and home-based summer reading interventions on the reading comprehension ability of low-

income children.

Keywords: meta-analysis, reading comprehension, low-income children, summer learning loss

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

According to the 2011 administration of the National Assessment of Educational

Progress (NAEP) in Grade 4 reading, low-income children scored approximately three-fourths of

a standard deviation lower, on average, than middle-income children; in Grade 8 reading, this

gap was 65% of a standard deviation (National Center for Education Statistics, 2011). Reardon

(2011) analyzed data from 19 nationally representative data sets and found that income-based

disparities in student reading achievement have grown larger over the past four decades.

Although there are many underlying causes of income-based disparities in reading, low-income

children are particularly at risk of falling behind their classmates in reading during the summer

months (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Entwisle, Alexander, & Olson,

1997).

Effective summer interventions may be critical to improving children's reading

achievement from kindergarten to Grade 8, particularly for low-income children. Policymakers

have adopted two primary intervention strategies for improving children's reading achievement

during the summer months: classroom- and home-based summer reading interventions.

Classroom-based summer reading interventions are designed to remediate children's academic

weaknesses through instructional activities led by schoolteachers, college and graduate students,

and university researchers. A meta-analysis of experimental studies (Cooper, Charlton,

Valentine, & Muhlenbruck, 2000) indicated that classroom-based summer reading programs

improved student achievement by .14 standard deviations. More recently, home-based summer

reading interventions have been implemented as a potentially cost-effective strategy for

preventing reading loss among low-income children (McCombs et al., 2011).

During the past decade, the federal No Child Left Behind Act of 2001 has created strong accountability pressures for schools to close achievement disparities through the implementation of out-of-school time policies and research-based reading instruction (Lauer, Akia, Wilkerson, Apthorp, Snow, & Martin-Glenn, 2006; National Reading Panel, 2000). As a result, policymakers and practitioners have sought to implement summer reading interventions that show strong evidence of efficacy and use research-based instructional practices. Given the national imperative to close income-based disparities in student achievement, there is a growing need to understand the programmatic characteristics of effective summer reading interventions and their potential benefits for low-income children (McCombs et al., 2011). This updated meta-analytic review synthesizes results from 41 summer reading interventions, involving children from kindergarten to Grade 8.

Defining Summer Reading Interventions

Summer reading interventions are usually implemented inside or outside classrooms (McCombs et al., 2011). Although context is only one characteristic of a summer reading intervention, theorists (Bronfenbrenner, 1999, 2005) have suggested that children's classrooms and homes shape the proximal processes that drive literacy development. In a summer reading intervention, the classroom or home context is likely to shape the instructional goals and activities, the roles of children and adults, and the quality of children's literacy experiences.

In classroom interventions, the quantity and quality of teacher-directed literacy instruction is the critical mechanism that promotes reading achievement (Tseng & Seidman, 2007). Classroom interventions emphasize teacher-managed instructional activities, in which teachers are responsible for focusing student's attention on the literacy activity (Connor, Morrison, & Katch, 2004). Therefore, teachers in classroom-based summer programs usually

implement literacy lessons that are designed to remediate past academic weaknesses (Cooper et al., 2000), or to preview skills and knowledge that students may encounter in the upcoming school year (McCombs et al., 2011). In classroom interventions sponsored by public school districts, teachers implement curriculum-based literacy activities that are designed to improve children's comprehension outcomes (Mariano & Martorell, 2011; Matsudaira, 2008; Jacob & Lefrgren, 2004). More recently, community-based and non-profit organizations have trained college and graduate students and other non-school personnel to implement classroom interventions that focus broadly on improving children's academic achievement (e.g., reading and mathematics), social and emotional learning, and leadership skills (Borman & Dowling, 2006; Chaplin & Capizzano, 2006). Because classroom interventions have diverse program goals that target multiple child outcomes, the amount of time devoted specifically to literacy instruction is likely to vary across programs.

In home interventions, the quantity and quality of child-initiated book reading is the critical mechanism that promotes reading achievement. Children must initiate book reading activities independently or with their family members to enjoy gains in literacy achievement (Senechal & Young, 2008). Home interventions are usually designed to improve children's reading comprehension by (a) providing access to a wide variety of narrative and informational texts, (b) promoting intrinsic motivation to read at home, and (c) increasing print exposure during the summer months (Allington et al., 2010; Guthrie & Humenick, 2004; Heyns, 1978; Mol & Bus, 2010). Home interventions are based on the hypothesis that children who have mastered basic decoding skills need to read widely in order to develop a fully specified orthographic representation of words encountered in text and to acquire word and world knowledge (Anderson, Wilson, & Fielding, 1988; Share, 1999, Stanovich, 2000). To enhance the

effectiveness of home interventions, some researchers have also scaffolded summer book reading by including teacher lessons right before the summer, by improving the match between a child's independent reading level and the readability of text, and by encouraging parent involvement in home literacy activities (McCombs et al., 2011). Although developers of home interventions may implement diverse approaches to scaffolding summer book reading, the combination of effective teacher-directed comprehension lessons, careful text leveling strategies, and opportunities to read books for multiple summers appear to enhance comprehension gains (Allington et al., 2010; Kim & White, 2008; Mesmer & Cumming, 2009).

To date, researchers have not conducted meta-analytic reviews of home-based summer reading interventions involving child-initiated book read activities. Nonetheless, numerous studies indicate that children's book reading activities outside school are an important predictor of comprehension and vocabulary gains during the elementary and middle school grades (Entwisle, Alexander, & Olson, 1997; Stanovich, 2000). For example, in a longitudinal study involving 1,128 sixth- and seventh-grade students, Heyns (1978) found that measures of independent reading—i.e., the number of books read and time spent reading during the summer months—were positively related to vocabulary scores, controlling for measures of prior achievement, family income, parent education, and household size. Heyns also found that "children in every income group who read six or more books during summer consistently gained more than children who did not" (p. 169). In addition, recent research indicates that children's access to books and home reading activities are malleable variables that explain individual differences among low-income children's literacy achievement (Chin & Phillips, 2004; Teale, 1986; Whitehurst, Arnold, Epstein, Angell, Smith, & Fischel, 1994). Past research suggests that measures of independent book reading predict disparities in reading achievement (a) between

low-income and middle-income children, and (b) among children within income groups. Given

these findings, there is a clear need for cost-effective interventions that promote child-initiated

book reading activities at home during the summer months (Public Agenda, 2010).

What is Known About the Impact of Summer Reading Interventions?

During the past 15 years, researchers have conducted two meta-analyses of summer

programs. In 2000, Cooper, Charlton, Valentine, and Muhlenbruck published a comprehensive

meta-analysis of classroom-based summer programs, which were designed to remediate

children's academic deficiencies. Both the characteristics of the study design and the income

characteristics of participating children moderated program effects. A random effects model

indicated that the mean effect size of single group pretest-posttest designs ($d = .30$, $k = 81$) was

significantly larger than two group designs ($d = .09$, $k = 44$). Because single group designs fail

to eliminate numerous threats to internal validity, Cooper et al. (2000) asserted that the mean

effect size from randomized experiments ($d = .14$, $k = 11$) provided the most credible estimates

of summer program effects. More recently, Lauer et al. (2006) conducted a meta-analysis of out-

of-school time (OST) interventions that were implemented outside the regular school day in an

after school, a Saturday, or a summer program. The review included only two-group designs

comparing the posttest reading scores of participants and non-participants. The review found no

difference in mean effects for programs implemented in the summer ($d = .05$, $k = 14$) or after

school ($d = .07$, $k = 15$).

Findings from these two previous meta-analytic reviews suggest that summer school

effects may differ based on the quality of the evaluation design. The previous reviews also left

unanswered several questions that guided our meta-analytic review of summer reading

interventions. Because summer reading interventions are not a unitary construct, it is unclear

whether classroom- and home-based summer reading interventions produce similar effects on reading achievement. In addition, classroom and home interventions usually target more than one domain of reading achievement, underscoring the need to measure program effects on reading comprehension and its component skills such as word reading ability, oral reading fluency, and reading vocabulary.

Research Hypotheses and Study Goals

Three hypotheses guided our meta-analytic review of summer reading interventions. First, we hypothesized that classroom and home interventions would improve diverse reading outcomes. This hypothesis was based on findings from two meta-analytic reviews of summer programs from 1966 to 2003 (Cooper et al., 2000; Lauer et al., 2006). The key findings suggest an upper bound estimate of $d = .14$ (Cooper et al., 2000) and a lower bound estimate of $d = .05$ (Lauer et. al., 2006) in reading achievement based on experimental and quasi-experimental evaluations of summer school. These two reviews imply a plausible mid-point effect size of $d = .10$ in total reading achievement using an aggregated effect size that combines student performance on multiple subtests (e.g., reading comprehension and vocabulary). There is less prior information, however, to make predictions about program effects on different components of reading comprehension. Although many studies of summer programs have evaluated program effects on diverse domains of reading achievement, previous researchers have used aggregated effect sizes in their meta-analytic review. Theories of text comprehension, however, suggest that reading interventions may have larger effects on proximal predictors of reading comprehension such as decoding ability and literal understanding of the explicit textbase (Gough & Tunmer, 1986; Kintsch, 1994). To test this hypothesis, research syntheses should measure diverse and disaggregated components of reading comprehension. In addition, it is unclear whether and how

home-based interventions improve diverse reading outcomes. Although home-based summer

reading interventions have been employed as a complementary strategy for reducing summer

reading loss among low-income children (McCombs et al., 2011), no study to date has

synthesized results to determine whether opportunities to read at home improve reading

outcomes. Despite the dearth of synthesis-generated evidence, we predicted that child-initiated

book reading would increase print exposure and improve reading comprehension during summer

vacation. This prediction flows from substantial empirical research indicating that print

exposure is an important mechanism driving children's acquisition of word and world knowledge

and verbal ability across the life span (Byrnes, 2000; Mol & Bus, 2011; Stanovich, 2000).

Second, we hypothesized that the implementation of research-based reading instruction

summarized by the National Reading Panel (2000) and subsequent syntheses of primary studies

(Duke & Pearson, 2002; National Institute for Literacy, 2006; Shanahan et al., 2010) would

moderate intervention effects on reading outcomes. In particular, there is broad agreement

among researchers that literacy instruction needs to build children's phonological awareness,

decoding ability, oral reading fluency, reading vocabulary, and comprehension (Pressley, 2002;

Snow, Burns, & Griffin, 1998; Snow & Juel, 2007). The scientific consensus regarding the

importance of research-based instruction is rooted in the 1998 National Research Council Report

(NRC), *Preventing Reading Difficulties in Young Children* (Snow et al., 1998). This report

encouraged teachers to integrate instruction that enabled children to master the alphabetic

principle and, at the same time, to read for understanding from a variety of narrative and

informational text. Meta-analytic findings from the 2000 National Reading Panel (NRP) found

that effective teacher-directed instruction was critical to improving children's decoding ability,

oral reading fluency, and comprehension outcomes. More recently, the Institute of Education

Sciences (Shanahan et al., 2010) and the National Institute for Literacy (2006) have published

reviews for practitioners in the elementary and middle grades, recommending the use of

research-based comprehension strategies summarized by the National Reading Panel.[1] Given the

findings of past syntheses, we predicted that the implementation of research-based instruction

would moderate intervention effects on reading outcomes.

Third, we hypothesized that the effects of summer reading interventions would be larger

for low-income children than for middle- and high-income children. In the absence of an

intervention, low-income children may lose ground in reading during the summer months.

Results from meta-analyses, nationally representative surveys, and ethnographic research

indicate that low-income children are particularly at-risk of falling behind in reading

comprehension during summer vacation. For example, Cooper, Nye, Charlton, Lindsay, and

Greathouse (1996) found that summer vacation had a larger negative impact on the reading

comprehension scores of low-income children ($d = -.27$) than middle-income children ($d = -.14$).

Longitudinal analyses involving the Early Childhood Longitudinal Survey, Kindergarten Cohort

of 1998, indicate that low-income children are more at-risk of falling behind middle- and high-

income children in reading during the summer months than during the academic school year

(Downey, von Hippel, & Broh, 2004). Findings from ethnographic research also indicate that

many low-income children have limited opportunities to participate in high quality summer

programs and to read appropriately challenging and interesting books (Chin & Phillips, 2004;

Lareau, 2003). As a result, a classroom- or home-based summer reading intervention may create

a stronger treatment-control contrast in program activities and outcomes when low-income

children comprise the majority of program participants.

Conversely, summer reading interventions may have smaller effects among more affluent children who have access to high quality summer programs and books at home. Indeed, there is growing evidence that high-income families have dramatically increased investments in their children's education over the past 40 years. Using data from the Consumer Expenditure Surveys from the 1972-73 to 2005-06, Kornrich, Gauthier, and Furstenberg (2011) found that parental investment in education increased sharply among wealthy families. In particular, the increase was $2,344 among families with incomes in the highest decile compared to $255 in the middle decile and $338 in the lowest decile. These figures imply that children from high-income families are more likely than low-income to have access to educational resources that foster learning. As a result, the contrast between a summer intervention and the counterfactual situation (i.e., children's experience in the absence of a summer intervention) may be smaller among high-income children than among low-income children.   To date, however, only Cooper and colleagues (2000) have formally tested the moderating role of income status on achievement by comparing mean effects from studies with children from a range of family income levels. The results from a random-effects analysis indicated that summer school effects were larger for middle-income children ($d = .44$, 95% CI = .14/.26) than for low-income children ($d = .20$, 95% CI = .13/.75). Given the substantively important policy implications of Cooper et al.'s (2000) findings, there is a clear need to test the robustness of student income status as a moderator of intervention effects. To pursue this goal, we examined whether mean effects differed for studies with mostly low-income children compared to studies with mixed-income samples of children; in addition, we conducted within-study comparisons of mean effects for children from different income groups.

To summarize, we hypothesized that (a) classroom- and home-based summer reading interventions would improve diverse reading outcomes, (b) the implementation of research-based reading instruction would moderate intervention effects, and (c) summer reading interventions would have larger effects for low-income children than for middle- and high-income children.

Method

Selection Criteria and Literature Search Procedures

The articles included in our review met five selection criteria. In particular, studies had to (a) evaluate the effects of a classroom- or home-based summer reading intervention in the United States or Canada, (b) evaluate effects on a measure of reading achievement, (c) provide sufficient empirical information to compute an effect size (Cohen's d-index), (d) include students who were in kindergarten to eighth-grade (K-8) prior to enrollment in a summer reading intervention, and (e) use an experimental or quasi-experimental design students to compare the post-program performance of treatment students to control students who did not systematically participate in an alternative intervention. If researchers published multiple reports based on the same data, we included only one of these reports in our analyses (i.e., the final evaluation report). Our review included both studies published in peer-reviewed journals and unpublished studies.

We focused on K-8 summer reading interventions because prior research suggests that the loss of reading skills during summer occurs across this grade span (Cooper et. al., 1996). Therefore, we excluded pre-kindergarten and high school programs because these programs tend to have different goals compared to K-8 interventions. We also excluded studies on the effect of supplemental educational services that included both summer school and after-school programs when the studies did not report results that allowed us to isolate the unique effect of the summer component on student outcomes. Finally, we excluded studies using single-group pre/posttest

designs because they fail to protect against most threats to internal validity (Shadish, Cook, &

Campbell, 2002).

To identify primary studies, we searched (a) electronic databases and targeted internet

sites, (b) reference lists of previous research syntheses, and (c) research reports from targeted

state and local education agencies. Because Cooper et. al.'s (2000) meta-analysis included

studies published between January 1966 and August 1998, we searched for studies published

after August 1998.

Electronic databases. We searched the electronic databases of Academic Search

Premier, Education Abstracts, ERIC, PsycINFO, EconLit, and ProQuest Dissertations and

Theses database. Our searches contained two sets of key words or phrases; the first set was

designed to identify studies that met our programmatic inclusion criteria ("summer program*,"

"summer school*," "summer reading," "summer literacy," "summer enrichment," "summer

remedia*," "summer instruction*," "summer education*," "summer learning") and the second

set was designed to narrow the results to studies more likely to meet our methodological

inclusion criteria ("*experiment*," "control*," "regression discontinuity," "compared,"

"comparison," "field trial*," "effect size," "evaluation"). We linked the search terms within each

set using the operator *or*; we linked the two sets of terms with the operator *and*. These searches

yielded 1691 results, which we exported to RefWorks for review and elimination of duplicates.

We then read each study's abstract and downloaded the complete study when appropriate. In

cases where a more thorough review revealed that a particular study did not meet our inclusion

criteria, we discarded the study. In the end, we retained 31 of these studies.   In addition, we

searched the public online databases of Child Trends LINKS, What Works Clearinghouse, and

the Harvard Family Research Project's Out-of-School Time Database. We used Google to

search within the websites of foundations and research organizations that could potentially have relevant reports (i.e., MDRC, NBER, RAND, Mathematica, SEDL, Wallace Foundation) and searched references on the National Association of Summer Learning website. These searches resulted in one additional study that met our inclusion criteria.

Reference lists of published reviews. We hand-searched the reference lists of four research reviews that were published after Cooper et al.'s 2000 meta-analysis (Bodilly & Beckett, 2005; Lauer et al., 2006; McCombs et al. 2011; Terzian, Moore, & Hamilton, 2009). We also reviewed the reference lists of studies found through our electronic searches. Through these reference lists, we uncovered three additional studies that had not surfaced during our electronic searches.

Direct contact with researchers and policymakers. Marsh, Gershwin, Kirby, and Xia (2009) provide a list of states and school districts that mandate or recommend summer school participation for students who fail to meet a particular performance threshold. We contacted each state and school district through e-mail seeking any evaluations they may have conducted of their policy. This strategy did not produce any additional reports. Our three search channels yielded a total of 35 studies (involving 41 interventions), only three of which were included in Lauer et al.'s (2006) review and none of which were in Cooper et al.'s (2000) review.

Procedures for Coding Studies

Major independent variables: Classroom or home interventions. In our meta-analysis, the key independent variable was the context of the summer reading intervention (i.e., classroom or home). Classroom interventions (65%) were more widely adopted than home interventions (35%). Classroom interventions were usually implemented in a K-12 public school campus, a college or university campus, a public library, or a community-based organization.

The most common goal among classroom interventions was the remediation of learning difficulties (75%) followed by the prevention of summer learning loss for low-income children (45%). In classroom interventions, teachers instructed students by using resources such as text and curriculum to enhance student engagement and to improve reading comprehension and its component skills. Most home interventions were designed to reduce summer learning loss (93%) or to increase parent involvement (29%).

Research-based instruction and other program moderator variables. To determine whether a classroom-based summer reading intervention used research-based instruction, we identified a list of recommended practices published in the National Reading Panel (2000) as outlined in Appendix A. In particular, we compared this list of effective practices in Appendix A to the instructional techniques described in each classroom intervention included in our meta-analysis. Each study was coded using a dichotomous and ordinal measure of research-based instruction. For the dichotomous measure (1 = yes, 0 = no), we coded whether a study reported implementing at least one research-based instructional recommendation. For the ordinal measure (0, 1, 2 or more), we coded for the total number of research-based instructional recommendations that were implemented in a single study. Finally, we coded for program context and instructor characteristics, including the class size, the number of program hours per day, the total program hours, instructor type (e.g. certified or uncertified teachers), and whether instructors were trained prior to a program.

Methodological moderator variables. To evaluate the influence of study methods on reading outcomes, we created codes for study design and study quality. Study design was coded dichotomously based on whether an experimental design, in which participants were randomly assigned to conditions, or a non-experimental design was implemented (Shadish, Cook, &

Campbell, 2002). Non-experimental designs included regression-discontinuity analyses or

methods used to match treatment and control groups on one or more pretest measures. Our

dichotomous code for study design was supported by prior research suggesting that randomized

controlled trials yield impact estimates that are different from non-experimental studies (Bloom,

Michaelopoulous, Hill, & Lei, 2002; Cook, 2002; Lipsey & Wilson, 1993). Study quality codes

were based on the What Works Clearinghouse (WWC) Standards for determining whether a

study (a) did not meet WWC standards, (b) met WWC standards with reservations, or (c) met

WWC standards without reservations (Institute of Education Sciences, 2010). Although it is

unclear whether quality scores moderate effects in meta-analytic studies (Greenland, 1994), we

used the WWC standards because they address the major threats to internal validity (i.e.,

randomization, attrition, equivalence) and are now widely used to evaluate intervention research

involving multiple domains of child development (Durlak, Weissberg, Dymnicki, Taylor &

Schellinger, 2010).

      Participant characteristic moderator variables. To code the characteristics of

participating students, we used the more narrow term "income status" rather than

"socioeconomic status" because none of the primary studies in our review used a composite

measure of socioeconomic status based on parents' income, education level, and occupational

status (Entwisle, Alexander, & Olson, 1997). Moreover, previous meta-analytic research (Sirin,

2005) indicates that the magnitude of the correlation between student free and reduced-price

lunch status and achievement are similar to correlations between parent socioeconomic status

measures (i.e., occupational status, income, education) and achievement. We coded each study

in our meta-analysis as having either a "low-income" sample, a "mixed-income" sample, or as

not reporting the income status of its participants. For a study to be classified as "low-income,"

it had to describe the studied intervention as being designed for low income students, or report

the percentage of students in the sample who were eligible for free/reduced-price lunch (FRL), or

who met some other measure of low-income status.[2] In the second type of studies, we coded

samples as "low-income" if 50% or more of the sample was classified as low-income or as FRL-

eligible. If a study did not report the FRL percentage for the sample, but did report the FRL

percentage for the school or district from which the sample was drawn, we used that information

to code the study's sample. No study reported the percentage of students who were "middle-

income" or "high-income." Therefore, in studies that reported that fewer than 50% of its

students were FRL-eligible, we could not determine the income status of the majority of the

sample. In such instances, we coded the study as having a "mixed-income" sample. In one

study (Paris et al., 2004), the researchers randomly selected 12 school districts in Michigan and

evaluated the summer school programs in those districts. Although this report did not include

income data, we coded this as a "mixed-income" study, under the assumption that a random

sampling of districts would result in a mix of income groups. Finally, if a study made no

mention of the income status of its participants, we coded it as "unreported."

      Of the 23 studies that report the percentage of low income students in their samples, the

median was 70 percent. We conducted an income moderator analysis on these studies using a

median split and the results were not substantially different from those found in the full sample

of studies. We therefore report the results of moderator analyses based on the original coding

described above, which include a larger sample size of studies. To assess the robustness of our

findings, we also identified seven studies that reported a separate effect size for low-income and

middle-income students. We compared the magnitude of the treatment effects for each subgroup

and tested the significance of the mean differences. Within-study subgroup comparisons control

for many study-level characteristics that may be confounded with the income characteristics of a study sample.

Student reading outcomes. We coded five student reading outcomes, including (a) total reading achievement, (b) reading comprehension total, (c) reading comprehension only, (d) fluency and decoding combined, and (e) reading vocabulary. First, to summarize the overall effect of each intervention, we created the *total reading achievement* outcome. Because two earlier reviews of summer programs (Cooper et al., 2000; Lauer, et al., 2006) used an aggregated measure of reading achievement that combined diverse measures, we created a total reading achievement outcome to be consistent with prior research and to compare our mean effects with the results of the two earlier reviews. Thus, for the first outcome measure, we generated one overall effect size per intervention that averaged together the effect sizes for each of the intervention's posttests. For example, a study may have separately assessed and reported posttest scores for fluency, reading comprehension, and vocabulary. If the effect sizes for these domains were $d = 0.13$, $d = 0.08$, and $d = 0.01$, respectively, this study's "total reading achievement" effect size would be the mean of these three effect sizes, or $d = 0.07$. Second, the *reading comprehension total* outcome included effect sizes from standardized tests that assessed reading comprehension as well as other reading skills. These effect sizes usually included the combined comprehension and vocabulary scores from a nationally norm-referenced test (e.g. the ITBS or Gates-MacGinitie) and total scores from a state's standardized test that assesses multiple literacy domains. Third, the *reading comprehension only* outcome was based on tasks that required children to read connected text and then answer multiple-choice questions.[3] Fourth, the *fluency and decoding combined* outcome includes the effect sizes from measures of oral reading fluency with effect sizes of decoding assessments. Oral reading fluency measures required

children to read connected texts with accuracy and speed, and decoding measures required children to read real words and pseudo-words from lists. Fifth, the *reading vocabulary* outcome assessed children's ability to identify the correct definition of a word that was not embedded in connected text.

Coder Reliability

We created a codebook to collect information from each included study and developed a procedure for estimating the reliability of the study codes. Two raters coded a random 20% sample of studies. Kappa coefficients adjust for chance agreement between raters and was high across coded study characteristics (mean $\kappa = .93$). All coding inconsistencies were resolved in follow-up meetings between coders.

Calculation of Effect Sizes and Analytic Strategy

The goal of meta-analysis is to combine the results of independent studies and to identify potential study-level moderators that explain variability in treatment effects. To conduct a meta-analysis, each study-level treatment effect must be converted to a standardized mean difference, or effect size. In this study, we computed Cohen's *d* for each study (i.e., the difference between the treatment and control group divided by the pooled standard deviation). We used a shifting unit of analysis to ensure that effect sizes were independent. For example, as noted earlier, we aggregated effects within a given intervention to generate a single effect size called "total reading achievement." For the analyses involving specific reading measures, we used the one effect size per intervention in order to maintain independent observations in the analytic models. For example, we identified studies that measured "reading vocabulary" and pooled these effects to generate the mean effect reported in the results section.

Random effects models. Because summer reading interventions vary along a number of

dimensions and because we were interested in making inferences back to the population of

studies from which our studies were sampled, we used a random effects model to pool effect

sizes. The random effects model includes both a within-study weight (inverse of the study

variance) and a between study-variance component. The random effects model can be viewed as

a special case of a multilevel model (Raudenbush & Bryk, 2002, pp. 209-210) in which the level

model 1 is given by the formula

$$d_j = \delta_j + \varepsilon_j, \text{(1)}$$

where d_j is the effect size (i.e., standardized mean difference between the treatment and control

group) for study j, δ_j is the parameter estimate for the true effect size, and ε_j captures error due to

the sampling of participants within study j and $\varepsilon_j \sim N(0, \sigma_j^2)$. The level 2 model can be written as

$$\delta_j = \ 0+ \qquad\qquad + \qquad , \text{(2)}$$

where   are coded study-level characteristics,   are parameter estimates, and   is the

study-level random error, where we assume that $\sim N(0, \tau)$. Substitution of the level 2 model

within the level 1 model yields a mixed effects model of the following form

$$d_j = \ 0+ \qquad\qquad + \qquad + \qquad \text{(3)}$$

To estimate the parameters in model 3, we used the *metan* command in Stata along with the

random option, which employs the method of moments procedure to estimate the between study

variance components (DerSimonian, & Laird, 1986). Ninety-five percent confidence intervals

that did not include $d = .00$ led us to reject the null hypothesis that the mean effect size was 0.

Because significance tests are sensitive to the number of studies, we also highlight the precision

of the 95% confidence interval in reporting the results.

Homogeneity and moderator analyses. For the homogeneity analysis, we computed a Q statistic to test the null hypothesis that mean effects were homogenous and estimating a common effect. The observed Q statistic follows a χ^2 sampling distribution and is essentially a weighted sum of squares statistic given by the following formula:

$$= =1 \quad (-)2 \ (4)$$

where   is a measure of precision based on the inverse of the within and between study variance estimate (Hedges & Olkin, 1985). The observed Q statistic is compared to the expected Q statistic, which is based on the degrees of freedom ($k - 1$), where k is the number of independent comparisons. We computed a Q_{total} statistic using the mean of combined weighted effect sizes for the full sample.

To conduct a moderator analysis, the Q statistic can be partitioned into a within group (Q_w) and between group (Q_b) component. More precisely, we examined whether variation among subgroup means was statistically significant (Q_b), and computed a 95% confidence interval for the difference between two mean effects (Hasselblad & Hedges, 1995). To supplement the null hypothesis tests involving the Q statistic, we computed an I^2 statistic to assess the amount of heterogeneity in effects between studies, ranging from 0 to 100%. Higgins et al. (2003) offer the qualitative benchmarks for describing the magnitude of heterogeneity in mean effects across studies (i.e., I^2 statistics near 15% reflect low heterogeneity, 25% to 50% reflect moderate heterogeneity, and 75% and above reflect high heterogeneity).

Within-study comparisons of subgroup mean effects. Although moderator analyses yield important information on study-level characteristics that explain differences in mean effects, they fail to protect against numerous threats to internal validity. For example, if studies with mostly low-income children have mean effects that differ from studies with middle-income children, it is unclear whether other study-level characteristics such as the quality of the study

design or the grade-level of participating students may be influencing treatment effects. To assess the robustness of the results from the income moderator analyses, we conducted within-study comparisons that enabled us to rule out study-level confounds. We employed a fixed-effect model for these comparisons because our goal was to make inferences about the subset of studies for which within-study comparisons were possible. .

In addition, a subsample of studies assessed program impacts on an immediate and a delayed measure of program effects.   For each comparison, we created a new effect size—the difference between the two subgroup mean effect sizes (2 - 1), where 2 is the delayed effect measured at time 2 and 1 is the immediate effect measured at time 1. For the comparison of mean effects that were measured immediately after an intervention and at follow-up, we created a variance of the difference between the two means by taking into account the correlation between the outcome measures. The variance of the difference is given by the formula

$$\text{var}(2 - 1) = 1 + 2 - (2r\ 1\ 2)\ (5)$$

where V_1 and V_2 represent the variance of each outcome measure (Borenstein et al., 2009). For the variance of the immediate and delayed effects, r is the estimated correlation between the two outcomes. We used an average value of the correlation ($r = .50$) to compute the variance of the difference and checked the robustness of the results using a lower-bound ($r = .25$) and upper bound ($r = .75$) estimate of the correlation between outcomes.

Ruling out rival hypotheses of findings. We addressed several alternative explanations for the findings by probing (a) the effects of study design, (b) the potential influence of publication bias, and (c) the impact of nested designs in classroom interventions. Finally, we compared the mean effects on immediate (< 1 month) and delayed measures (3+ months) for a subset of studies that reported post program effects on two measurement occasions.

Results

The results are reported in four main sections. To address the first hypothesis, we examined whether classroom and home interventions improved diverse reading outcomes. To address the second hypothesis, we assessed the moderating role of research-based instruction and other program characteristics on reading outcomes. To address the third hypothesis, we compared the magnitude of mean effects for studies with a majority of low-income samples and for studies with mixed income samples. In addition, we conducted within study analyses by comparing the magnitude of the mean effect size for children from low-income backgrounds and mixed income backgrounds. Finally, to check the robustness of our findings, we addressed several rival hypotheses that could potentially explain the main findings.

Descriptive Characteristics of Interventions and Studies

Table 1 summarizes the characteristics of the 35 studies that met our inclusion criteria. Most studies were either journal articles (31%) or dissertations (49%). Over one-third of the interventions occurred in urban settings and 40% used an experimental design or regression discontinuity design.   Most studies involved K-5 students, and low-income children comprised the majority of participants in 60% of the studies. In addition, approximately two-thirds of the interventions were classroom-based programs and 35% of those reported using at least one research-based instructional practice summarized by the National Reading Panel. The studies measured a variety of reading outcomes, including fluency and decoding, vocabulary, and comprehension, and 56% of the effect sizes included in this study measured intervention effects 1 month after the conclusion of the intervention. Four studies in our meta-analysis reported effects for multiple interventions, yielding independent effect sizes from 41 interventions. Table 2 provides descriptive characteristics for each of the 41 summer reading interventions. The

sample sizes of the studies ranged from large regression-discontinuity analyses of district

sponsored summer programs (Jacob & Lefgren, 2004; Mariano & Martorell, 2011; Matsudaira,

2008) to smaller studies involving home interventions.

Mean Effects of Classroom and Home Interventions

Table 3 reports the mean effect size and the associated 95% confidence interval, Q

statistic, and I^2 statistic for all summer reading interventions and separately for classroom and

home interventions. Combining results from 41 interventions yielded a grand mean effect on

total reading achievement of $d = 0.10$ (95% $CI = 0.04/0.15$). The statistically significant Q

statistic of 82.44 ($p < .001$) and the I^2 value of 52% revealed moderate heterogeneity in effect

sizes among studies. In addition, mean effects were also positive and significant for reading

comprehension total ($d = .13$), reading comprehension only ($d = .23$), and fluency and decoding

combined ($d = .24$). In the seven studies that reported effects for a decoding measure only, the

effect size ($d = .43$) was larger than for the other reading outcomes (result not shown in Table 3).

The magnitude of the effect size across the five outcome measures was similar for

classroom and home interventions. More precisely, there was no significant difference in the

mean effects of classroom and home interventions on each of the five outcome measures. In

addition, the disaggregated findings show that classroom and home interventions improved

reading comprehension total scores by approximately one-tenth of a standard deviation.

Although both types of interventions improved reading comprehension only outcomes by

approximately one-fourth of a standard deviation, the mean effect size for home interventions

was not statistically significant ($d = .22$, 95% CI $= -.03/.48$). The magnitude of the treatment

effect on the fluency and decoding combined outcome was also similar for both intervention

settings, although the effect size for classroom interventions ($d = .22$) included $d = .00$. For both

types of interventions, there was no significant effect on reading vocabulary. In sum, these results indicate that both classroom and home interventions improved total reading achievement and reading comprehension total outcomes, and the magnitude of the treatment effects on each of the five outcome measures was similar.

One important difference between classroom and home interventions is related to the degree of between study heterogeneity in effect sizes. In general, both the Q statistics and the I^2 values were larger for classroom than for home interventions. Among classroom interventions, the Q statistic was significant for all outcomes except reading vocabulary, leading to the rejection of the null hypothesis that effects were homogeneous. The moderate to large I^2 values ranged from 58% to 79% suggesting substantial between study heterogeneity in effects among classroom interventions. For home interventions, however, the Q statistics for reading comprehension only and fluency and decoding combined outcomes led us to reject the assumption of homogenous effects, and the I^2 value for both outcomes were smaller in magnitude than the I^2 value for classroom interventions.

Research-based Instruction and Program Moderators of Reading Outcomes

Table 4 displays the results of moderator analyses involving research-based instructional practices, as summarized in the National Reading Panel (2000) report. More precisely, there was a positive impact of classroom interventions using research-based instruction on reading comprehension total ($d = .38$). However, interventions that did not report using research-based instruction had no significant impact on four of the five outcome measures. Inspection of the magnitude of the mean effects on four outcomes revealed moderate to large effects ($d = .25$ to $d = .63$) in classroom interventions reporting the use of research-based instruction and smaller mean effects ($d \leq .18$) for those not reporting the use of research-based instruction. For reading

comprehension total, there was suggestive evidence that research-based instruction moderated

mean effects, $Q_b(1) = 3.16$, $p = .075$.[4]

To probe the source of heterogeneity in mean effects among classroom interventions, we

examined whether coded characteristics of programs and instructors moderated treatment effects.

First, we created a median split for the 14 studies reporting class sizes (≤ 13 students or > 13

students), the 23 studies reporting the number of program hours per day (≤ 4.0 hours or > 4.0

hours), and the 20 studies reporting the total program hours (≤ 70 hours or > 70 hours). There

was no significant difference in the magnitude of the effect size for small class sizes ($d = .17$,

95% CI = -.02/.37) and large class sizes ($d = .02$, 95% CI = -.10/.13). There was no significant

difference in the mean effect of shorter programs ($d = .09$, 95% CI = -.02/.20) and longer

programs ($d = .15$, 95% CI = .03/.27) using an hour per day measure. There was also no

significant difference between less intensive programs ($d = .21$, 95% CI = -.02/.43) and more

intensive programs ($d = .11$, 95% CI = .01/.20) using a total program hour measure. We also

conducted a follow-up analysis to examine mean effects for resource intensive classroom

interventions that had (a) fewer than 13 students per class, (b) 4 to 8 hours of instruction per day,

and (c) 70 to 175 hours of total instruction. Twelve studies provided sufficient information (i.e.,

codes for all three program characteristics) to compare mean effects based on whether classroom

interventions were resource intensive. There was a positive effect on total reading achievement

for the five studies ($d = .25$, 95% CI = .01/.48) that met the criteria for being resource intensive,

and a non-significant effect for the seven studies ($d = .03$, 95% CI = -.12/.18) that failed to meet

the criteria.

Second, we also used a categorical measure for instructor credentials and for whether

teachers received program-specific training. Instructor type did not moderate outcomes (certified

teachers: d=.06, 95% CI = -.06/.17; college/graduate student: d=.39, 95% CI -.34/1.12; mix:

d=.06, 95% CI -.00/.12). Finally, there was no difference in mean effects for interventions that

provided training for instructors (d = .03, 95% CI = -.05/.11) and those that did not provide

training (d = .17, 95% CI = .04/.30).

Income Status Moderators of Reading Outcomes

Table 5 presents the results of moderator analyses based on the income status of

participating children. Inspection of the effect size and 95% confidence intervals shows that

intervention effects were positive and significant for majority low-income samples for total

reading achievement (d = .10), reading comprehension total (d = .20), reading comprehension

only (d = .33), and fluency and decoding combined (d = .23). Among mixed-income samples,

however, only the effect size for fluency and decoding (d = .27) was positive and statistically

significant. Most importantly, income status moderated effects on reading comprehension. For

reading comprehension total, the mean effect size for majority low-income samples (d = .20,

95% CI = .11/.29) was significantly higher than the mean effect size for mixed-income samples

(d = .00, 95% CI = -.11/.10), $Q_b(1)$ = 8.81, p = .04. For reading comprehension only, the mean

effect size for majority low-income samples (d = .33, 95% CI = .14/.53) was significantly higher

than the mean effect size for mixed-income samples (d = -.05, 95% CI = -.23/.14), $Q_b(1)$ = 7.58,

p = .006.[5]

We conducted a within-study sensitivity analysis to check the robustness of our

moderator analyses involving student income status. To conduct this analysis, we used data from

a subset of seven studies that included separate effects on total reading achievement for children

from low-income backgrounds and children from a mix of income backgrounds. The results of

the fixed-effect analysis indicated that mean effects were .28 standard deviations higher for

children from low-income backgrounds ($d = .14$, 95% CI = .06/.22) than for children from a mix

of income backgrounds ($d = -.14$, 95% CI = -0.19/-0.10). Results from our moderator analyses

and our within-study comparisons of mean effects are convergent, suggesting that intervention

effects were largest for children from low-income families.

These results, however, provide limited information on the specific reasons why student

income moderates treatment effects. To shed light on the possible mechanisms driving income-

based differences in mean effects, we compared spring-to-fall change in reading scores for

control group students in low-income samples and mixed-income samples of children. The goal

of this analysis was to understand whether student income moderated the magnitude of summer

loss (or gain) in reading scores. In this analysis, we identified studies that reported a pre- and

posttest score and computed standardized mean gains to understand whether fall scores were

different from spring scores (Cooper et al., 1996). Table 6 displays standardized mean gains by

income status on three outcome measures. For total reading achievement, income status

moderated gain in spring to fall scores for control students, Q_b (1) = 5.40, $p = .02$. More

precisely, among samples with a majority of low-income students, children in the control group

showed no change from spring to fall on the total reading achievement measure ($d = -.05$, 95%

CI = -.22/.12). Among the mixed-income samples, however, control children enjoyed a positive

reading gain from spring to fall in the total reading achievement ($d = .26$, 95% CI = 0.07/0.45).

The income characteristic of the sample was a marginally significant moderator of reading

comprehension total, $Q_b(1) = 3.34$, p = .068. Consistent with the previous results, mixed-income

samples enjoyed larger spring to fall gains in reading comprehension total than majority low-

income samples.   In sum, these findings indicate that summer vacation had a larger negative

effect on the readings scores of control group children in low-income samples than mixed-income samples.

Addressing Rival Hypotheses

To rule out rival hypotheses for the main findings, we probed (a) the effects of study design, (b) the possible influence of publication bias, (c) the impact of nested designs in classroom interventions, and (d) the effects of delayed measurement on the magnitude of treatment effects.

Effects of study design. First, we conducted moderator analyses based on study design using multiple approaches. In the first approach, we found no evidence that mean effects on total reading achievement differed for experimental (d = 0.09, 95% CI = 0.02/0.17) and non-experimental designs (d = 0.11, 95% CI = 0.03/0.18). In addition, we also used the What Works Clearinghouse (WWC) evidence standards to create a quality scale based on whether a study employed a randomized controlled design, showed evidence of baseline equivalence, and had low overall and differential attrition (Institute of Education Sciences, 2010). Specifically, we examined whether evidence standards moderated mean effects (a) for all studies and (b) for a subset of studies that employed a randomized controlled design or a regression-discontinuity design.

Using data from all studies, we conducted moderator analyses that revealed statistically equivalent mean effects for studies meeting WWC standards with or without reservation (d = .08, 95% CI = .02/.13) and for studies not meeting WWC standards (d = .16, 95% CI = .05/.27). In addition to comparing mean effects based on study quality, we examined whether mean effects were homogenous based on the WWC standards. Among studies meeting WWC standards (with and without reservations), we were not able to reject the assumption of homogenous effects for

studies meeting standards, $Q(15) = 23.33$, n.s., suggesting that the variability in effects was driven largely by sampling error. We were, however, able to reject the assumption of homogenous effects for studies not meeting standards, $Q(24) = 58.44$, p < .001, suggesting that there was more variability in mean effects among studies not meeting WWC standards than studies meeting WWC standards. In other words, studies that did not meet WWC standards yielded more heterogeneous effects than studies meeting WWC standards. Using data from studies using only a randomized controlled design or a regression discontinuity analysis, we found that mean effects were positive for studies meeting WWC standards without reservations ($d = .08$, 95% CI = .03/.14). However, there was a non-significant treatment effect in studies that either met WWC standards with reservations ($d = .00$, 95% CI = -.17/.18) or studies that did not meet WWC standards ($d = .07$, 95% CI = -.07/.20).

Potential influence of publication bias. Second, published studies in peer-reviewed journals ($d = .11$, 95% CI = .03/.18) and unpublished studies ($d = .13$, 95% CI = .03/.22) yielded statistically equivalent effects on reading comprehension outcomes. We also conducted a failsafe N analysis, which indicates the number of non-significant effects that would be needed to overturn the positive and significant results (Orwin, 1983; Rosenthal, 1979). The failsafe N of 264 exceeded the 215 cutoff for our sample, providing suggestive evidence that publication bias was not driving our main findings. The failsafe N analysis, however, is based only the statistical significance of results. As a follow-up test of publication bias, we used the trim and fill method for a subset of data that yielded homogenous results (Duval & Tweedie, 2000). We focused on results that yielded homogenous results because the trim and fill method may underestimate the true population effects if there is significant heterogeneity (Peters, Sutton, Jones, Abrams, & Rushton, 2007). For the homogenous effects reported in Table 3 (i.e., home intervention effect

size for reading total, $d = .12$, and reading comprehension total, $d = .11$), the trim and fill

analysis yielded mean effect sizes that remained statistically significant and similar in magnitude

to the original results. These analyses suggest that our results were robust to publication bias.

Impact of nested designs. Third, we assessed the impact of nested designs by adjusting

the standard errors in classroom interventions and then assessed the significance of the mean

effects reported in Table 3. In particular, we re-analyzed the data using a variance estimate that

takes into account the clustering of students within classrooms. When adjustments were made to

the variance of the effect size, the effect size for classroom interventions remained positive and

statistically significant in three of the outcomes in table 3 (total reading achievement, reading

comprehension total, reading comprehension only).[6]

Effects of delayed measurement of treatment effects. Fourth, we conducted a within-

study comparison of mean effects to rule out the possibility that positive effects stemmed largely

from the immediate measurement of program effects. For this analysis, we found seven studies

that administered an immediate (1 month) and delayed (3 or more months) measure of post-

program effects. The combined mean weighted effect size was larger on immediate measures (d

$= .52$, 95% CI $= .32/.73$) than on delayed measures ($d = .20$, 95% CI $= .00/.41$). The magnitude

of the delayed measures of program impact was approximately one-third of a standard deviation

lower than the magnitude of immediate measures of program impact. When using an upper

bound ($r = .75$) and lower bound ($r = .25$) estimate of the correlation between the immediate and

delayed measures, the results showed that the delayed effects were significantly smaller than the

immediate effects. In our sample of seven studies for which within-study comparisons of

immediate and delayed effects were possible, the magnitude of intervention effects clearly

diminished over time.

Discussion

Three major hypotheses motivated this meta-analytic review of classroom- and home-based summer reading interventions involving children from kindergarten to Grade 8. In particular, we hypothesized that (a) both classroom and home interventions would improve diverse reading outcomes, (b) the implementation of research-based reading instruction would moderate intervention effects, and (c) summer reading interventions would be most effective for low-income children. We review the results related to each hypothesis, place the findings in a broader research context, and discuss the study limitations and research implications.

What is the Impact of Classroom and Home Interventions on Diverse Reading Outcomes?

Combining results from 41 independent samples yielded a mean effect size of $d =.10$ on a composite measure of total reading achievement. Furthermore, the average effect size ([.23 + .04] / 2 = .135) based on reading comprehension only ($d = .23$) and reading vocabulary ($d = .04$) is quite similar to the effect size for the reading comprehension total outcome ($d = .13$). This finding implies that composite measures of reading achievement used in two earlier meta-analytic reviews (Cooper et al., 2000; Lauer et al., 2006) may have obscured the comparatively larger effects on reading comprehension than reading vocabulary. To place the magnitude of the mean effects in a broader research context, it is useful to compare our findings with research on summer programs in particular and education interventions in general. Cooper et al. (2000) found that randomized experiments of summer school programs focused on the remediation of learning difficulties improved reading achievement scores by .14 standard deviations, and Lauer et al. (2006) found that out-of-school time programs in the summer improved reading achievement by an average of .05 standard deviations. The magnitude of the effect size for the composite reading outcomes (i.e., reading achievement total and reading comprehension total)

was within the lower and upper bound estimates generated by these two earlier reviews of

summer programs. In addition, the comparatively larger effect size for reading comprehension

only outcome ($d = .23$) was within the lower and upper bound estimates of the mean impact of

76 educational interventions from kindergarten to Grade 12 (Hill, Bloom, Black, & Lipsey,

2007).[7]

Although classroom and home interventions had a positive impact on composite

measures of reading achievement, there was clear evidence that the magnitude of the mean effect

was larger for decoding ability than reading vocabulary. The mean effect for decoding ability (d

$= .43$; $k = 7$) was substantially larger than the mean effect for reading vocabulary ($d = .04$; $k =$

12), although these estimates are based on a small number of independent samples. How do we

explain these differences? The simple view of reading (Gough & Tunmer, 1986) suggests that

reading comprehension is the product of decoding ability and linguistic comprehension.

Moreover, procedural skills such as a child's ability to phonologically decode new and unknown

words are susceptible to loss without extensive practice (Cooper et al., 1996; Geary 1995; Share,

1999). Despite the positive effects on reading comprehension and decoding outcomes, neither

classroom nor home interventions had a positive impact on reading vocabulary. Perhaps the

most obvious explanation for this finding is that only 3 of the 12 studies that measured

vocabulary outcomes actually reported including teacher- or child-managed instructional

activities that were designed to improve vocabulary outcomes (Chaplin & Capizzano, 2006;

Pagan, 2010; Paris et al., 2004). Furthermore, most home interventions provided children with

opportunities to read books at home for a single summer. The one study (Allington et al., 2010)

that provided children with books for three consecutive summers did not measure reading

vocabulary. Given the low probability that a reader will learn a new word during normal reading

(Swanborn & de Glopper, 1999), low-income children may need frequent opportunities to read

connected text for multiple summers to enjoy a significant improvement in reading vocabulary.

Because acquisition of new words through wide reading is an incremental process (Swanborn &

de Glopper, 1999), a summer reading intervention carried out over three months is unlikely to

improve reading vocabulary.

Does Research-Based Instruction Moderate the Effects of Summer Reading Interventions?

Given limited information on the quality of classroom instruction, there is a clear need to

understand the precise research-based instructional practices that moderate treatment effects in

classroom interventions. Moreover, the presence or absence of research-based instruction is a

binary distinction with limited information on the degree to which teachers actually implement

research-based instruction in classroom lessons. Despite these data limitations, the general

pattern emerging from our moderator analyses indicate that classroom interventions using

research-based instruction produced more positive effects, ranging from $d = .25$ on total reading

achievement to $d = .63$ on fluency and decoding combined. Classroom interventions that did not

employ research-based instruction yielded smaller effects ($d \leq .18$) on each of the five reading

outcomes.

Among classroom interventions that reported using research-based instruction, the I^2

values were greater than 70% for three outcomes (i.e., total reading achievement, reading

comprehension total, reading comprehension only), reflecting a high degree of heterogeneity

(Higgins et al., 2003). What are the sources of heterogeneity in these mean effects? One

possible explanation is that classroom interventions using research-based instructional practices

vary in their program goals and the amount of time devoted to literacy instruction, resulting in

heterogeneous effects on student reading outcomes. Given the limited duration of summer

programs and the challenge of maintaining high attendance, classroom interventions that emphasize a variety of goals may devote less time to literacy instruction than programs with more focused goals. We did not formally assess whether variation in the quantity of time devoted to literacy instruction explain variation in mean effects, because few studies reported the amount of time devoted to literacy instruction. In the future, we encourage more primary study authors to provide descriptive information on the quantity and quality of literacy instruction and its relation to student reading outcomes.

Are Summer Reading Interventions Most Effective for Low-Income Children?

The results of this review suggest that summer reading interventions may be particularly effective for low-income children. Previous meta-analytic evidence indicated that summer school had larger effects for children from middle-income than low-income backgrounds (Cooper et al., 2000). Our study, however, did not replicate these earlier results. In our meta-analytic review, the mean effect size was positive and statistically significant in four of five outcomes in studies with a majority of low-income children. In addition, student income characteristics moderated effects on reading comprehension. For the reading comprehension total outcome, the mean effect for low-income samples ($d = .20$) was significantly higher than for mixed-income samples ($d = .00$). For the reading comprehension only outcome, the mean effect for low-income samples ($d = .33$) was also significantly larger than for mixed-income samples ($d = -.05$). Data from seven studies that disaggregated results by student income status were used to replicate the results of the moderator analyses.

These results revealed that mean effects were .28 standard deviations higher for children from low-income backgrounds than for children from mixed-income backgrounds. There may be several potential reasons why our results differ from the results of Cooper et al (2000). Our

review evaluated effects for both classroom and home interventions, focused exclusively on

reading outcome measures, and included only two-group experimental and quasi-experimental

evaluations. It is possible that differences in the intervention setting, the outcome measures, and

the study design of the primary studies that were included in the two meta-analyses yielded

different conclusions about the moderating role of student income status.

        Why do low-income children seem to enjoy the greatest benefit from participating in

summer reading interventions? To address this question, it is important to understand what

happens to low-income children's reading achievement in the summer months in the absence of

an intervention. The results reported in Table 6 indicate that control children in majority low-

income samples made no gains in total reading achievement scores from spring to fall (i.e.,

summer months). These findings provide some clues into the underlying mechanisms driving

income-based disparities in summer reading loss, although the results require replication given

the small sample sizes. Numerous studies indicate that income-based disparities in measurable

aspects of children's home literacy environments may contribute to disparities in reading

achievement. For example, descriptive findings from the National Longitudinal Survey of Youth

(NLSY) suggest that poor families (those meeting the federal definition of poverty) are less

likely than non-poor families to own 10 or more books (Bradley, Corwyn, McAdoo, & Coll,

2001). More precisely, the rich-poor gap in the proportion of children owning 10 or more books

was .57 SD in early childhood (3-5 years) and .25 SD in middle childhood (6-12 years).

Ethnographic research also indicates that low-income parents spend less time discussing books

with their children and have less knowledge about their children's reading interests and levels

than middle-income parents (Chin & Phillips, 2004). Furthermore, cognitive psychologists have

also noted that children need extensive experience reading expository texts to acquire

background knowledge (Geary, 1995; Kintsch, 1994). In the absence of an effective summer

reading intervention, low-income children may have limited opportunities to practice reading

connected text with speed and accuracy and to acquire conceptual and background knowledge.

Limitations and Implications for Future Research

      The results of our review highlight several limitations in the design of summer reading

interventions and the quality of previous evaluation studies. For example, if one goal is to

improve vocabulary outcomes during the summer months, classroom-based interventions should

implement explicit, teacher-directed instruction of high-utility words that enable children to read

proficiently during the school year (Beck & McKeown, 1991; Snow, 2002). It is striking to find

studies that measure reading vocabulary outcomes but provide very little direct vocabulary

instruction in the context of a summer reading intervention. In addition to improving reading

vocabulary, another challenge for researchers and policymakers alike is to sustain short-term

improvement in reading achievement over time. The sensitivity analyses based on within study

comparisons suggest that positive short-term effects diminish over time. Effect sizes measured

six or more months after the conclusion of an intervention were approximately one-third of a

standard deviation smaller than effects measured immediately after an intervention (i.e., less than

one month). Although this finding is based on a small subsample of only seven studies, the fade

out in the magnitude of the treatment effect of summer reading interventions is consistent with

findings from research on other compensatory education interventions (Barnett, 1992).

      Our findings raise questions about the instructional practices that improve reading

outcomes. To open the black box of summer reading interventions, there is a clear need to

identify the variables that mediate improvement in reading outcomes. Tseng and Seidman

(2007) have suggested that better measurement of classroom-level processes might shed light on

the interactions between youth and adults that improve student outcomes. We found emerging

evidence that teachers' use of research-based instructional practices may promote larger gains in

reading comprehension. However, few researchers used direct measures of the quality of

teacher-student interactions in classroom-based summer reading interventions. There are many

advances in theory and measurement of classroom interventions during that school year that

could be applied to summer interventions (Cohen, Raudenbush, & Ball, 2003; Pianta, 2008).

Doing so would illuminate the critical mechanisms inside classrooms—most notably, the quality

of teacher's instructional practice and emotional support for learning—that underlie the observed

improvements in reading achievement during the summer.

In addition to examining the relationship between the quality of classroom instructional

practices and reading outcomes, researchers should examine whether resource intensive summer

programs enhance reading achievement. There was suggestive evidence that the five resource

intensive programs with small class sizes of 13 or fewer children, 4 to 8 hours of daily program

time, and 70 to 175 hours of total program time had a positive effect on reading achievement (d

$= .25$, 95% CI $= .01/.48$). However, the seven studies ($d = .03$, 95% CI $= -.12/.18$) that failed to

meet the criteria being resource intensive had no effect on reading achievement. Caution should

be exercised in interpreting these findings, because the analyses were based on a small number of

studies ($n = 12$). Furthermore, previous meta-analytic reviews have used inconsistent criteria for

determining whether policymakers implemented small class sizes or longer, and more intensive

programs (Cooper et al., 2000; Lauer et al., 2006).[8] Despite these limitations, experimental

studies have shown that the combination of effective instructional practices, reduced class sizes,

and more intensive compensatory education policies are critical to improving the academic

outcomes of low-income children (Krueger & Whitmore, 2001; Ramey & Ramey, 1998, St.

Pierre, Ricciuti, & Rimdzius, 2005). Consistent with findings from prior research, our results suggest that the implementation of research-based instruction and resource intensive programs may enhance effects on student reading outcomes. In future work, researchers should compare the benefits and costs of different summer reading interventions, ranging from resource intensive classroom interventions to potentially less costly home interventions.

More research is also needed to understand how and why student income characteristics moderate the effects of summer reading interventions. One hypothesis emerging from our review is that a summer reading interventions may create a strong treatment-control group contrast among samples with a majority of low-income children. In addressing this hypothesis, researchers might consider the many ways in which parenting practices and family resources shape children's experiences outside school, especially during the summer months. For example, Lareau (2003) has shown that low-income parents promote the accomplishment of their children's natural growth by providing basic needs (e.g., food, shelter, safety) whereas middle-income parents promote the concerted cultivation of their children's talents and skills. Using data from the 2005-06 Consumer Expenditure Surveys, Kornrich, Gauthier, & Furstenberg (2011) found that high-income parents spent $1,373 more, on average, than low-income parents on a variety of educational expenses (e.g., school fees, books). In many ways, then, it seems plausible that the counterfactual situation—i.e., children's literacy experiences in the absence of a summer reading intervention—is substantially different for low- and middle-income. As a result, a summer reading intervention may create a small treatment-control contrast in program activities and outcomes if the majority of children are from middle- and high-income families. However, a summer reading intervention may create a large treatment-control contrast in program activities and outcomes if the majority of children are from low-income families.

Clearly, a direct test of this hypothesis is needed through mixed-methods designs that embed observational measures in an experimental study (Grissmer, Subotnik, & Orland, 2008). In particular, researchers could use observational measures that provide richer descriptions of children's home literacy environment (Chin & Phillips, 2004; Lareau, 2003; Purcell-Gates, 1996), and illuminate the mechanisms driving improvement in low-income children's reading outcomes during the summer months.

In addition, very few interventions were designed to integrate effective elements of both classroom- and home-based summer reading interventions. Although most home interventions do not include a school-based event prior to the summer, it is critical to strengthen the home-school connection (Bronfrenbrenner, 2005). Right before summer vacation, policymakers could implement a school-based family literacy event, in which teachers equip parents and children with skills and knowledge to engage in home literacy activities (Senechal & Young, 2008). To date, however, researchers and policymakers have largely reinforced the notion that classrooms and homes are separate spheres for children's development and distinct settings where summer programs are usually implemented (Cooper et al., 2000; McCombs et al., 2011). The findings of our review, however, suggest the importance of involving both teachers and parents in children's home literacy activities. Toward this end, it would be desirable to test an intervention including classroom teacher-directed comprehension lessons during the last month of school, and home-based literacy activities involving independent book reading and parent-child discussions about books. Parent-child discussions that promote dialogic reading activities, extended discourse about text, and elaborative reminiscing may support oral language, comprehension and vocabulary outcomes (Hart & Risley, 1995; Reese, Sparks, & Leyva, 2010). Although these parent-child activities have been studied in the context of preschool and emergent literacy

interventions, they could be adapted for use in a summer program and with children from a wider range of developmental reading levels.

Future evaluations of summer reading interventions should also include more cost-effectiveness analyses using long-term child outcomes. Limited data on cost-effectiveness constrains the ability of policymakers to invest in interventions that improve student achievement at the lowest per pupil cost. The RAND Corporation (McCombs et al., 2011) recently undertook a comprehensive analysis of the costs of summer programs for classroom-based programs involving teacher-directed instruction of academic skills such as reading and mathematics. According to the RAND report, the per pupil costs of classroom programs ranged from a low of $1,109 to a high of $2,801 depending on whether programs were led by school districts or external community-based organizations. Although it is tempting to conclude that policymakers interested in preventing reading loss among low-income children should invest in home-based summer interventions, there are many outcomes that home interventions are unlikely to improve. The RAND report also suggested that "classroom-based programs may result in additional positive outcomes…such as mathematics achievement and improvements in safety, nutrition, behavioral or social outcomes, or recreational opportunities during the summer" (p. 43). In fact, two of the largest classroom interventions in our review were based on large-scale regression-discontinuity analyses of mandatory summer programs, which showed improvement in both reading and mathematics scores (Jacob & Lefgren, 2004; Matsudaira, 2008). In addition, other summer interventions like the Building Educated Leaders for Life (BELL) are designed to improve children's social skills, academic self-efficacy, and leadership skills (Chaplin & Capizzano, 2006).

Because improvement in non-cognitive skills are important predictors of long-term improvement in students' social and economic outcomes, more rigorous cost-benefit analyses may show that classroom interventions are more likely than home interventions to improve a wide range of cognitive, social, and economic outcomes (Fifer & Krueger, 2006). Unfortunately, no study to date has employed either cost-effectiveness or cost-benefit analyses to show how scarce resources should be allocated to advance diverse societal goals, including efforts to reduce summer loss in reading comprehension, to improve health outcomes during summer, and to improve social and emotional learning and youth leadership skills. The limitations of the current review highlight fruitful areas for additional research.

APPENDIX A
Additional details on codes used to operationalize research-based reading instruction
summarized by the National Reading Panel (NRP)

Domain of reading instruction	Operationalization using NRP-based definitions
Phonemic awareness	This variable refers to studies in which instructors (a) teach students to manipulate phonemes with letters, (b) focus on one or two types of manipulations at a time, (c) teach in small groups.
Phonics	This variable refers to studies in which instructors (a) teach phonics systematically, (b) use the analogy method, (c) use the analytic method, (d) use embedded methods, (e) use the synthetic method.
Fluency	This variable refers to studies in which instructors teach guided repeated oral reading strategies.
Comprehension	This variable refers to studies in which instructors (a) relate readings to students' prior experiences, (b) help students create mental representations, (c) explicitly model strategies for students, (d) teach multiple strategies, (e) teach comprehension monitoring, (f) employ graphic organizers, (g) teach question-generation, (h) teach question-answering, (i) teach story structure, or (j) teach summarizing.
Vocabulary	This variable refers to studies in which instructors employ (a) multiple methods, (b) direct and indirect methods, (c) restructuring, (d) word substitution, (e) graphic organizers, (f) analogies, (g) pictures, or (h) sentence-generation.

Notes

We thank Soojin Oh for assistance in double-coding a random subsample of studies and Syndi Kim and Thomas G. White for commenting on earlier drafts of this article. An earlier version of this article was presented at the Annual Meeting of the Society for Research on Educational Effectiveness, Washington, DC, March 9, 2012.

[1]Since 2000, the recommendations of the National Reading Panel report have been echoed in subsequent reviews of reading instruction. Most recently, the 2010 Institute of Education Sciences (IES; Shanahan et al., 2010) practice guide recommended comprehension strategies that parallel recommendations in the National Reading Panel report, including activating prior knowledge, questioning, visualizing, monitoring and fix-up strategies, drawing inferences, and summarizing and retelling. Both synthesis reports reached convergent recommendations based on studies that used experimental and quasi-experimental designs.

[2]In their evaluation of BELL (Building Educated Leaders for Life), Chaplin and Capizzano (2006) did not report free lunch data but did report that the average family income of participating students was below $30,000; we coded this as a low income sample. Linder (2004) did not report free lunch percentages but reported that the program was offered to Title 1 students. Finally, the three Canadian studies in our meta-analysis (Pagan, 2010; Van Andel, 2011; Seward, 2009) report income data for the students in their samples; our codes for these studies were based on the social class status that the authors ascribed to the reported income level.

[3]Two studies used retell assessments in which children were prompted to summarize a text.

[4]We also conducted two additional analyses based on the extent to which a classroom intervention implemented research-based instruction summarized by the NRP report. First, there

was no evidence that the magnitude of the effects on reading total were significantly different among studies with two or more research-based instructional practices (d = .31, 95% CI = -.20/.83; k = 3), one research-based practice (d = .23, 95% CI = -.12/.57; k = 4), or 0 research-based practices (d = .09, 95% CI = .06, 95% CI = .00/.12, k = 16). Second, we compared only those classroom interventions that used at least one research-based practice with home-based interventions on the reading total outcome, which provides the largest sample size for the analysis. There was no significant difference between the mean effect size for total reading in classroom-based interventions using at least one research-based practice (d = .25, 95% CI = .00/.50; k = 8) and home-based interventions (d = .12, 95% CI = .02/.21; k = 13). Because these results are based on small samples, they require further study and replication.

[5]As noted in table 2, both classroom and home interventions enrolled children with a range of achievement labels. Because the income characteristics of samples may covary with the reading levels of samples, we conducted a follow-up moderator analysis to compare mean effects based on the achievement level targeted. Specifically, we compared mean effects for programs serving all students and programs that specifically served at-risk students, including students who had failed a state reading test, scored below some threshold on national norms, or were designated as below grade level on district reading tests or teacher assessments. The results of this analysis indicated that summer reading interventions were equally effective in improving total reading achievement outcomes for all students (d = .12, 95% CI = .04/.19, k = 25) and for at-risk students (d = .08, 95% CI = .00/.15, k = 16).

[6]We adjusted the standard errors based on an intra-class correlation of .20.

[7]Hill et al. (2007) specifically conducted a synthesis of meta-analyses. The mean effect size from 192 meta-analyses of educational interventions ranged from .20 to .30. The mean effect

size for elementary, middle, and high school grades was .23, .27, and .24, respectively. Given

the consistency of these mean effects, the authors noted that there was "remarkably little

variation in the means across grade levels, despite considerable variation in the interventions and

outcomes represented for the different grades" (p. 176).

[8]For example, Cooper et al. (2000) used a median split of 20 students to compare mean

effects in small and larger class sizes. To compare effects based on program duration, Lauer et

al. (2006) created four categories and found that out-of-school time programs providing 44 to 84

hours ($d = .28$) and 85 to 210 hours ($d = .15$) produced positive effects in reading achievement.

Studies that provide fewer than 43 hours or more than 210 hours did not produce statistically

significant, positive effects.

References

References marked with an asterisk indicate studies included in this meta-analysis.

*Allen, L. C. S. (2003). *At-risk kindergartners: Effects of an intervention program on early literacy acquisition.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3101730)

*Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., & Nowak, (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411-427. doi:10.1080/02702711.2010.505165

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23,* 285–303. doi: http://dx.doi.org/10.1598/RRQ.23.3.2

*Bakle, B. R. (2010). *Summer learning loss: The influence of summer school programs on student achievement in language usage, math, and reading.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3404432).

Barnett, W. S., (1992). Benefits of compensatory pre-school education. *Journal of Human Resources, 27*, 279-312.

Beck, I., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *The handbook of reading research, vol. 2* (pp. 789-815). New York: Longman.

Bloom, H. S., Michaelopoulos, C., Hill, C. J., & Lei, Y. (2002). Can non-experimental comparison group methods match the findings from a random assignment evaluation of

mandatory welfare-to-work pro-grams? New York: Manpower Demonstration Re-search

Corporation.

Bodilly, S., Beckett, M.K. (2005). Making out-of-school-time matter: Evidence for an action

agenda. Santa Monica, CA: RAND Corporation. Retrieved from:

http://www.rand.org/pubs/monographs/MG242. Also available in print form.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to

meta-analysis*. West Sussex, United Kingdom: John Wiley & Sons.

*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*(1), 25-48. doi:10.3102/01623737028001025

*Borman, G. D., Goetz, M. E., & Dowling, N. M. (2009). Halting the summer achievement

slide: A randomized field trial of the KindergARTen summer camp. *Journal of Education for

Students Placed at Risk,14*(2), 133-147. doi:10.1080/10824660802427652

Bradley, R.H., Corwyn, R.F., McAdoo, H.P., & Coll, C.G. (2001). The home environments of

children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child

Development*, *72*, 1844-1867. doi:10.1111/1467-8624.t01-1-00382

Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and

operational models. In S. L. Friedman & T. D. Wachs (Eds.), *Measuring environment across

the life span: Emerging methods and concepts* (pp. 3 − 28). Washington, DC: American

Psychological Association.

Bronfenbrenner, U. (2005). *Making human beings human: Bioecological perspectives on human

development*. Thousand Oaks, CA: Sage Publications.

*Butler, T. (2010). *The effects of summer reading programs on low-income english language

learners' and low-income english L1s' reading achievement in grades 2-4. (Doctoral

dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No.

3407348)

Byrnes, J. P. (2000). Using instructional time effectively. In L. Baker, M. J. Dreher, & J. T.

Guthrie (Eds.), *Engaging young readers* (pp. 188–208). New York, NY: Guilford Press.

*Chaplin, D., & Capizzano, J. (2006). *Impacts of a summer learning program: A random

assignment study of building educated leaders for life (BELL).*Retrieved from:

http://www.urban.org/uploadedPDF/411350_bell_impacts.pdf

Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social

class, children's agency, and the summer activity gap. *Sociology of Education, 77*, 185-210.

doi:10.1177/003804070407700301

*Cleary, T. T. (2001). *Providing phonemic awareness instruction to pre-first graders: An

extended-year kindergarten program.* (Doctoral disseration). Retrieved from ProQuest

Dissertations and Theses database. (UMI No. 3039075).

Cohen, D. K., Raudenbush, S. W. & Ball D. L. (2003). Resources, instruction and research.

Educational Evaluation and Policy Analysis, 25, 119-142. doi:10.3102/01623737025002119

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the Reading Wars: Exploring the

Effect of Child-Instruction Interactions on Growth in Early Reading. *Scientific Studies of

Reading, 8*, 305-336. doi**:**10.1207/s1532799xssr0804_1

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical

examination of the reasons the educational evaluation community has offered for not doing

them. *Educational Evaluation and Policy Analysi*s, 24, 175- 199.

doi:10.3102/01623737024003175

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of

   summer school: A meta-analytic and narrative review. *Monographs of the Society for*

   Research in Child Development, 65(1, Serial No. 260).

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer

   vacation on achievement test scores: A narrative and meta-analytic review. *Review of*

   Educational Research, *66*, 227-268. doi:10.3102/00346543066003227

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical*

   Trials, 7, 177-188. doi: 10.1016/j.cct.2006.04.004

Downey, D.B., von Hippel, P.T., & Broh, B.A. (2004). Are schools the great equalizer?

   Cognitive inequality during the summer months and the school year. *American Sociological*

   Review, *69*, 613-635. doi:10.1177/000312240406900501

Duke, N., & Pearson, D. (2002). Effective practices for developing reading comprehension. In A.

   Farstrup & J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.)

   (pp. 205-242). Newark, DE: International Reading Association.

*Durand, B. A. C. (2002). *The effect of the summer reading academy on the reading*

   achievement of struggling third grade readers. (Doctoral disseration). Retrieved from

   ProQuest Dissertations and Theses database (UMI No. 3042433)

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2010).

   The impact of enhancing students' social and emotional learning: A meta-analysis of

   school-based universal interventions. *Child Development, 82*, 405-432. doi: 10.1111/j.1467-

   8624.2010.01564.x

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing

and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455-463. doi:

10.1111/j.0006-341X.2000.00455.x

*Dwight, L. G. (2010). *Using a summer extended year program to increase learning for title I students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3423319).

*Ellers, S. (2009). *The effects of a standards-based middle-level summer school program as an intervention to increase academic achievement as measured by standards-based assessments.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 33804800)

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview Press.

Fifer, M. & Krueger, A. (2006). *Summer opportunity scholarships: Narrowing the skills gap*. Washington, DC: The Brookings Institution.

Geary, D. D., (1995). Reflections of evolution and culture in children's cognition. *American Psychologist, 50*, 24-37. doi: 10.1037/0003-066X.50.1.24

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6-10. doi:10.1177/074193258600700104

Greenland, S. (1994). Quality scores are useless and potentially misleading (invited commentary). *American Journal of Epidemiology, 140*, 300-301.

Grissmer, D. W., Subotnik, R. F., & Orland, M. (2008). *A guide to the use of Randomized Controlled Trials (RCTs) in assessing intervention effects: The promise of multiple methods*. Retrieved September 8, 2012, from http://www.apa.org/ed/cpse/multmethod08.pdf

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

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

(Eds.), *The voice of evidence in reading research* (pp. 329–354). Baltimore: Paul H. Brookes.

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

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117*, 167-178.doi:10.1037/0033-2909.117.1.167

*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., & 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.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. doi:10.1136/bmj.327.7414.557

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172–177. doi:10.1111/j.1750-8606.2008.00061.x

Institute of Education Sciences. (2010). What Works Clearinghouse procedures and standards workbook, Version 2.1. Retrieved June 1, 2012 from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handb ook.pdf

*Jacob, B.A., Lefgren, L., 2004. Remedial education and student achievement: a regression-discontinuity analysis. *Review of Economics and Statistics* 68, 226–244. doi:10.1162/003465304323023778

*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:10.3102/01623737028004335

*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., & 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

 *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

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294-303. doi:10.1037/0003-066X.49.4.294

Kornrich, S., Gauthier, A. H., & Furstenberg, F. F. (2011). Changes in private investments in children across three liberal welfare states: Australia, Canada, and the United States. 2011 Annual meeting of the Population Association of America.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college test taking and middle school test results: Evidence from Project STAR. *Economic Journal, 111*, 1-28

Lareau, A. (2003). *Unequal childhoods*. Berkeley, CA: University of California Press.

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

*Li, Y., Alfred, C., Kennedy, R., & Putallaz, M. (2009). Effects of summer academic programs in middle school on high school test scores, course-taking, and college major. *Journal of Advanced Academics, Vol. 20, No. 3,* 404-436. doi:10.1177/1932202X0902000303

*Linder, C. A. (2004). *Informing policymakers of the value of summer school in the recoupment of reading skills.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database.(UMI No. 3136432).

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209. doi:10.1037/0003-066X.48.12.1181

*Mariano, L., & Martorell, P. (2011). *The academic effects of summer instruction and retention in New York City.* Santa Monica, CA: RAND Corporation. WR-857.

Marsh, J.A., Gershwin, D., Kirby, S.N., & Xia, N. (2009). *Retaining Students in Grade: Lessons Learned Regarding Policy Design and Implementation.* Santa Monica, CA: RAND Corporation.

*Matsudaira J. (2008) Mandatory summer school and student achievement. *Journal of Econometrics* 142(2): 829–850. doi:10.1016/j.jeconom.2007.05.015

McCombs, J.S., Augustine, C.H., Schwartz, H.L., Bodilly, S.J., McInnis, B., Lichter, D.S., Cross, A.B. (2011). *Making summer count: How summer programs can boost children's learning.* Santa Monica, CA: RAND Corporation.

*Meehan, R. C. (2005). *The impact of an intensive literacy program on the sustained reading achievement of second-grade students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3156499).

*Melosh, G. (2003). *Summer books! stemming reading loss in high-poverty second graders through access to appropriate books during the summer vacation.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3105640).

Mesmer, H. A. E., & Cumming, S. (2009). Text-reader matching: Meeting the needs of struggling readers. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts* (pp. 149-177). New York: Guilford Press.

Mol, S. E. & Bus, A. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267-296. doi:10.1037/a0021890

National Center for Education Statistics. (2011). *The nation's report card: Reading 2011* (NCES 2012-457). Institute of Education Sciences. U.S. Department of Education, Washington, DC.

National Institute for Literacy. (2006). *What content-area teachers should know about adolescent literacy*. Washington, DC: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

*Opalinski, G. (2006). *The effects of a middle school summer school program on the achievement of NCLB identified subgroups.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3224110).

Orwin, R. F. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational*

Statistics, 8, 157-159. doi:10.3102/10769986008002157

*Pagan, S. L. (2010). *Children reading for pleasure: Investigating predictors of reading achievement and the efficacy of a paired-reading intervention to foster children's literacy skills.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (ISBN:978-0-494-70556-8)

*Paris, S. G., Pearson, P. D., Cervetti, G., Carpenter, R., Paris, A. H., DeGroot, J., Bashore-Berg, T. (2004). Assessing the effectiveness of summer reading programs. In G. D. Borman, M. Boulay, G. D. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs.* (pp. 121-161). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*, 4544-4562. doi: 10.1002/sim.2889

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system, K-3*. Baltimore, MD: Paul H. Brookes.

Pressley, M. (2002). *Reading instruction that works (2ⁿᵈ ed.)*. New York: Guilford Press.

Public Agenda. (2010). *A time to learn, a time to grow*. New York: Author.

Purcell-Gates, V. (1996). Stories, coupons, and the TV Guide: Relationships between home literacy experiences and emergent literacy knowledge. *Reading Research Quarterly, 31*, 406-428. doi:10.1598/RRQ.31.4.4

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*, 109-120. doi:10.1037/0003-066X.53.2.109

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Reardon, S. (2011). The widening academic achievement gap between the rich and the poor:

New evidence and possible explanations. In Duncan, G., & Murnane, R. (Ed.) *Whither*

opportunity? New York: Russell Sage.

*Reed, G. (2001). *The relationship between participation in a developmental reading*

summer school program and reading achievement among low-achieving first grade

students. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database.

(UMI No. 3014295).

Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool

children's language and emergent literacy. *Journal of Early Childhood Literacy, 10*, 97-117.

doi:10.1177/1468798409356987

*Roman, S., Carran, D., & Fiore, C. (2010). *The Dominican study: Public library summer*

reading programs close the reading gap. Retrieved from:

http://www.dom.edu/gslis/downloads/DOM_IMLS_book_2010_FINAL_web.pdf

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological*

Bulletin, 86, 638-641. doi:10.1037/0033-2909.86.3.638

*Schacter, J. (2001). *Reducing inequality in elementary reading achievement: Establishing*

summer literacy camps for disadvantaged children. Santa Monica, CA: Milken Family

Foundation. Retrieved from: http://www.mff.org/publications/publications.taf?page=297

*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*(2),

158-169. doi:10.1111/j.1467-9817.2005.00260.x

Senechal, M., & Young, L. (2008). The effect of family literacy interventions on children's

    acquisition of reading from Kindergarten to Grade 3: A meta-analytic review. *Review of*

    Educational Research, 78, 880-907. doi:10.3102/0034654308320319

*Seward, C. (2009). *Evaluating the effectiveness of a short-duration reading intervention*

    on grade one phonological awareness and word reading. (Doctoral dissertation). Retrieved

    from ProQuest Dissertations and Theses database. (ISBN:978-0-494-49971-9)

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental*

    designs for generalized causal inference. Boston: Houghton-Mifflin.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., &

    Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A*

    practice guide (NCEE 2010-4038). Washington, DC: National Center for Education

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

    Education. Retrieved from whatworks.ed.gov/publications/practiceguides

Share, D.L. (1999). Phonological recoding and orthographic learning: A direct test of the self-

    teaching hypothesis. *Journal of Experimental Child Psychology*, *72*, 95–129.

    doi:10.1006/jecp.1998.2481

Sirin, S. R., (2005). Socioeconomic status and academic achievement: A meta-analytic review

    of research. *Review of Educational Research, 75*, 417-453.

    doi:10.3102/00346543075003417

Snow, C. E., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in youth*

    children. Washington, DC: National Reading Council.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program*

    in reading comprehension. Santa Monica, CA: RAND Reading Research Group.

Snow, C. E., & Juel, C. (2007). Teaching children to read: What do we know about how to do

 it? In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 501-521).

 Malden, MA: Blackwell Publishing.

St. Pierre, R. G., Ricciuti, A. E., & Rimdzius, T. A. (2005). Effects of a family literacy program

 on low-literate children and their parents: Findings from an evaluation of the Even Start

 Family Literacy Program. *Developmental Psychology, 41*, 953-970. doi:10.1037/0012-

 1649.41.6.953

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and*

 new frontiers. New York, NY: Guilford Press.

*Sunmonu, K., Larsen, J., Van Horn, Y., Cooper-Martin, E., & Nielsen, J. (2002). Evaluation of

 the extended learning opportunities summer program. Rockville, MD: Montgomery County

 Public Schools. Retrieved from:

 http://www.montgomeryschoolsmd.org/departments/sharedaccountability/reports/2002/elo20

 02.pdf

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading:

 A meta-analysis. *Review of Educational Research, 69,* 261–285.

 doi:10.3102/00346543069003261

Teale, W. H. (1986). Home background and young children's literacy development. In W. H.

 Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 173-206). Norwood,

 NJ: Ablex.

Terzian, M., Moore, K. A., Hamilton, K., Foundation, W., & Trends, C. (2009). *Effective and*

promising summer learning programs and approaches for economically-disadvantaged children and youth: A white paper for the Wallace Foundation. New York: Wallace Foundation.

Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology, 39,* 217–228. doi:10.1007/s10464-007-9101-8

*Ugel, N. (1999). *The effects of a multicomponent reading intervention on the reading achievement of middle school students with reading disabilities.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9959597).

*Van Andel, A. K. (2011). *Summer library reading programs and literacy: An assessment of children's reading progress after having participated in a summer library reading program.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (ISBN:978-0-494-75392-7)

*Waters, B. D. (2004). *The effect of the second grade summer school program on the reading achievement of at-risk second-grade students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3122369)

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30,* 679-689. doi:10.1037/0012-1649.30.5.679

Table 1

Descriptive Characteristics of Studies

General publication characteristics	N	%
Date of report		
1999	1	3
2001	4	11
2002	2	6
2003	2	6
2004	3	9
2005	2	6
2006	4	11
2007	1	3
2008	2	6
2009	5	14
2010	7	20
2011	2	6
Source of report		
Journal article	11	31
Book or book chapter	1	3
Dissertation	17	49
MA thesis	1	3
Private report	4	12
Govt. report	1	3
Program setting		
Urban	14	40
Suburban	4	11
Rural	3	9
Mix	3	9
Unreported	11	31
RCT or regression discontinuity design		
Yes	14	40
No	21	60
Sample characteristics		
Grade level		
Early elementary (K-2)	12	34
Upper elementary (3-5)	8	23
Mix of elementary grades	7	20
Middle school (5-8)	5	14
Elementary and middle grades	3	9
Income characteristics		
Low income	21	60
Mix of income levels	11	31
Unreported	3	9
Home-based Interventions		
Yes	14	34
No	26	63
Mix (home and classroom-based)	1	2

Did study report that the program used reading practices summarized by the NRP?		
Yes	9	35
No	17	65
Outcomes assessed (interventions may assess >1 outcome)		
Total reading achievement	41	100
Reading comprehension total	36	88
Reading comprehension only	18	44
Fluency and decoding	18	44
Vocabulary	12	29
Time elapsed between program end and posttest (studies may have multiple posttest dates)		
One month	49	56
Two months	16	18
3-10 months	16	18
>10 months	4	5
Unreported	2	2

Note: Study-level characteristics reported at the study level ($N = 35$); Intervention-level characteristics reported at the intervention level ($k = 41$). Percentages may not sum to 100 due to rounding. NRP = National Reading Panel.

Table 2

Descriptive Characteristics for Each of the 41 Summer Reading Interventions

Author(s), publication year, and source of report	Grade level	Sample Size Treatment *n* (Total sample *n)*	Achievement Level Targeted	Income Status	Outcome Categories	RCT or RD design	Total Reading Effect Size (*d*) 95% CI
Classroom interventions							
Allen (2003), Dissertation	K	18 (30)	"At risk," as identified by PALS assessment and teacher recommendation	Low income	Reading comp only Reading comp total Reading Mixed	N	0.099 (-0.707 to 0.904)
Bakle (2010), Dissertation	Grades 1-4	860 (1720)	All students	Mixed Income	Reading comp total	N	-0.112 (-0.303 to 0.078)
Borman & Dowling (2006), Journal Article	Grades 2-3	420 (658)	All students (from low-income urban community)	Low income	Vocabulary Reading comp only Reading comp total	Y (RCT)	0.069 (-0.150 to 0.287)
Borman et al. (2009), Journal Article	K	73 (100)	All students (from high-poverty urban schools)	Low income	Reading comp total Fluency & decoding	Y (RCT)	0.208 (-0.247 to 0.663)
Chaplin & Capizzano (2006), Report	Grades 1-7	417 (835)	All students (from "underserved communities"; low-achievers targeted)	Low income	Vocabulary Reading comp only Reading comp total	Y (RCT)	0.065 (-0.071 to 0.201)
Cleary (2001), Dissertation	K	80 (124)	Students with "literacy and phonemic awareness deficits," as measured by district assessments	Unreported	Fluency & decoding Reading comp total	N	0.116 (-0.407 to 0.639)
Durand (2002), Dissertation	Grade 3	88 (215)	Students scoring below 70% on Texas Assessment of Academic Skills	Mixed Income	Vocabulary Reading comp only Reading comp total	N	-0.013 (-0.285 to 0.258)
Dwight (2010), Dissertation	Grades 3 and 4	36 (54)	"At risk Title 1 students" scoring below proficiency on local assessments	Low income	Reading comp total	N	0.331 (-0.243 to 0.905)
Ellers (2009), Dissertation	Grades 7 and 8	60 (120)	"Low achieving and at-risk students"	Low income	Reading comp total	N	-0.088 (-0.446 to 0.270)
Haymon (2009), Dissertation	Grade 7	30 (60)	Available to all students; over 70% of participants were below proficient on state standardized test	Low income	Reading comp only Reading comp total	N	0.404 (-0.125 to 0.933)

Study	Grade	N	Students	Income	Outcome	Exp	Effect size
Jacob & Lefgren (2001), Journal Article	Grades 3 and 6	12,175 (total sample)	Students who had failed state test	Low income	Reading comp total	Y (RD)	0.023 (-0.008 to 0.054)
Li et al. (2009), Journal Article	Middle School	141 (2790)	"Academically gifted and talented"	Unreported	Reading comp total	N	0.257 (0.087 to 0.426)
Linder (2004), Dissertation	Grade 1	47 (83)	Students scoring below 25th percentile on district assessment	Low income	Fluency & decoding	N	-0.038 (-0.472 to 0.396)
Mariano & Martorell (2011), Report	Grade 5	36,481 (total sample)	Students who had failed state test	Low income	Reading comp total	Y (RD)	0.042 (-0.026 to 0.11)
Matsudaira (2008), Journal Article	Grades 3-5, 7	199,164 (total sample)	Students who had failed state test	Low income	Reading comp total	Y (RD)	0.157 (0.008 to 0.307)
Meehan (2005), Journal Article	Grade 2	41 (57)	Students more than 6 months below grade level, as measured by running record	Unreported	Reading comp total	N	0.114 (-0.705 to 0.932)
Opalinkski (2006), Dissertation	Grade 8	167 (331)	Students in danger of retention due to failing grades	Mixed Income	Reading comp total	N	-0.197 (-0.415 to 0.020)
Paris et al. (2004), Book chapter	Grades K-3	319 (551)	Varying criteria - "generally considered at risk," based on teacher recommendation	Mixed Income	Fluency & decoding Vocabulary Reading comp only Reading comp total	N	-0.078 (-0.450 to 0.295)
Reed (2001), Dissertation	Grade 1	30 (74)	Students scoring below 50th percentile on reading Terra Nova	Mixed Income	Reading comp only Reading comp total	N	-0.144 (-0.609 to 0.321)
Schacter (2001), Report	Grade 1	21 (51)	Students from schools "whose reading scores were below the 25th percentile"	Low income	Vocabulary Reading comp only Reading comp total Fluency & decoding	N	0.33 (-0.231 to 0.892)
Schacter & Jo (2005), Journal Article	Grade 1	54 (118)	All students (from "economically disadvantaged" backgrounds)	Low income	Reading comp only Reading comp total	Y (RCT)	0.696 (0.313 to 1.080)
Seward (2009), Dissertation	K	55 (86)	"Children showing signs of early reading delay," as identified by teacher	Mixed Income	Fluency & decoding Vocabulary	N	0.472 (0.024 to 0.921)

Seward (2009), Dissertation	K	53 (84)	Children showing signs of early reading delay, as identified by teacher	Mixed Income	Fluency & decoding Vocabulary	N	0.475 (0.025 to 0.925)
Sunmonu et al. (2002), Report	Grades K-2	1472 (2402)	All students (in Title 1 schools with high concentrations of FRL and ELL students)	Low income	Fluency & decoding	N	-0.077 (-0.197 to 0.043)
Ugel (1999), Dissertation	Grades 5-7	24 (48)	Students with reading disabilities	Low income	Fluency & decoding Reading comp only Reading comp total	N	0.821 (0.221 to 1.421)
Waters (2004), Dissertation	Grade 1	23 (46)	Students with scores of 24 or below on the DRA	Mixed Income	Reading comp total	N	0.804 (0.195 to 1.413)
Home interventions							
Allington et al. (2010), Journal Article	Elementary	852 (1330)	All students (from high-poverty schools)	Low income	Reading comp total	Y (RCT)	0.139 (0.027 to 0.251)
Butler (2010), Dissertation	Grades 2-4	26 (67)	Below grade level	Low income	Fluency & decoding Reading comp only Reading comp total	N	0.708 (0.202 to 1.214)
Butler (2010), Dissertation	Grades 2-4	27 (68)	Below grade level	Low income	Fluency & decoding Reading comp only Reading comp total	N	0.621 (0.125 to 1.118)
Kim & Guryan (2010), Journal Article	Grade 4	110 (216)	All students (from high-poverty, language minority schools)	Low income	Reading comp total Reading comp only Vocabulary	Y (RCT)	-0.143 (-0.414 to 0.129)
Kim & Guryan (2010), Journal Article	Grade 4	103 (211)	All students (from high-poverty, language minority schools)	Low income	Reading comp total Reading comp only Vocabulary	Y (RCT)	-0.021 (-0.323 to 0.281)
Kim & White (2008), Journal Article	Grades 3-5	93 (200)	All students	Mixed income	Fluency & decoding Reading comp total	Y (RCT)	-0.115 (-0.403 to 0.174)
Kim & White (2008), Journal Article	Grades 3-5	100 (207)	All students	Mixed income	Fluency & decoding Reading comp total	Y (RCT)	0.083 (-0.200 to 0.365)
Kim & White (2008), Journal Article	Grades 3-5	100 (207)	All students	Mixed income	Fluency & decoding Reading comp total	Y (RCT)	0.089 (-0.194 to 0.371)
Kim (2006), Journal Article	Grade 4	252 (486)	All students	Mixed Income	Fluency & decoding Reading comp total	Y (RCT)	0.083 (-0.095 to 0.261)

Study	Grade	N	Population	Income	Outcome	Design	Effect size
Kim (2007), Journal Article	Grades 1-5	166 (331)	All students	Mixed Income	Reading comp total	Y (RCT)	0.185 (-0.051 to 0.420)
Melosh (2003), Dissertation	Grade 2	26 (46)	All students (from high-poverty schools; 2/3 were scoring below proficient on state tests)	Low income	Fluency & decoding Reading comp only Reading comp total	Y (RCT)	0.108 (-0.505 to 0.721)
Pagan (2010), Dissertation	Grades 3 and 5	28 (57)	Students scoring below age expectations on measure of expressive vocabulary	Mixed income	Vocabulary Reading comp only Reading comp total Fluency & decoding	Y (RCT)	0.11 (-0.410 to 0.630)
Seward (2009), Dissertation	K	27 (58)	Children showing signs of early reading delay, as identified by teacher	Mixed income	Fluency & decoding Vocabulary	N	0.553 (0.025 to 1.081)
Van Andel (2011), Dissertation	Grades 2-5	16 (69)	All students (from low SES schools)	Low income	Fluency & decoding Vocabulary Reading comp only Reading comp total	N	0.05 (-0.511 to 0.611)

Mix of intervention contexts

Study	Grade	N	Population	Income	Outcome	Design	Effect size
Roman (2010), Report	Elementary	149 (total sample)	All students (from 50% FRL schools, no more than 15% ELL)	Low income	Reading comp only	N	0.226 (-0.097 to 0.548)

Note. A total of 35 studies contributed descriptive information for 41 interventions. Studies reporting the effects of multiple interventions are listed more than once. Multiple intervention effects were reported by four studies, including Seward (2009), Butler (2010), Kim & Guryan (2010), and Kim & White (2008). In the meta-analysis of classroom and home interventions, independence of observations was maintained by using a single intervention effect size from each study. The data source for each cited study is reported in the reference list. Effective sample sizes for each effect size may differ depending on the information reported in the study. All interventions contribute to the total reading achievement effect size. Some interventions may contribute additional assessments to the total reading achievement effect size not listed here. RCT = randomized controlled trial; RD = regression-discontinuity design; Reading comp = Reading comprehension.

Table 3

Mean Effect Size (ES), 95% Confidence Intervals (CI), and Homogeneity Statistics for the Total Sample, and for Classroom and Home Interventions

Moderator variable		Total reading achievement	Reading comprehension total	Reading comprehension only	Fluency and decoding combined	Reading vocabulary
All interventions combined	ES	0.10*	0.13*	0.23*	0.24*	0.04
	CI	0.04 to 0.15	0.06 to 0.19	0.07 to 0.40	0.08 to 0.40	-0.04 to 0.12
	k	41	36	18	18	12
	Q total	82.44***	101.7***	56.38***	49.09***	6.92
	I^2	51.50%	65.60%	69.80%	65.40%	0.00%
Classroom Interventions	ES	0.09*	0.13*	0.25*	0.22	0.06
	CI	0.02 to 0.15	0.05 to 0.22	0.02 to 0.49	-0.03 to 0.46	-0.04 to 0.16
	k	26	22	10	9	7
	Q within	57.78***	79.04***	41.88**	27.38**	1.65
	I^2 within	56.70%	73.40%	78.50%	70.80%	0.00%
Home Interventions	ES	0.12*	0.11*	0.22	0.26*	-0.02
	CI	0.02 to 0.21	0.01 to 0.21	-0.03 to 0.48	0.04 to 0.47	-0.2 to 0.17
	k	14	13	7	9	5
	Q within	19.34	18.36	14.19*	18.3*	4.45
	I^2 within	32.80%	34.60%	57.70%	56.30%	10.20%

*$p < .05$, **$p < .01$, ***$p < .001$

Table 4

Findings for Research-based Instruction Moderator Analyses for Classroom-based Summer Reading Interventions

Moderator variable		Total reading achievement	Reading comprehension total	Reading comprehension only	Fluency and decoding combined	Reading vocabulary
Research-based instruction (yes)	ES	0.25	0.38*	0.53	0.63*	0.04
	CI	-0.01 to 0.50	0.05 to 0.71	-0.03 to 1.09	0.04 to 1.21	-0.09 to 0.18
	k	9	9	5	1	1
	Q within	29.08***	51.97***	37.02***	0	0
	I^2 *within*	72.50%	84.60%	89.20%	n/a	n/a
Research-based instruction (no)	ES	0.06	0.08*	0.05	0.18	0.08
	CI	-0.00 to 0.12	0.01 to 0.15	-0.09 to 0.20	-0.08 to 0.43	-0.06 to 0.21
	k	17	13	5	8	6
	Q within	27.37*	24.51*	2.93	23.38**	1.54
	I^2 *within*	41.50%	51.00%	0.00%	70.1%	0%

*p < .05, **p < .01, ***p < .001

Table 5

Findings for Student Income-based Moderator Analyses

Moderator variable		Total reading achievement	Reading comprehension total	Reading comprehension only	Fluency and decoding combined	Reading vocabulary
Majority low-income sample	ES	0.10*	0.20*	0.33*	0.23*	0.02
	CI	0.04 to 0.17	0.11 to 0.29	0.14 to 0.53	0.01 to 0.46	-0.08 to 0.11
	k	23	21	14	9	6
	Q within	48.28**	75.26***	50.54***	19.77*	4.85
	I^2 within	54.40%	73.40%	74.30%	59.50%	0%
Mixed income sample	ES	0.08	0.00	-0.05	0.27*	0.10
	CI	-0.04 to 0.19	-0.11 to 0.10	-0.23 to 0.14	0.00 to 0.54	-0.06 to 0.26
	k	15	12	4	8	6
	Q within	28.02*	18.05	1.08	25.72**	1.35
	I^2 within	50.00%	39.10%	0%	72.80%	0%

*$p < .05$, **$p < .01$, ***$p < .001$

Table 6

Findings for Income Status Moderators of Reading Achievement from Spring to Fall for Control Group Children

Moderator variable		Total reading achievement	Reading comprehension total	Reading comprehension only
Majority low-income sample	ES	-0.05	-0.08	-0.10
	CI	-0.22 to 0.12	-0.22 to 0.06	-0.26 to 0.05
	N	10	8	6
	Q within	13	5.46	4.89
	I^2 within	30.70%	0%	0%
Mixed-income sample				
	ES	0.26*	0.20	0.32
	CI	0.07 to 0.45	-0.06 to 0.45	-0.38 to 1.03
	N	6	5	2
	Q within	7.87	11.03*	5.42*
	I^2 within	37%	64%	82%

*p < .05, **p < .01, ***p < .001