The Effectiveness of Out-of-School-Time Strategies in Assisting Low-Achieving Students in Reading and Mathematics: *A Research Synthesis*

Regional Educational Laboratory
Contract #ED-01-CO-0006
Deliverable #2003-19

Prepared for
Institute of Education Sciences
Department of Education
Washington, D.C.

Prepared by
Patricia A. Lauer
Motoko Akiba
Stephanie B. Wilkerson
Helen S. Apthorp
David Snow
Mya Martin-Glenn

Updated January 2004



©2004 MCREL

Acknowledgements

The authors wish to thank several individuals at McREL who contributed to this research synthesis. Thanks go to Zoe Barley who provided helpful advice and review comments. We appreciate the efforts of Rebecca Van Buhler who conducted literature searches and examined research articles and of Terry Young who obtained the articles. We acknowledge Barbara Aiduk, Barbara Gaddy, and Robyn Alsop for their editing and logistical support in preparing the document. Finally, we also thank our external reviewers for their invaluable input. These included Harris Cooper of Duke University, Sarah Lubiensky of Iowa State University, and Patricia Mathes of Southern Methodist University.

This document has been funded at least in part with federal funds from the U.S. Department of Education under contract number ED-01-CO-0006. The content of this publication does not necessarily reflect the views or policies of the Department of Education nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. Government.

Table of Contents

Executive Summary 1

Preface 5

Chapter 1

Background and Methods 7
- Background 7
- Methodology 13
- Analyses and Results 21
- Overview of Synthesis 24

Chapter 2

Studies of the Effectiveness of Out-of-School-Time Strategies for Reading Achievement 25
- Background 26
- Methodology 27
- Overview of Studies 30
- Meta-Analysis Results 36
- Narrative Review of Studies 42
- Common Features Highlighted in Studies 44
- Discussion of Findings 48
- Implications for Policy and Practice 50

Chapter 3

Studies of the Effectiveness of Out-of-School-Time Strategies for Mathematics Achievement 51
- Background 52
- Methodology 53
- Overview of Studies 55
- Meta-Analysis Results 59
- Narrative Review of Studies 65
- Common Features Highlighted in Studies 66
- Discussion of Findings 70
- Implications for Policy and Practice 72

Chapter 4

Summary and Conclusions 73
- Summary and Interpretation of Results 73
- Research Issues 76
- Conclusions and Implications 76

References 79

Appendix A: Coding Instrument 91
Appendix B: Meta-analysis Methods 103
Appendix C: Annotated Bibliography 109

Executive Summary

The No Child Left Behind (NCLB) Act of 2001 requires states to ensure that all students achieve proficiency in reading and mathematics. States must provide supplementary education services to low-income students in Title 1 schools that do not achieve adequate yearly progress toward this goal. Because the instruction for supplementary services must occur outside the regular school day, there is interest among educators in the effectiveness of out-of-school-time (OST) strategies for improving student achievement. Thus, the current synthesis addresses the following research problem: Based on rigorous research and evaluation studies, what is the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading and mathematics?

OST programs vary greatly in their goals and characteristics, and the research on OST has been equally varied. Although some prior reviews of research on after-school programs and summer schools have been conducted, none has systematically examined outcomes in relationship to methodological rigor and content area. To address this need, the current synthesis reviews only studies that used comparison or control groups to reach conclusions, and it provides separate analyses of OST strategies for student achievement in reading and in mathematics.

An exhaustive literature search was conducted to identify both published and unpublished research and evaluation studies conducted after 1984 that addressed the effectiveness of a program, practice, or strategy delivered outside the regular school day for low-achieving or at-risk K–12 students. The search resulted in 1,808 citations, from which 371 reports were obtained. Among the criteria for synthesis inclusion were that studies had to measure student achievement in reading and/or mathematics and employ control/comparison groups. Fifty-three studies met the inclusion criteria, 47 with reading outcomes and 33 with mathematics outcomes. Of the 53 studies, 27 addressed outcomes in both subject areas.

Researchers used a coding instrument to describe the following for each study: characteristics of the OST strategy and the students it addressed, research design and methods, data analyses and findings, and research quality. The latter concerned the degree to which studies had four types of validity: construct, internal, external, and statistical. To produce consistency among judgments, researchers trained on the use of the coding instrument and used procedures for double checking their coding results.

The studies were analyzed through meta-analyses and supplemented by narrative descriptions. Results were further analyzed for the influence of moderators on the effectiveness of OST strategies. Program moderators included timeframe (after school or summer school), grade level of the students, focus of the OST activities (academic or academic plus social), duration of the OST program, and grouping of students (large or small groups or one-on-one tutoring). Study moderators included research quality (high, medium, or low), publication type (conference paper, dissertation, or peer-reviewed journal article), and score type (gain score or posttest score).

The synthesis resulted in statistically significant positive effects of OST on both reading and mathematics student achievement. The overall effect sizes ranged from .06 to .13 for reading and from .09 to .17 for mathematics, depending on the statistical model used for meta-analysis. Though numerically small, these results are important because they are based on strategies to supplement the regular school day and to prevent learning loss. Positive findings for supplementary programs that address the needs of low-achieving or at-risk students are therefore encouraging. Together, the results for reading and mathematics suggest that OST programs can significantly increase the achievement of these students by an average of one-tenth of a standard deviation compared to those students who do not participate in OST programs.

With regard to moderators of effectiveness, the timeframe for delivery of OST strategies did not have a statistically significant influence. Grade level was a statistically significant moderator of effect sizes for both reading and mathematics outcomes. For reading, the largest positive effect size (.26) occurred for students in the lower elementary grades (K–2), while for mathematics the largest positive effect size (.44) was for students in high school (9–12). For reading outcomes, activity focus was not a statistically significant moderator of effect size, while for mathematics outcomes, strategies that were both academic and social had a slightly higher mean effect size than those that were mainly academic. For both reading and mathematics, effect sizes were larger for OST programs that were more than 45 hours in duration, but the programs with the longest durations (more than 210 hours for reading and more than 100 hours for mathematics) had effect sizes that were not significantly different from zero.

Only the reading studies had sufficient information to analyze the statistical influence of the way in which students were grouped in OST programs. The largest positive effect (.50) occurred for the reading studies that used one-on-one tutoring. Thus, the moderator results suggest that certain program features can result in higher positive effects of OST on student achievement.

Most of the studies reviewed were rated as medium in research quality because they did not adequately describe the OST intervention or its implementation. For mathematics, there was a statistically significant result in favor of higher quality studies, but quality ratings did not significantly influence the effect size for reading. Type of publication was a statistically significant moderator of effectiveness of OST for reading achievement but not for mathematics. The effect size for reading studies reported in peer-reviewed journals was larger than for unpublished reports and dissertations. The type of score had a significant influence on the effect sizes for mathematics but not for reading. For mathematics outcomes, the average effect size for gain scores was significantly greater than zero, while this was not true for the average effect size based on posttest scores.

In addition to the analyses of study outcomes, the syntheses of reading and mathematics studies described some common features among the studies in each content area. In reading, these were the links between student attendance and student achievement, the importance of staff quality, the development of academic and social skills, the implementation of a well-defined reading curriculum, and the prevention of learning loss. Common features highlighted in the mathematics studies were additional time for remediation, the use of tutoring, the use of counseling and mentoring, and the combination of recreation with mathematics instruction.

Overall, the meta-analytic and narrative results lead to the following conclusions and implications for practice and policy related to OST and its evaluation:

- OST strategies can have positive effects on the achievement of low-achieving or at-risk students in reading and mathematics.
- The timeframes for delivering OST programs (i.e., after school or summer school) do not influence the effectiveness of OST strategies.
- Students in early elementary grades are more likely than older elementary and middle school students to benefit from OST strategies for improving reading, while there are indications that the opposite is true for mathematics.
- OST strategies need not focus solely on academic activities to have positive effects on student achievement.
- Administrators of OST programs should monitor program implementation and student learning in order to determine the appropriate investment of time for specific OST strategies and activities.
- OST strategies that provide one-on-one tutoring for low-achieving or at-risk students have strong positive effects on student achievement in reading.

- Research syntheses of OST programs should examine both published and unpublished research and evaluation reports.

- Future research and evaluation studies should document the characteristics of OST strategies and their implementation.

Preface

Although there have been after-school and summer school programs for school-age children for many years, the No Child Left Behind (NCLB) Act of 2001 has focused new attention on children's out-of-school-time (OST) activities. Children in schools that fail to help all children reach proficiency are eligible to receive supplementary education services. These services must occur outside the school day and be backed by evidence that the services are effective in raising student achievement. Thus, NCLB gives new emphasis to the use of OST strategies for improving academic achievement and stresses the need to examine evaluation results for these strategies. Our study responds to this need through a review and synthesis of research on the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading and mathematics, the content areas emphasized by NCLB.

This report is the third annual research synthesis that Mid-continent Research for Education and Learning (McREL), a Regional Educational Laboratory, has conducted in its laboratory leadership area of standards-based educational practice. In 2001, McREL published a synthesis of research on standards-based classrooms (Apthorp et al.). That report used narrative reviews to examine research on standards-based instruction in literacy and mathematics and on the practices and policies needed for professional development and school organizations in a standards-based education system. In 2002, McREL conducted a research synthesis on the effectiveness of strategies designed to assist low-achieving or at-risk students during the school day so that all students can ultimately achieve standards (Barley et al.). The 2002 synthesis provided reviews of research on six classroom strategies: general instruction, cognitively oriented instruction, grouping structures, tutoring, peer tutoring, and computer-assisted instruction. Findings were described in relationship to both research outcomes and the research quality of the studies.

This year's synthesis complements the previous year's work through a review of research on strategies to assist low-achieving or at-risk students outside the school day — OST strategies. Due to the range in goals and outcomes of OST strategies and based on NCLB's emphasis on reading and mathematics, we limited our synthesis to research on reading and mathematics outcomes. In keeping with our emphasis and that of NCLB's on research quality, we again examined findings in relationship to quality criteria.

The goals for the current research synthesis are the following:

1. To identify effective OST strategies in assisting low-achieving or at-risk students in reading and mathematics based on a collection of research and evaluation studies gathered through an exhaustive search process

2. To assess the effectiveness of OST strategies and the influences of strategy and study characteristics using meta-analytic techniques and narrative reviews

3. To describe study findings in relation to the quality of the research

4. To describe the implications of the findings for researchers and policymakers

This synthesis is organized into four chapters. Chapter 1 describes the research problem, provides background information on OST strategies used to improve academic achievement, and describes the methods used to search the literature, code studies, and synthesize results. Chapters 2 and 3 review research on the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading and mathematics, respectively. Chapter 4 summarizes the findings across reading and mathematics and provides general conclusions. Appendices include the instrument used to code studies, a description of the meta-analysis methods, and an annotated bibliography of selected references.

The authors of this document worked as a team to conduct the synthesis and produce the report. They made individual contributions based on their areas of expertise. Patricia Lauer was the author of chapters 1 and 4 and led the synthesis team. Stephanie Wilkerson and Helen Apthorp wrote chapter 2, and Motoko Akiba and David Snow wrote chapter 3. Motoko Akiba also conducted the meta-analyses for the synthesis. Mya Martin-Glenn directed the search for and documentation of synthesis research studies.

The primary audience for this document includes education researchers and state education administrators who have a general understanding of scientifically based evidence. The secondary audience includes policymakers and district and school administrators who have some background in research. Although this document is not intended for practitioners, the findings reported inform education practice.

CHAPTER 1

Background and Methods

States and districts are experiencing pressure to ensure that all students achieve proficiency on standards-based achievement tests in reading and mathematics. The No Child Left Behind (NCLB) Act of 2001 requires states to ensure that children reach high standards of learning so that all students will be proficient after 12 years. Low-income students in Title 1 schools that do not achieve adequate yearly progress toward this goal for three or more years are eligible to receive supplementary educational services. The instruction for these services must occur outside the regular school day, and states must approve providers of supplementary services based on their evidence of effectiveness in raising student achievement.

Thus, according to NCLB, children's out-of-school-time (OST) activities, such as after-school programs and summer school instruction, can be used for delivering supplementary education services when schools do not adequately fulfill their responsibilities to students. Though some educators question whether this is a developmentally appropriate solution for improving children's learning (Halpern, 1999, 2000), others question the effectiveness of OST strategies in raising student achievement. As we and other researchers have found, programs that use OST strategies abound, but many evaluations of such programs are not methodologically rigorous (Scott-Little, Hamann, & Jurs, 2002). Thus, we conducted this synthesis to address the following research problem: Based on rigorous research and evaluation studies, what is the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading and mathematics?

BACKGROUND

OST refers to the hours in which school-age children are not in school (National Institute on Out-of-School Time, 2003). OST does not imply a specific time, schedule, or duration, but it does mean that during those hours, children are doing something other than activities mandated by school attendance. Researchers have discussed OST with reference to the *timeframes* in which OST programs are

delivered, the most common of which are after-school programs and summer schools.[1]

According to De Kanter (2001), six million of the 54 million K–8 children in the United States participate in after-school programs that are school based or community sponsored. De Kanter reported that since 1994, the number of schools that offer programs after school has doubled, but according to the National Institute on Out-of-School Time (2003), there are still eight million children between the ages of 5 and 14 who are unsupervised after school on a regular basis. De Kanter and other advocates for after-school programs (The After-School Corporation, 1999; Fashola, 2002) have cited increasing public support for the development and funding of after-school programs in public schools.

Halpern (2002) traced the origins of after-school programs to societal concerns for the safety and care of children who live in unsafe neighborhoods and to the need for childcare due to the growth in maternal employment. Halpern noted that only recently have policymakers suggested after-school programs as ways to improve student achievement, a policy that Halpern opposes due to its interference with developmental play. According to Kugler (2001), three societal concerns have contributed to the recent growth in after-school programs: the lack of caregivers in the home after school, the belief that disadvantaged children can improve their learning given more time and opportunities, and the high incidence of teen crime after school. Similarly, The After-School Corporation (1999) cited statistics to suggest that after-school programs are needed to prevent maladaptive behaviors by children, such as crime and drug abuse. Fashola (2002) added that after-school programs are needed to provide enriching experiences that can improve children's socialization.

Thus, after-school programs have a long history, and the conditions that shape their development reflect societal concerns regarding child development. Because these concerns compete for focus, after-school programs vary widely in goals and practices, making it difficult to assess their impacts as interventions. Adding to this complexity is the need for after-school programs to be developmentally appropriate and attractive to participants. Proponents of after-school programs have emphasized that older children and youth, as well as children in early elementary school, need adult supervision and access to enrichment activities. Because it is more difficult to recruit older children than younger children to after-school programs, implementers

[1] Extended-day programs are after-school programs that are connected to a specific school (Fashola, 2002).

have devised creative programming strategies (Grossman, Walker, & Raley, 2001), a result that has contributed to the variation in content among after-school programs.

A report by Cooper, Charlton, Valentine, and Muhlenbruck (2000) described the history and goals of summer school. Similar to after-school programs, the original reason for summer schools was the prevention of behavior problems. In the 1950s, the view emerged among educators that summer school could address students' learning deficits through remedial activities. Cooper et al. cited Title 1 of the Elementary and Secondary Education Act (ESEA) of 1965 as an early federal initiative for the delivery of supplemental education help to low-income students in the form of extended time. As a result, Title 1 funds have been used to fund summer schools. In more recent years, summer schools also have provided enrichment activities and opportunities for students to graduate early. The authors cited the following societal factors influencing the push to create summer school programs: family influences, such as maternal employment and single parent households; the need for the United States to maintain a globally competitive education system; and the emphasis on high learning standards and minimum student proficiency requirements. Cooper et al. noted, "Although additional purposes for summer school will emerge, the primary focus is likely to remain academic" (p. 8). Thus, compared to after-school programs, summer school programs tend to be more oriented toward academic improvement and less oriented toward multiple goals.

Historically, the needs of low-income children have been a major influence on the development of OST programs. Because their neighborhoods tend to be less safe than those of middle-income children, there is a greater need for their OST to be structured by adults. In addition, there is less likely to be an after-school caregiver in the homes of low-income children. Title 1 of the ESEA was created in part because of data indicating that low-income children are at risk for academic failure and therefore need additional time in education activities to supplement what they experience during regular school hours (Cooper et al., 2000; Borman & D'Agostino, 1996). Researchers of after-school programs also have indicated that compared to middle-income children, low-income children are more in need of after-school opportunities and more likely to benefit from them (Miller, 2003; Cosden, Morrison, Albanese, & Macias, 2001).[2] The histories of after-school programs and summer schools suggest that the current emphasis on OST is due to the perceived failure of societal institutions, particularly the family and the school, to fulfill their responsibilities to all children. This research synthesis examines the effectiveness of OST strategies in assuming some of the responsibilities of schools.

---

[2] However, Cooper et al. (2000) found that both middle-income and low-income students benefited from summer school, but the effect was greater for the former.

Research Context

Although in recent years research and evaluation of OST have increased dramatically, as a whole the studies tend to be as varied as OST strategies, particularly with respect to after-school programs (Scott-Little et al., 2002). As described previously, improved student achievement is only one of the goals of OST strategies. Furthermore, many of the studies that address student achievement have not disaggregated outcomes by subject area. This is problematic because, for example, if students' GPAs increase as a result of an after-school program, the increase might be due to higher grades in non-core subjects, such as physical education or art. Though non-core subject areas make important contributions to children's education and development, reading and mathematics are the main concerns of current policymakers and school administrators.

Another element of the current research context that influences this research synthesis is the emphasis on what is referred to as scientifically based research. As supported by the U.S. Department of Education and defined by NCLB, scientifically based research is research that is systematic, rigorous, objective, empirical, appropriate for peer-reviewed journal publication, and relies on multiple reliable and valid measurements and observations, preferably through experimental or quasi-experimental methods. In general, reviews of OST have not based conclusions on the methodological quality of studies. As described in the next section, studies were screened for inclusion in the current synthesis based on the degree to which methods approximated those of rigorous research, and synthesis results were examined in relationship to research quality.

Prior reviews related to OST strategies informed this synthesis. Cooper et al. (2000) reported on a comprehensive synthesis of summer school research using both meta-analysis and narrative review. The results indicated positive academic effects of summer school for both middle-income and low-income students. In addition, results favored programs run for smaller numbers of students and those that provided more individualized and small-group instruction to students. Also, students in the early elementary grades and secondary grades benefited more from summer school compared to students in late elementary grades. The current synthesis adds to Cooper et al.'s findings by examining summer school effects in relationship to other types of OST strategies, primarily after-school programs.

McComb and Scott-Little (2003) provided a narrative review of 27 studies of after-school programs. The authors concluded that large variations in program content, size, goals, and research designs prevented a simple answer to the question of the effects of after-school programs on academic outcomes. Instead, McComb and Scott-Little emphasized the conditions that favored positive outcomes. For example, there

were indications that low-achieving students benefited more than did students who entered programs with higher achievement, and that students who attended the programs more frequently benefited more. Overall, the results of this review were inconclusive about the effects of after-school programs on academic achievement. In addition, the review did not examine in depth the influences of content area or participant grade level as the current synthesis does.

Fashola (1998) reviewed evaluations of 34 programs delivered in extended-day or after-school formats. Fashola concluded that with regard to academic after-school programs for elementary and secondary students, the research has been limited:

> We find that there are a number of promising models in existence, many of which have encouraging but methodologically flawed evidence of effectiveness. Among programs intended to increase academic achievement, those that provide greater structure, a stronger link to the school-day curriculum, well-qualified and trained staff, and opportunities for one-to-one tutoring seem particularly promising, but these conclusions depend more on inferences from other research than from well-designed studies of the after-school programs themselves. (p. 55)

Fashola's report provided guidelines for implementing effective after-school programs based on the "rudimentary stage" (p. 54) of the research at that time. The current synthesis adds to this knowledge base by including more studies and more systematic examination of the methodological quality of studies and the influence of student grade level.

A report by Redd, Cochran, Hair, and Moore (2002) examined studies of 12 academic-oriented programs for adolescents, half of which the authors classified as experimental studies and half as quasi-experimental. Most of the programs were delivered after school. The researchers were interested in program effects on both academic and developmental outcomes such as self-sufficiency. As in other reviews, the researchers found variations in program focus and duration. They reported limited evidence of positive academic and developmental outcomes and considerable variation in type of outcomes measured. The current synthesis examines OST strategies with academic and other foci across all grade levels.

Recently, Miller (2003) reported on a comprehensive narrative review of after-school programs for middle school children. The purpose of Miller's report was to examine the roles of after-school programs in promoting academic success and positive early adolescent development. Miller described the effects of different after-school programs on academic outcomes and on outcomes that Miller and others connect

with academic success, such as students' attitudes toward school. Although the report provided valuable information related to all facets of how after-school programs can benefit adolescent development, questions about specific effects on achievement in reading and mathematics were left unanswered.

One recent study of OST that is receiving national attention is the first-year evaluation of the 21st Century Community Learning Centers program (U.S. Department of Education, 2003). Congress authorized this program in 1994 to promote broader use of schools by communities and, in 1998, repurposed the program to provide academic as well as recreational activities to students outside of regular school hours.

The evaluation compared the academic and developmental outcomes of elementary and middle school students who attended a 21st Century program with those who did not attend. The unit of analysis was the school district grantee that received program funds to implement one or more centers. In general, first-year findings were discouraging; no statistically significant impacts on achievement were found in reading or mathematics for elementary or middle school students. However, the evaluation documented great variation in the characteristics of centers across school districts, particularly in the range of activities offered and in the emphasis on academic assistance. As a result, it is not possible to link a specific 21st Century program to outcomes of the students served by that program. As the authors (U.S. Department of Education, 2003) noted, "The study was designed to examine the characteristics and outcomes of typical programs and did not attempt to define the characteristics of the best programs" (p. xi). In a footnote they added, "This study focuses on school-based programs that are part of the 21st Century program. Results do not extrapolate to all after-school programs in general" (p. xi). Thus, the evaluation addressed the effectiveness of the 21st Century grant program as a funding source and not the effectiveness of after-school strategies.[3]

President George W. Bush's administration interpreted the results as indicative of problems with the program and requested a decrease in program funding ("After-School Grants," 2003). Some researchers and evaluators of OST have criticized this proposal as premature, contending that one year of findings is an insufficient basis on which to pass judgment about program effectiveness (Harvard Family Research Project, 2003). They also have pointed out methodological weaknesses in the evaluation, despite its use of a randomly selected control group for students in

[3] This evaluation was not included in the current synthesis because student results were not disaggregated for specific OST programs, which was one of our criteria for inclusion of studies.

elementary grades. These critics have called for consolidating knowledge gleaned from many individual evaluations to better approximate the effects of after-school interventions, the approach used for this synthesis.

In summary, the current synthesis contributes to the knowledge base about OST strategies for low-achieving or at-risk students in the following ways:

- This synthesis examines research on OST strategies delivered in all timeframes, including summer school, after school, extended day, before school, vacation sessions, and Saturday schools.
- This synthesis includes the results of separate analyses of the effectiveness of OST strategies for student achievement in reading and in mathematics.
- Both meta-analyses and narrative reviews and descriptions of studies are used to analyze and report findings.
- Studies are included in the review only if they used a comparison group of students who did not experience the OST strategy under investigation.
- Studies are coded for alignment with criteria of research quality, and synthesis results are described in relationship to these ratings.

METHODOLOGY

As described in the next section, both meta-analytic and narrative techniques were used to review research on the effectiveness of OST strategies in assisting low-achieving or at-risk students. For guidance, we consulted other researchers who have published on synthesis methodology (Cooper, 1998; Cooper et al., 2000; Shanahan, 2000).

Literature Searches

The goal of the literature searches was to conduct an exhaustive search for research and evaluation studies of OST strategies for K–12 students within the parameters of our criteria for including studies. We began with a preliminary search of the ERIC database from 1985 through 2003 using keyword search terms of "supplementary education" and "at-risk" or "remediation." The search yielded 1,940 citations; we read the abstracts for the first 50 of these and sorted them into the subject areas addressed in the studies. Based on these findings, we concluded that there was sufficient research on OST strategies related to reading and mathematics to conduct a synthesis, and we identified formal search terms. In May 2003, we conducted several

searches of the ERIC database using FirstSearch and the following parameters: 1985–2003, *not* college, and English-language-only documents. Separate searches were conducted using specific keywords, and citations were identified: "supplementary" – 1,926 citations, "summer school" – 260 citations, "after school" – 1,254 citations, and "vacation" – 254 citations. The four searches resulted in 3,694 citations, which were entered into a master library using EndNote software. We next conducted separate searches of the master library for the terms "literacy" and "reading" and "math" and "algebra" anywhere in the citation. This resulted in a reading library of 880 citations and a math library of 391 citations.

The PsychInfo database subsequently was searched with the following results: "supplementary" – 41 citations, "summer school" – 57 citations, "after school" – 207 citations, and "vacation programs" – 3 citations, for a total of 308 citations. We searched Dissertations Abstracts with parameters of 1985–2003, *not* college, English language only, and PhD dissertations only. We searched in the titles only due to the inordinately large number of irrelevant citations that resulted when the texts of the abstracts were searched. The results were "supplementary" – 64 citations, "summer school" – 36 citations, "after school" – 67 citations, and "vacation programs" – 0 citations, for a total of 167 citations from Dissertation Abstracts.

We next read abstracts of the 1,746 citations obtained from the searches, except when the titles indicated that the studies would be excluded from the synthesis, for example studies of undergraduates or international students. After examining abstracts for relevance to the synthesis based on the criteria described in the next section, we ordered 309 articles.

In addition to the above databases, another major source was the research reviews and syntheses related to OST described in a previous section of this chapter. We examined descriptions of studies in the following research reports and ordered those that met our inclusion criteria: Fashola (1998), Cooper et al. (2000), Redd et al. (2002), Scott-Little et al. (2002), and Miller (2003). We also reviewed the following websites for OST evaluation studies and ordered reports on those that were relevant: Afterschool Alliance, The After School Corporation, Harvard Family Research Project, and National Institute on Out-of-School Time. We ordered 62 additional research studies from reference citations on websites and in research articles and evaluation reports. In sum, the total number of articles that we ordered and read was 371 from a total of 1,808 citations.

Criteria for Inclusion of Studies

The criteria for including studies in this synthesis reflect the research problem and our goal of addressing it through rigorous research and evaluation studies. To operationalize the research problem, we defined an OST strategy as a program, practice, or intervention delivered outside the regular school day.[4] We defined low-achieving or at-risk students as those in grades K–12 who are identified as low performing based on an academic assessment or who are at risk for being low performing based on previously identified risk factors, such as high poverty (Slavin & Madden, 1989). Based on these definitions and the goals of the synthesis, we used the following criteria for including studies:

- Studies had to concern K–12 students.
- A research or evaluation study had to be published or reported in or after 1985. (We chose this date as the approximate start of the standards movement in the United States.)
- The study had to be implemented in the United States.
- Quantitative studies had to include some type of direct assessment of students' academic achievement in reading, mathematics, or both. Examples include classroom assessments, standardized tests, and grades in subject areas. Measures of dropout and student motivation did not qualify as measures of academic achievement. Guided by NCLB requirements, we were more interested in documented achievement than in the prevention of achievement deficits or the potential for achievement.
- Qualitative studies had to include documentation of students' learning in reading, mathematics, or both.
- The study had to examine the effectiveness of an OST strategy for low-achieving students or students at risk for school failure. The study could include students performing at other achievement levels, but it had to disaggregate effects for those entering an OST program with low achievement or at risk for low achievement. Our goal was to assess the effectiveness of OST strategies for those students who are most likely to need them. Low-achievement could be determined by student performance on standardized tests or classroom assessments or through teacher-assigned grades or recommendation for assistance. At-risk status could be determined by characteristics typically associated with lower student achievement and school dropout in large-scale data collections

[4] Because the literature on OST does not differentiate strategies, programs, practices, and interventions, we use the terms interchangeably throughout the synthesis.

including low socio-economic status (SES), racial or ethnic minority background, a single parent family, a mother with low education, and limited proficiency in English (Slavin & Madden, 1989; Miller, 1993).

- Quantitative studies had to include a control/comparison group, which we defined as a group of students who did not participate in the OST strategy under investigation and whose achievement results are compared with those for students who did participate.[5] Thus, in keeping with our emphasis on rigor, included studies had experimental designs or quasi-experimental designs with comparison groups. The primary type of study excluded based on this criterion included only students who participated in the OST strategy. Examples of this type of study include a one-group posttest-only design or a one-group pretest-posttest design (Shadish, Cook, & Campbell, 2002).

- Studies had to disaggregate student results for specific OST programs. Five studies were excluded because they aggregated data state-wide or nationally so that results could not be connected to specific programs and our follow-up queries for disaggregated data and/or local evaluations were not successful.

- Studies were *not* included if they examined OST strategies designed for and delivered *only* to special populations such as special education students, English language learners, and migrant students. Although such OST strategies are important, they are too specific in strategy design and implementation for treatment in the current synthesis.

We included both published and unpublished studies, including evaluation reports, conference presentations, and dissertations. Through this approach, we attempted to avoid the null hypothesis problem (Cooper, 1998) whereby studies that do not find effects from an intervention are excluded from the synthesis because they are not published. This problem tends to bias a synthesis in favor of finding positive results. It is particularly important to examine unpublished studies on OST programs because many of them are evaluations that are disseminated as technical reports for organizations rather than published in peer-reviewed journals. As a counterbalance, we rated each study for research quality and described findings in relationship to this quality.

We read each article that was ordered and received by July 16, 2003. Fifty-three studies met the criteria for inclusion, 47 with reading outcomes and 33 with mathematics outcomes. Of the total, 27 studies addressed outcomes in both subject

[5] Qualitative studies did not require a control/comparison group for inclusion because qualitative approaches use other methods to reach valid conclusions.

areas. There were 250 studies excluded from the synthesis. The main reasons for exclusion were lack of a control/comparison group, lack of student achievement data in reading or mathematics, or the fact that the study did not target low-achieving or at-risk students.

Coding of Studies

The instrument used to code studies for content and quality was a version of the instrument used for a previous research synthesis published by McREL (Barley et al., 2002). We refined the coding instrument to align with the research problem for the current synthesis. The instrument has an initial overview of the study and four major sections: program/intervention and subject/client information, research design/methodology, quantitative analysis (effect sizes and study outcomes) and quality rating. The coding instrument can be found in Appendix A.

Program/Intervention and Subject/Client Information. Each study was coded for descriptive information about the OST strategy that the study examined. This information included the nature of the strategy (e.g., homework help, one-on-one tutoring), content foci (e.g., reading, math, recreational, cultural), timeframe (e.g., after school, summer school), and descriptions of specific strategies related to reading or mathematics. We described how the study identified students as low achieving, the qualifications of those implementing the strategies, how implementers were assigned to different groups in the study, strategy duration defined as the amount of students' average daily exposure to the strategy[6], and student characteristics of grade level, gender, and ethnicity.

Research Design/Methodology. To code the research design of the study, we identified the predominant methodology as the one on which study conclusions were based. We described quantitative research as either experimental or quasi-experimental. To be classified as experimental, students had to be randomly assigned to treatment or control/comparison groups. Studies classified as quasi-experimental did not randomly assign students to comparison groups but often used procedures to equate or match the different groups, which we described. Quantitative designs were coded for whether students were pretested on achievement prior to strategy implementation and posttested afterward or only posttested. We coded qualitative research designs as case studies, action research/field studies, studies using grounded theory, and ethnographic studies, and we noted when qualitative studies used more

[6] Due to inconsistent reporting of the frequency of OST strategies, the duration of strategies was used to indicate the amount of participants' exposure to the strategies.

than one of these approaches. For both quantitative and qualitative research, we described any secondary methods that the study used.

Quantitative Analysis. Statistical results for quantitative studies were coded for each outcome measure for each student group in the study and included the information needed to conduct a meta-analysis: group means, and standard deviations, effect sizes, and inferential test statistics. For both quantitative and qualitative studies, we described the relevant findings and conclusions that related to the research problem addressed by the synthesis.

Quality Rating. To code the quality of quantitative studies, we used Shadish et al.'s (2002) framework on threats to validity and the Study Design and Implementation Assessment Device proposed for the What Works Clearinghouse (Valentine & Cooper, 2003). Both examine research studies for four types of validity: construct, internal, external, and statistical. For example, related to construct validity, we examined whether the intervention (i.e., the OST strategy) was properly defined and whether fidelity of the intervention was measured or discussed.

We assigned points to a study based on the degree to which research methods addressed each type of validity as indicated by the information provided in the article. In assigning points, we judged that for the purposes of this synthesis, there should be more weight given to internal validity and construct validity than to external and statistical conclusion validity. These criteria resulted in the following quality scale for quantitative studies: low (0–14 points), medium (15–21 points), and high (22–26). Tables 1.1 and 1.2 describe the characteristics of quantitative studies rated as "low" and "medium" respectively. These examples are for studies that rated on the high end of their rating categories. A study with the minimum points for a medium rating would have characteristics that fall in-between the two example studies. A study rated as high would have the characteristics of the study in Table 1.2 but would meet *all* of the requirements for at least one of the four types of validity.

Table 1.1 Characteristics of a Quantitative Study with a Low Quality Rating of 11 Points

Type of Validity	Study Characteristics
Construct Validity	The description of the intervention is incomplete. Treatment fidelity is discussed, but there is no report of its assessment. There is evidence for face validity of the outcome measure but not for the construct it represents.
Internal Validity	The steps taken to make student groups comparable may have been inadequate. Although alternative explanations for results are not readily apparent, some remain plausible.
External Validity	Only some of the important characteristics of the participants, settings, and outcomes are represented in the sample. The intervention was tested for effectiveness with only a few important subgroups of participants.
Statistical Validity	Effect sizes can be calculated for only some outcome measures due to insufficient reporting.

Note: Rating scale: low (0–14 points), medium (15–21 points), high (22–26 points)

Table 1.2 Characteristics of a Quantitative Study with a Medium Quality Rating of 21 Points

Type of Validity	Study Characteristics
Construct Validity	The description of the intervention is adequate and largely reflects commonly held ideas about its definition. Treatment fidelity is discussed and its assessment is reported. There is evidence for the alignment of the outcome measure with the intervention and for construct validity of the outcome measure.
Internal Validity	There were adequate steps taken to make student groups comparable. Alternative explanations for results are ruled out.
External Validity	The most important characteristics of the participants, settings, and outcomes are represented in the sample. The intervention was tested for effectiveness with most but not all important subgroups of participants.
Statistical Validity	Effect sizes can be calculated for most but not all outcome measures.

Note: Rating scale: low (0–14 points), medium (15–21 points), high (22–26 points)

We coded qualitative studies for whether the research had characteristics of dependability, credibility, confirmability, and transferability (Miles & Huberman, 1994), and gave greater weight to the first two characteristics. For example, related to dependability, we coded the studies for whether the constructs used for analyses of

qualitative data were clearly defined and whether data were collected across the full range of settings, times, and respondents as suggested by the research questions. The resulting quality scale for qualitative studies was low (0–9 points), medium (10–21 points), and high (22–31 points). Table 1.3 lists the characteristics of a qualitative study rated as being of high quality, of which there were two. The qualitative study rated as medium only partially met each study characteristic for the four types of validity. There were no low-quality qualitative studies in the synthesis.

Table 1.3. Characteristics of Qualitative Studies Rated as High

Type of Validity	Study Characteristics
Confirmability	The study used at least two methods to verify findings, such as member checking and an audit trail. The study used at least two methods to control for researcher effects, such as triangulation of data and the use of unobtrusive measures.
Dependability	The research questions are completely clear and congruent with features of the study design. Data were collected across the full range of appropriate settings, times, and respondents. Paradigms and analytic constructs are clearly specified.
Credibility	There are multiple sources of evidence used to produce converging conclusions. The study used at least two methods to support findings, such as a search for disconfirming evidence and the generation of rival explanations. The presented data and measures reflect constructs of prior theory.
Transferability	The characteristics of the sample and setting are fully described so that potential transferability to other samples and settings can be assessed. The researcher fully defined the scope and boundaries of generalization from the study.

Coding Procedures. Coding procedures were designed to help the authors of the current synthesis reach a common understanding of the codes used to describe each study and to check for the reliability of coding results among the authors. Coding procedures incorporated Stock's (1994) recommendations for reducing coding errors.

Each of the synthesis authors participated in coder training, which involved an overall description of the coding form, explanations for items in each section, and examples of information from studies to be extracted and judged. The authors confirmed that they had a common understanding of terms used for coding and that the instrument included sufficient information for adequate description of study characteristics and quality.

Following initial training, each author independently coded two studies that had both reading and mathematics student outcomes. The authors then compared completed forms and identified and resolved discrepancies. Based on the resolutions, revisions were made to the coding form; for example, more detailed distinctions were added concerning how to code the strategies used in the intervention or programs. The quality rating section also was revised to include an item pertaining to whether intervention fidelity was assessed. Each author then independently coded two additional studies, which resulted in improved coding consistency. The authors reached consensus on the quality ratings for the four studies, and confirmed the face validity of the ratings — that is, a study rated as high quality based on points was a study considered high in overall quality for the purposes of this synthesis.

Coding procedures and decisions were double-checked at several points in the analysis within each pair of authors for the reading and mathematics chapters or among authors across chapters. Double-checking occurred during data entry for the meta-analysis, in preparation of chapter tables and reporting of findings, during internal review of the chapter drafts, and during chapter revisions. Prior to data entry of the program/intervention information for each study in the reading and mathematics chapters, the pair of authors for that chapter reached consensus on the type of strategy, content focus, and quality rating. During data entry, the coding results for studies included in both the reading and mathematics chapters were compared to confirm consistency across chapters. Any discrepancies were resolved among the four authors of those chapters.

ANALYSES AND RESULTS

Based on their background knowledge and expertise, two-person teams of researchers analyzed and synthesized studies of OST strategies that measured reading and mathematics outcomes. This approach aligned with our goal of describing the effectiveness of OST strategies in assisting low-achieving or at-risk students in the two content areas. The teams followed common procedures for evaluating and analyzing studies and presenting results. These procedures were jointly developed prior to data analyses, and written presentations were modified through frequent discussions.

Because sufficient numbers of studies provided the quantitative information needed to compute effect sizes, separate meta-analyses for reading and mathematics were conducted.[7] It also was determined that effect sizes would provide meaningful and

[7] Effect size refers to the magnitude of the effect of a strategy/intervention on an outcome such as student achievement. In general, the larger the effect size, the stronger the relationship between the

useful information within the context of each outcome category. Appendix B describes the methods used to conduct the meta-analyses.

Moderators

Based on the research literature related to OST, the following strategy characteristics were identified as possible moderators of effect sizes: timeframe, grade level, strategy focus, strategy duration, and student grouping. Timeframe refers to whether the OST strategy was delivered to students after school, in summer school, or in some other time-related format. Much of the OST research has been organized around when program delivery occurs, as in Cooper et al.'s (2000) synthesis of summer school research and Fashola's (1998) review of research on after-school programs. There has been little discussion of OST effectiveness related to variations in timeframe. By examining this variable, we hoped to learn about the relationship of time of program delivery and the strategy being used during the program.

Several researchers have suggested that the effectiveness of OST might vary depending on the grade levels of the students. Cooper et al. (2000) documented more benefit from summer school for students in early elementary grades and secondary grades compared to students in late elementary grades. Grossman et al. (2001) indicated that secondary students are less attracted to after-school programs than are elementary students and are more difficult to recruit. Other researchers have suggested that the focus of OST needs to differ depending on the ages of the participants. For example, OST programs for older students should be more recreational than those for younger students (Miller, 2003).

Due to the wide variation in the foci and goals of OST programs, it is logical to conclude that the degree to which an OST strategy focuses on academics might influence the effectiveness of a strategy in improving student achievement. According to a report by Policy Studies Associates (1995) for the U.S. Department of Education, connecting OST activities to regular academic programs in schools is a feature of promising practices that extend learning time for disadvantaged students. However, others suggest that to be effective, strategies for disadvantaged students should "not be too closely identified with schools and, hence linked to the uncaring and unknowing attitudes that neighborhood parents and youths characterized as typical of local schools" (Heath, 1994, p. 32). Miller (2003) agreed that for low-income students, experiencing the same learning strategies that they experience in

---

strategy/intervention and the outcome. For an explanation of the practical use of effect size, consult Marzano, Pickering, and Pollock, 2001.

school is not likely to be beneficial. Miller supports a wide variety of activities for OST learning programs.

Based on prior research, we identified the duration of an OST strategy as another possible moderator of OST effectiveness. McComb and Scott-Little's review (2003) suggested that students who attend OST programs more, and therefore experience more exposure, benefit more. (Although we were unable to analyze student attendance, OST programs that are longer in duration provide students with more exposure to OST activities and might be more effective than those that are shorter in duration.) However, other research has shown that with regard to academic learning, the amount of time is less important than what occurs during that time (WestEd, 2002), and that extending the time *for* learning does not mean that students will spend that time *in* learning (Ascher, 1990).

The final strategy characteristic we examined was how students were grouped for OST activities. Fashola's (1998) review indicated that individualization through one-on-one tutoring is a promising practice among programs designed to improve academic achievement. A research synthesis by Barley et al. (2002) found that both tutoring and peer tutoring can be effective strategies for improving achievement during the school day, so it is likely that the same benefits would occur during OST. However, a report by Policy Studies Associates (1995) on promising after-school practices concluded that the key is to engage students' attention, which can occur through traditional classroom instruction.

In addition to characteristics of OST strategies, we also looked at characteristics of studies. As mentioned previously, researchers (Scott-Little et al., 2002; Fashola, 1998) have identified the need for higher quality research of OST strategies. Only quantitative studies with control/comparison groups were included in the current synthesis. In addition, recognizing that research quality reflects criteria related to different types of validity, we examined how study findings related to our quality ratings.

Another study characteristic that was a moderator in this synthesis was the type of publication, such as a peer-reviewed journal article or a dissertation. As Cooper (1998) indicated, peer-reviewed journals are more likely to publish research that reports statistically significant effects than those that support the null hypothesis.

A final study characteristic was the type of score used to calculate effect sizes for studies in the meta-analyses. Studies reported one of two types of achievement scores: gain scores based on the differences between pretests and posttests for each comparison group of students, or the posttest scores of each comparison group. Type

of score was included as a moderator so that its influence on effect sizes could be assessed.

Thus, prior research on the relationships of OST strategy characteristics to strategy effectiveness has been inconclusive. By examining these characteristics in the current synthesis, we aimed to better understand their influences. By including study characteristics as moderators, we sought to present research findings in relation to method and publication contexts.

OVERVIEW OF SYNTHESIS

Chapters 2 and 3 describe the analyses and results for a synthesis of research on OST strategies that address reading and mathematics respectively. Each chapter describes the studies that were analyzed and presents results from meta-analysis and moderator analysis. There is also a narrative review[8] of studies that met the inclusion criteria but had insufficient data for meta-analysis. Synthesis findings are supplemented by narrative descriptions of relevant research studies. Conclusions about the results in each chapter are based on the extent and quality of the research. Each chapter also discusses implications for policy and practice. Chapter 4 suggests some overall conclusions across the chapters in relation to the research problem.

A final note concerns approaches to research syntheses. Researchers have published on different types of syntheses and provided guidelines for their conduct (Cooper, 1998; Shanahan, 2000). However, there has been disagreement about which synthesis methods are most appropriate (Wayne & Youngs, 2003). Given the identified research problem, the goals of this synthesis, the nature of the studies that met the inclusion criteria, and the audience, we chose to use a meta-analytic approach to the studies. In addition to the results from these analyses, this report includes narrative reviews and descriptions of informative studies that did not have the necessary data for meta-analysis, including qualitative studies, as well as summaries of individual studies that we judged as informative concerning the nature of programs that deliver OST strategies. Through this multi-method approach, we hoped to inform our audience about the research base related to the use of OST strategies to improve achievement and the types of OST programs that are successful.

[8] The methods used for locating and coding studies for the narrative reviews and meta-analyses were equally systematic. The primary difference was the greater precision in reporting results of the meta-analyses.

CHAPTER 2

Studies of the Effectiveness of Out-of-School-Time Strategies for Reading Achievement

> *The ongoing literacy development of adolescents is just as important, and requires just as much attention, as that of beginning readers. The expanding literacy demands placed upon adolescent learners includes more reading and writing tasks than at any other time in human history. They will need reading to cope with the escalating flood of information and to fuel their imaginations as they help create the world of the future.* (International Reading Association, 1999)

Given the critical role that literacy plays in a child's future, programs, strategies, and interventions designed to help develop basic and advanced reading skills need close examination. This chapter presents a synthesis of current research that addresses the effectiveness of out-of-school-time (OST) strategies in improving the reading achievement of low-achieving or at-risk students.

The chapter first presents background information related to important constructs of this synthesis including the focus and timeframe of OST strategies and the developmental aspects of becoming a proficient reader. This section is followed by a description of the methodology employed in reviewing the research and evaluation studies on OST and reading. Studies selected for inclusion in the synthesis are then reviewed. Results from meta-analysis and moderator analysis to address the following research questions are then presented:

1. What is the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading?

2. How does the effectiveness of OST strategies differ by *program* characteristics such as timeframe, grade level of students, activity focus, program duration, and student grouping?

3. How does the effectiveness of OST strategies differ by *study* characteristics such as research quality, publication type, and score type?

Narrative summaries of relevant research studies that met the selection criteria also are provided to supplement the meta-analysis results. The chapter concludes with a discussion of findings and implications for policy and practice.

BACKGROUND

According to Slavin, Karweit, and Madden (1989), "The negative spiral that begins with poor achievement in the early grades can be reversed" (p. 4). The authors suggested that by utilizing programs and instructional strategies geared at helping all children achieve adequate basic skills, the school success of many children can be increased. In the current context of standards-based reform and accountability, we know that all children by the end of grade 3 need to be able to read and understand both literary and informational texts. Reading is a component of *literacy,* which is defined as the "complex, dynamic, interactive and developmental process of making meaning with text" (Davidson & Pulver, 1991, as cited in Davidson & Koppenhaver, 1993, p. 12). Simply put, *reading* is the "process of understanding written language" (New Standards Primary Literacy Committee, 1999, p. 19). The National Institute for Literacy includes speaking, gathering information, thinking critically, understanding others, and expressing oneself in its definition of reading (Hynes, O'Connor, & Chung, 1999). In most states and districts, reading is a strand of the content standards and benchmarks in the area of Language Arts — along with the strands of listening and speaking, writing, viewing, and media.

Achieving reading proficiency requires that students master certain knowledge and skills at or before critical grade levels. During the primary years (K–2), children need to master all of the reading fundamentals, for example associating sounds with written words. During the intermediate grades (3–5), children need to develop and use, in some cases effortlessly, all word identification concepts and skills, as well as comprehension strategies such as recognizing confusion, adjusting one's strategies, and identifying and summarizing main ideas and important details (McREL, n.d.). As children prepare for and progress through middle school and high school, they are expected to develop and use advanced reasoning for reading so that they can understand and interpret texts well enough to take and pass a college-preparation sequence of courses (Committee for Economic Development, 2000).

Results from the National Assessment of Educational Progress (NAEP) indicate that a large percentage of students are not meeting reading standards. For example, *The Nation's Report Card*: *Reading 2002* reported that 69 percent of fourth graders did not demonstrate proficiency in reading and were unable to read a fourth-grade text and make inferences, draw conclusions, and make connections to their own experiences (Grigg, Daane, Jin, & Campbell, 2003). Among fourth-grade students

from low-income homes, 87 percent failed to meet these same benchmarks. At eighth grade, 85 percent of the NAEP sample failed to demonstrate proficiency in reading.

In order to help students be proficient in academic standards for reading, many educators are considering the utility and effectiveness of OST strategies and programs. The purposes of using OST strategies for assisting low-achieving students in reading are varied. These include the prevention of summer learning loss, early intervention, remediation of skill deficiencies, acceleration of learning, increased motivation to read, and preparation of students for the intellectual challenges of later schooling and work. In addition to an academic focus, OST strategies and programs enable educators to address the safety, behavioral, cultural, vocational, emotional, and social needs of students. The timeframes for delivering OST strategies that are discussed in this chapter include after school, Saturday school, and summer school. The variation among the purposes and formats of OST strategies reflects how interventions address the different academic and social learning needs of students. The National Institute on Out-of-School Time "believes that high-quality after-school programs focus on the development of the whole child, integrating academic supports such as literacy skills into programming that also promotes children's social, emotional, and physical development" (Hynes et al., 1999, p. 1). Others have emphasized the informality of after-school programs as being well suited to developing the social and cultural dimensions of literacy, such as helping children see how reading and writing can be intrinsically rewarding and relevant to their lives (Speilberger & Halpern, 2002). One purpose of this review is to examine evidence of the effectiveness of OST strategies and programs designed to address the academic and/or social-emotional needs of students.

METHODOLOGY

Chapter 1 described the review process and inclusion criteria for both quantitative and qualitative studies regarding strategies for improving the reading performance of low-achieving or at-risk students. This chapter synthesizes this research. This chapter also includes information from background articles that reflect current thinking related to reading and OST strategies, findings from previously conducted meta-analyses and syntheses on summer school and after-school programs, and evidence from the primary quantitative and qualitative studies described in the following section. The primary studies served as our data sources for addressing the research questions.

In order to address the first research question regarding the effectiveness of OST strategies in reading, we calculated an overall effect size for studies (Appendix B describes the methods used for meta-analysis). We then conducted homogeneity analyses in order to examine if the average effect sizes significantly differed by

moderators of program and study characteristics. Finally, we conducted a narrative review of studies not included in the meta-analysis, and we described noteworthy themes that emerged during our review of all reading studies. These themes are intended to supplement the meta-analysis findings.

Study Selection

As described earlier, reports related to OST strategies for reading were located in an initial search of ERIC and other databases. Researchers identified additional studies for possible inclusion through report and article reference lists, and other online reports and databases such as the Harvard Family Research Project's Out-of-School-Time program evaluation database.

The literature search and review of abstracts resulted in 47 reports on OST strategies on reading that met the synthesis inclusion criteria. Of these, 44 were quantitative studies that employed the use of comparison or control groups, and 3 were qualitative studies that focused on student learning in reading. Most of the reports that did not meet inclusion criteria were program descriptions, did not use control or comparison groups, or focused on students outside of our target population of K–12 students. A few reports were excluded because they dealt with international programs or focused solely on special populations (e.g., Limited English Proficient students, migrant populations, or learning disabled). Researchers coded the 47 studies for a variety of information including data on specific strategy characteristics that might influence program effectiveness on student learning such as student demographics, strategy timeframe (e.g., summer school or after-school), focus (e.g., academic, social, recreational, cultural), and duration of the intervention. (Appendix A contains the instrument for coding studies.)

Data Analysis

As the chapter authors for the synthesis of research on OST strategies that address reading, we reviewed each study and discussed how we coded them to ensure the reliability of coding. As part of this process, we determined if studies reported sufficient data for conducting a meta-analysis. Twenty-seven studies on OST strategies for reading reported effect sizes or data that could be used to calculate effect sizes. If a study included sufficient data to calculate effect sizes and the results were non-significant, it was still included in the meta-analysis. These 27 studies yielded 43 independent samples for the meta-analysis. The number of independent samples from a single study varied from one to five. Twenty studies were determined to be inappropriate for the meta-analysis either because they were qualitative studies (n = 3) or did not report sufficient data to calculate effect sizes (n = 17). It is

important to note, however, that these 20 studies included measures of student learning; the findings of these studies, whether significant or non-significant, are presented following the meta-analysis section.

Using the meta-analytic approach described in Appendix B, effect sizes weighted by sample sizes (weighted *d*s) were calculated for each study that reported sufficient data. To address our first research question on the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading, we computed an overall effect size based on the 43 independent samples. We used 95 percent confidence intervals to determine if the effects of OST strategies on reading achievement were significantly greater than zero. Our second and third research questions address how the effectiveness of OST strategies varies by strategy moderators of timeframe, grade level, focus, duration, student grouping and by study moderators of research quality, publication type, and score type. In conducting moderator analyses, we used independent samples as the unit of analysis for computing effect sizes for grade level and studies as the unit of analysis for all the other moderator analyses.

We coded the grade levels of sample students using four categories: lower elementary (K–2), upper elementary (3–5), middle school (6–8), and high school (9–12) levels. When an independent sample overlapped two categories, we chose the category in which the majority of grade levels fell. For example, the Bergin, Hudson, Chryst, and Resetar (1992) study included kindergarten through third graders and was categorized as lower elementary rather than upper elementary. The grade level of one independent sample overlapped categories (it included all elementary and middle school grades), so its effect size was excluded from the moderator analysis for grade level.

We coded strategy focus either as "academic" or "academic and social." Studies in which the OST strategy focused purely on academic enrichment in reading, including homework assistance, study skills, and remedial lessons, were coded as "academic." We coded studies as "academic and social" if the OST strategy focused not only on academic enrichment, but also on social enrichment including music, art, social skills, recreational activities, and vocational activities.

Strategy duration was based on the total hours of treatment and was coded using four categories: less than 44 hours, 44 to 84 hours, 85 to 210 hours, and more than 210 hours. Five studies did not report sufficient information to compute the total hours; thus they were excluded from this analysis.

We examined two publication characteristics related to OST strategies: study quality and publication type. As described in Chapter 1, studies were categorized as high, medium, or low based on the indicators of research quality the project team

developed. The studies also were categorized by publication type: conference paper/report, dissertation, and peer-reviewed journal. A final characteristic coded was the type of score reported — gain score or posttest score.

OVERVIEW OF STUDIES

Table 2.1 describes the characteristics of the 47 studies selected for this chapter on OST strategies that address reading achievement. The publication year of these studies ranged from 1985 to 2003; 20 of these studies were published in 2000 or later. Twenty-four studies examined the impact of summer school programs on participants' reading achievement; 19 involved research on after-school strategies; two involved research on Saturday school; and two involved research on a mix of strategies (e.g., summer school and Saturday school). The majority of studies (32) concerned programs that emphasized only academics, whereas 14 studies involved programs that focused on both academic and social skills. The latter programs often included recreational, cultural, or vocational components in addition to their emphasis on academic and social skills. The studies included in the meta-analysis versus the narrative review did not differ greatly on study characteristics such as grade level(s), timeframe, program focus, or grouping strategies. The main difference between the studies included in the meta-analysis versus the narrative review was that the narrative review studies did not report sufficient data to calculate effect sizes.

As stated previously, to be included in this synthesis, studies had to measure student learning in reading. The three qualitative studies included pre/post assessments and also included observations, interviews, or self-report surveys to measure student learning. Of the 44 quantitative studies, seven employed norm-referenced assessments that measured and reported on specific reading dimensions such as vocabulary, phonemic awareness, and reading comprehension. Thirty other studies reported aggregated reading scores from standardized assessments, and seven studies employed other outcome measures, including teacher grades and end-of-grade tests.

Nine of the 44 quantitative reading studies used random assignment to treatment and control groups. One study matched groups with a pretest, 21 matched groups using other criteria such as demographics, and 13 studies did not report any matching. For the 27 studies included in the meta-analysis, we computed effect sizes based on 14 studies that reported gain scores or pretest-posttest difference scores and 13 studies that reported only posttest scores.

All of the studies examined low-achieving or at-risk students, although each study defined students according to different characteristics such as low performing, low income, and not promoted. The grade level of the students in the studies ranged from

kindergarten to the 12th grade. Twenty-three percent (n = 11) of the studies involved students across several grades spanning elementary, middle, and high school levels. Twenty-eight percent (n = 13) of the studies targeted lower elementary students (e.g., kindergarten through second graders), 19 percent (n = 9) involved upper elementary students (e.g., third through fifth graders), 23 percent (n = 11) focused on middle school students (e.g., sixth through eighth graders), and 7 percent (n = 3) included high school students (e.g., ninth through twelfth graders).[9]

The duration of OST programs reflected in the studies ranged from three weeks to the entire school year over a period of one, two, or three consecutive years. The duration of programs ranged from nine hours to 750 hours, with an average duration of 127 hours and a median of 78 hours. [10] For the studies included in the meta-analysis, the total number of hours offered by each program ranged from 9 to 450 hours; for these studies, the median program duration was 84 hours.

Table 2.1. Studies of Out-Of-School-Time (OST) Reading Strategies

Author(s) and Year	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Baker & Witt (1996)	302	3rd–6th	low SES[a]	Academically oriented activities in the context of a goal-oriented, fun, recreational experience; teacher-directed, large- and small-group instruction; focus on activities that promote cultural awareness and positive self-esteem and attitude	after school
Bergin, Hudson, Chryst, & Resetar (1992)	10	K–3rd	low SES	Phonics-based, direct instruction model with child-centered, culturally sensitive teaching methods and materials; Sing, Spell, Read & Write curriculum	after school
*Borman, Rachuba, Fairchild, & Kaplan (2002)	438	K–1st	low SES	Integrated read-aloud and math activities, recreation, art, foreign language, and drama; 8 students maximum per class	summer school
Branch, Milliner, & Bumbaugh (1986)	752	6th–8th	low performing	STEP (Summer Training and Education Program) combined an existing federal work program with drop-out prevention strategies	summer school
Cosden, Morrison, Albanese, & Macias (2001)	90	4th–6th	low performing	Homework time and support	after school
D'Agostino & Hiestand (1995)	1,006	4th	low performing	Academic focus emphasizing higher order thinking, questioning, and problem-solving skills	summer school

[9] Some studies in each of these categories only focused on one grade rather than the entire grade span (e.g., 3rd–5th grade).

[10] Some studies only reported the number of hours per week and indicated that the intervention occurred for the entire school year. Excluding the first and last weeks of a 180-day school year, we used 30 weeks as the duration of an entire school year.

Author(s) and Year	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
*Duffy (2001) (qualitative design)	10	2nd	low performing	Balanced, accelerated, and responsive literacy program; whole-group reading and sorting; individual reading and writing; book talk and read aloud; instructional-level support reading	summer school
Foley & Eddins (2001)	1,978	2nd–5th	educator identified	Virtual Y, YMCA program; literacy-based activities; addresses socio-emotional behaviors and four core values: respect, responsibility, honesty, and caring	after school
Gentilcore (2002)	114	8th	educator identified	Preparation to help students pass state assessment; 8-10 hours total; workbook practice in reading passages and writing responses	after school
Grimm (1997)	19	5th–8th	educator identified	Residential summer program with follow-up mentoring from shipyard workers; summer school and follow-up activities included academic classes to support or remediate skills, dinners with mentors, and field trips	summer school & after school
Hansen, Yagi, & Williams (1986)	871	3rd–7th	not promoted	Arts and crafts and academic remediation	summer school
Harlow & Baenen (2001)	86	7th–8th	have high potential but are at-risk	An intensive enrichment program stressing academic excellence, leadership, creativity, and diversity; small classes to allow individual attention to students	summer and Saturday school
Hausner (2000)	128	K	low performing	Scaffold instruction; shared & guided reading; independent learning and teacher-directed, small- and large-group instruction	after school
Hink (1986)	48	1st –9th	educator identified	Teacher-directed, remedial, large-group instruction. Summer program teachers consulted with teachers from prior school year.	summer school
Holdzkom (2002)	3,043	3rd–8th	low performing	A summer academy designed by and implemented at individual schools provided by the district	summer school
Howes (1989)	22	1st	low SES and low performing	Remedial instruction to groups of 10 to 15 students for 10 hours/week for 3 weeks total; focus on developing phonics, comprehension and writing skills	summer school
Huang, Gribbons, Kim, Lee, & Baker (2000)	4,312	2nd–5th	low performing	Homework time and support; academic, recreational, and social and motivational components	after school
Jacob & Lefgren (2001)	147,894	3rd & 6th	low performing	Teacher-directed instruction in groups of 15 students	summer school
King & Kobak (2000) (qualitative design)	13	7th	low performing	Direct instruction in strategic reading for understanding; keeping reading response journals; game-like cooperative activities; parent involvement	summer school
Kociemba (1995)	192	2nd & 5th	low performing	Academic focus including reading comprehension	summer school

Author(s) and Year	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Kushmuk & Yagi (1985)	67	3rd–7th	not promoted	Arts and crafts and academic remediation program for public school students in Portland, Oregon (see also Hansen, Yagi, & Williams, 1986)	summer school
Leboff (1995)	40	3rd	low performing	Remedial Chapter 1 program for urban youth	summer school
Legro (1990)	49	1st–2nd	low SES	One-on-one homework tutoring; parent involvement, partnership program; social and communication skills component	after school
Leslie (1998)	73	6th–8th	low performing	One-on-one tutoring, homework support, and incentives (e.g., students earned tickets to purchase tickets to play games)	after school
Levinson & Taira (2002)	1,289	3rd–5th	not promoted & low performing	Homework support; computer-assisted instruction; teacher-directed lg. group instruction; leveled trade books; word study, reading, vocabulary, writing	summer school
Lodestar Mgmt. Research (2003)	160	2nd–8th	low performing	Homework time and support; cultural and recreational activities with reading and writing exercises interwoven	after school
Luftig (2003)	34	K– 4th	educator identified	Small-group tutoring; phonics instruction tied to district curriculum	summer school
McKinney (1995)	47	1st–2nd	low performing	One-on-one tutoring program; self-concept and non-academic enrichment component	after school
Mooney (1986)	15	4th	low performing	Trained, 8th-grade peer tutors helping 4th graders with understanding and completing reading homework assignments	after school
Morris, Shaw, & Perney (1990)	30	2nd & 3rd	low performing	One-on-one tutoring; shared reading, word study, writing personal stories, reading to child; basal sets and trade books	after school
*Ortiz (1993) (qualitative design)	3	1st	low performing	Parent and student collaborative learning; teacher-directed small-group instruction; parent coaching & support; writing and reading in a risk-free environment	after school
Paeplow, Baenen, & Banks (2002)	116	2nd–8th	low performing	One-on-one tutoring and cooperative learning leadership program; teacher-directed, small-group instruction	summer school
Phelan (1987)	17	7th & 8th	at-risk for dropping out	Remediation and enrichment activities including development of computer skills	Saturday school
Prenovost (2001)	271	6th–8th	District-wide low performance	Homework support, enrichment, field trips, and sports	after school
Pyant (1999)	30	K– 4th	low performing	Tutoring with focus on reading, spelling, & student attitudes; social skills component includes modeling, role playing, & real-life situations	after school

Author(s) and Year	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Rachal (1986)	9,675	2nd–5th	low performing	Compensatory/remedial program in Louisiana	summer school
Raivetz & Bousquet (1987)	141	9th	low performing	One-on-one tutoring and teacher-directed, large-group instruction	summer school
Reed (2001)	30	1st	low performing	Individualized instructional programs using the "Prescription for Reading Improvement" thru four class periods: (1) language development, (2) phonics instructional time, (3) fluency in reading, and (4) reading potpourri	summer school
Rembert, Calvert, & Watson (1986)	87	10th–12th	educator identified	College prep through classroom instruction that mimicked college courses, mentoring, and computer-assisted instruction	summer school
Roderick, Engel, & Nagaoka (2003)	21,000	3rd, 6th, & 8th	low performing	Preparation for passing state assessment through practice and instruction on types of problems and reading comprehensions tasks on the assessment. Some teachers provided individualized attention (e.g., assigning extra reading) and consultation with teachers from prior school year	summer school
Ronacher, Tullis, & Sanchez (1990)	1,072	9th–12th	low performing	Study and life skills program	Saturday school
Ross, Lewis, Smith, & Sterbin (1996)	328	2nd–4th	low performing	Small-group tutoring program based on components of Success For All; cooperative learning & teacher-directed instruction; focus on reading, writing, & computer skills	after school
*Schacter (2001)	21	1st	low performing	Systematic reading curriculum with camp activities that promote social & emotional growth; one-on-one tutoring, teacher-directed instruction; Open Court Reading series, word study, daily phonics instruction, journal writing, reading, computer-assisted instruction	summer school
Schinke, Cole, & Poulin (2000)	283	5th–8th	low SES	Homework assistance; mentoring; incentives	after school
Sipe, Grossman, & Milliner (1988)	1,272	5th–7th	low SES and performing	A work-study program providing basic skills remediation (including silent sustained reading and computer-assisted instruction) and life skills instruction; includes data from five urban demonstration sites	summer school
Smeallie (1997)	31	6th–8th	low performing and educator identified	Homework assistance; teacher-directed instruction on study skills; incentives; parent seminars on homework issues	after school
Ward (1989)	385	3rd & 6th	low performing	Teacher-directed instruction with an emphasis on minimal skill achievement; no basals allowed, hands-on activities instead	summer school

[a] SES: socio-economic status
[b] The n for the meta-analysis could be smaller based on the data available to calculate effect sizes.
*Studies rated as "high" based on quality indicators used for this synthesis

Research Quality Review

As previously described, studies considered for inclusion in the synthesis were rated on the quality of the research based on separate indicators for quantitative and qualitative methodologies. We used the indicators as descriptors of the research included in this synthesis. Table 2.2 presents the number of studies in each rating category (i.e., high, medium, and low).

Table 2.2. Ratings of Reading Studies Based on Quality Indicators

Methodology	Rating	Number of Meta-analysis Studies	Number of Narrative Review Studies	Total Number of Studies
Quantitative	High	2	1	3
	Medium	18	9	27
	Low	7	7	14
Qualitative	High	-	2	2
	Medium	-	1	1
	Low	-	-	-
Total		27	20	47

Based on the quality indicators, the majority of quantitative studies included in this chapter were rated as being of "medium" quality. The three studies that received "high" ratings presented thorough descriptions of the intervention and implementation fidelity measures; used comparable treatment and control groups; ruled out potential effects caused by concurrent events; targeted appropriate participants, settings, outcomes, and occasions in the study; tested effectiveness within important subgroups of the sample; and accurately estimated and reported effect sizes. In general, the medium-rated studies addressed most of these indicators, but with less sufficiency or clarity. All 14 studies with a "low" rating omitted a measure or discussion about implementation fidelity of the intervention. Other reasons for a "low" rating included limited or missing descriptions of strategies or interventions used incomplete description of the target population of students, incomplete reporting of results, no report on steps taken to make treatment and control groups comparable, and/or no tests of the intervention for its effectiveness within subgroups.

The two qualitative studies rated as "high" presented methods for confirming study results and controlling for researcher effects; specified clear research questions aligned with the study's design and analytic approach; used multiple sources of

evidence; employed techniques to rule out alternative explanations; and defined the scope and the boundaries of reasonable generalization from the study.

META-ANALYSIS RESULTS

This section presents the findings from the meta-analysis and moderator analysis. We begin with a report on the overall effect size for studies included in the meta-analysis and the results from the homogeneity analysis, which determines whether the effect sizes from selected studies varied more than expected by sampling error alone. Next, we present results from the analysis of moderators of the effect sizes, which includes moderators from program characteristics and from study characteristics. (See Appendix B for a description of the meta-analysis methodology.)

Overall Effect Size of OST Strategies in Reading and Homogeneity Analysis

In order to determine the effectiveness of OST strategies in assisting low-achieving or at-risk students in reading, we calculated effect sizes (weighted *d*s) for each of 43 independent samples yielded from 27 studies. Table 2.3 presents information on each independent sample, including the number of treatment students (those who received the OST strategy); defining characteristics of the independent sample such as grade level or gender[11]; the effect size; the lower and upper limits of the 95 percent confidence interval for the effect size; and a graphic display of the effect sizes and confidence intervals. When we examined the effect sizes for statistical outliers, there was only one outlier ($d = 2.35$) and its adjustment had no influence on the results, so the original analysis is reported here. (See Appendix B for a description of the outlier analysis.)

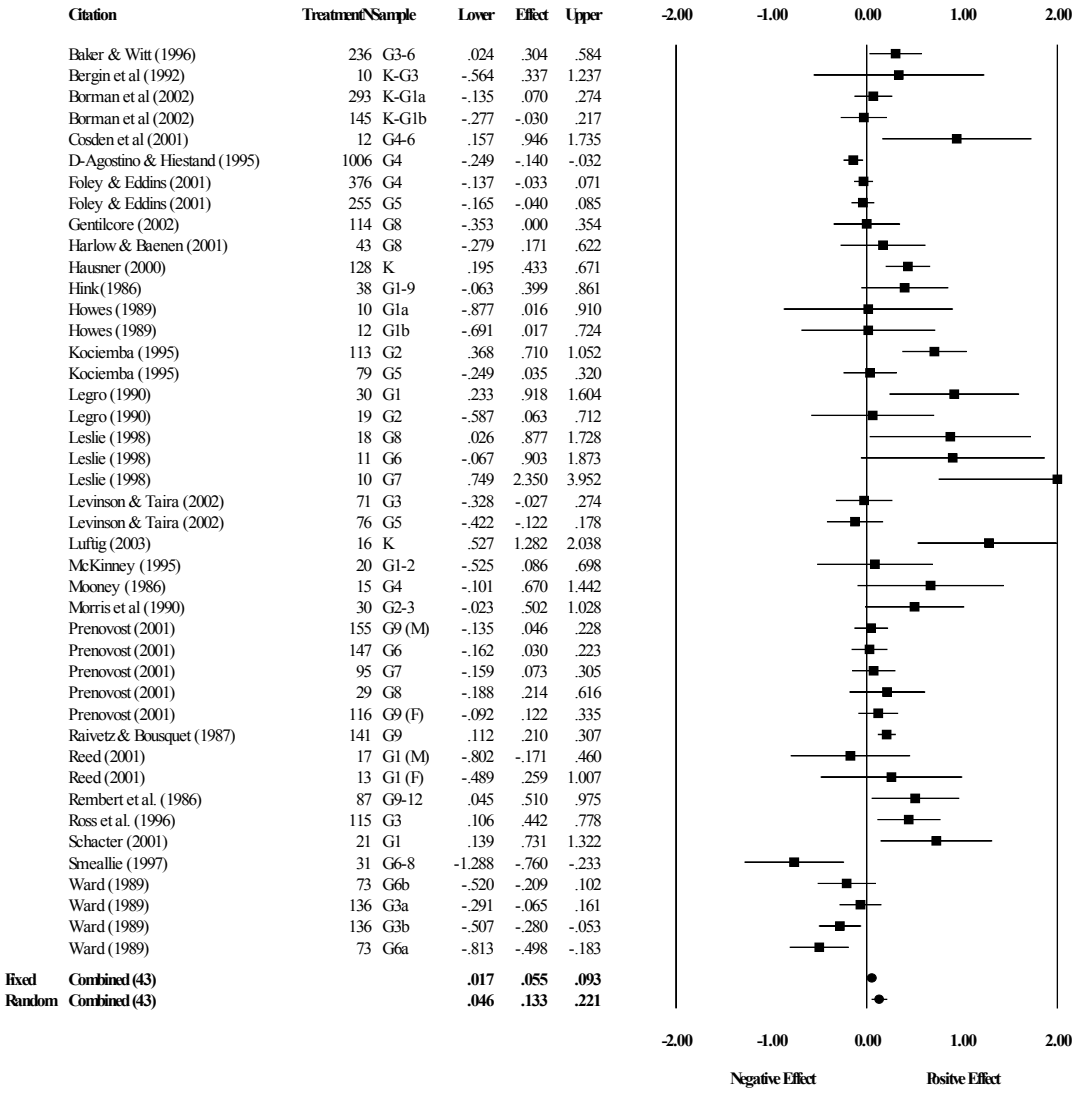
As the display in Table 2.3 indicates, there is an overall tendency toward positive effects of OST strategies for improving the reading achievement of low-achieving or at-risk students. The overall effect size based on a fixed-effects model is .06, and the overall effect size based on a random-effects model is .13.[12] The 95 percent confidence intervals around these effect sizes do not include zero, which supports the conclusion that the OST strategies examined through this meta-analysis had a significantly positive effect on the reading achievement of at-risk students ($p < .05$).

[11] Although gender is not a moderator, we indicated gender in the table if the data for calculating effect sizes were available only at this level.

[12] The two effect sizes are different because weighting by sample size has less impact in the random-effects model compared to the fixed-effects model (Cooper et al., 2000).

The homogeneity analysis resulted in a Q value of 103.7, which is statistically significant ($p < .0001$). This indicates that the variation among the effect sizes is significantly more than expected by sampling error alone. Therefore, we conducted additional analyses based on identified moderators in order to explain the variation among the effect sizes.

Table 2.3. Effectiveness of OST Strategies for Improving Student Achievement in Reading



Program Characteristics as Moderators of Effect Sizes of OST Strategies for Reading

We analyzed five program characteristics for influences on the overall effect size previously reported: (1) timeframe, (2) grade level (3) activity focus, (4) program duration, and (5) student grouping. Table 2.4 presents the average effect sizes for these five moderators weighted by the sample size. The table reports the total number of effect sizes analyzed for each moderator, which depended on the unit of analysis and whether there was sufficient information to code the study for the moderator. The unit of analysis for the moderator of grade level was the effect sizes of independent samples of students at the different grade levels. The unit of analysis for all other moderators was the overall effect size of the study.

Table 2.4. Program Characteristics as Moderators of Effect Sizes of OST Strategies for Reading

Moderator	*k*[a]	*Q*	Effect Size[b]	95% Confidence Interval	
				Lower Bound	Upper Bound
OST Timeframe		1.08			
After school	14		.12	.04	.20
Summer school	12		.07	.01	.13
Summer & Saturday school	1		.17	-.28	.62
Grade Level[c]		40.65**			
Lower elementary (K–2)	14		.26	.16	.37
Upper elementary (3–5)	13		-.04	-.10	.01
Middle (6–8)	13		.01	-.07	.10
High (9–12)	2		.22	.13	.32
Focus		2.30			
Academic	20		.12	.06	.17
Academic + Social	7		.04	-.05	.12
Duration		16.45**			
<44 hrs	7		.02	-.14	.18
44–84 hrs	7		.25	.16	.34
85–210 hrs	5		.19	.06	.32
>210 hrs	3		-.01	-.11	.09
Grouping		12.30**			
Large group (11 or more)	6		.16	.08	.25
Small group (10 or less)	5		.04	-.05	.14
One-on-one tutoring	5		.50	.21	.80
Mixed	7		.24	.10	.38

[a] Number of effect sizes included in the analysis
[b] Fixed-effects model
*$p < .05$ ** $p < .01$
[c] The unit of analysis (k) for this moderator is within-study effect sizes — one or more per study.

In Table 2.4, when the 95 percent confidence interval does not include zero, the average effect size of the moderator is significantly different from zero. The *Q* statistic examines the amount of variation in the average effect sizes for the different levels of a moderator. A statistically significant *Q* statistic indicates that the moderator accounts for variation among the average effect sizes.[13]

For the moderator of program timeframe, we coded studies as "after-school" (n = 14), "summer school" (n = 12), or "summer and Saturday school" (n = 1). The unit of analysis was the study. The average effect size was .12 for after-school programs and .07 for summer school programs. The one study with both summer and Saturday school had an effect size of .17. The average effect sizes of after-school and summer school programs were significantly greater than zero. Although the effect size of after-school programs was slightly larger than the effect size of summer school programs, based on the *Q* statistic, the overall effect size of OST strategies in reading was not affected by the timeframe of programs. This might indicate the importance of the nature of the strategies used during summer school and after school rather than the timeframe in which they occur.

Ten studies included programs for lower elementary grade students (K–2), ten studies included programs for upper elementary students (3–5), six studies included middle school students (6–8), three studies included high school students, and only one study included a program for students in grades K–9. This one independent sample was omitted from these analyses (Hink, 1986); therefore, results are based on 42 independent samples, which served as the unit of analysis. As indicated in Table 2.4, programs targeting lower elementary students had the largest positive effect size (.26) and a negative effect size was observed for upper elementary students (-.04). There was an average effect size of .22 for high school students, which is larger than the effect size of .01 for middle school students. The 95 percent confidence intervals indicated that the effect sizes for lower elementary and high school students were significantly greater than zero, whereas the effect sizes for upper elementary and middle school students were not significant. These results suggest a possible tendency for OST strategies to be more effective for students at the lowest and highest grade levels. The homogeneity analysis yielded a *Q* value of 40.65 ($p < .0001$), indicating that grade level accounts for some of the variance in the overall effect size estimation.

When we examined the focus of activities in OST programs, we again used the study as the unit of analysis. Twenty programs focused solely on academic enrichment, and seven programs focused on both academic and social enrichment. As shown in Table

[13] We conducted homogeneity analyses of effect sizes based on the fixed-effects model only.

2.4, the average effect sizes were .12 for academic focus and .04 for academic and social focus and only the former was significantly different from zero. The *Q* statistic was not statistically significant, indicating that focus did not influence variation in the overall effect size.

The duration of a program reflected the number of hours students participated in OST strategies. The duration of programs that addressed reading achievement ranged from 9 to 480 hours. The total number of hours for each program was calculated and divided into quartiles of less than 44 hours, 44–84 hours, 85–210 hours, and more than 210 hours. Although we generally assume that a longer implementation of a program produces a larger effect on student achievement, this did not occur in our analysis. The programs with 44 to 84 hours had the largest effect size of .25, followed by programs with 85 to 210 hours for which the effect size was .19. Both of these effects sizes were significantly greater than zero. The effect sizes were -.01 for programs longer than 210 hours and .02 for programs less than 44 hours. The *Q* value of 16.45 indicated a statistically significant influence of program duration on the variation among the effect sizes ($p < .001$).

For studies reporting a grouping structure, five programs worked with students one-on-one, five programs used small groups, six used large groups, and seven used a mix of all three grouping structures. The unit of analysis was the study. Working with students one-on-one had the largest average effect size of .50, and a combination of student grouping structures had the next largest average effect size of .24; both effect sizes were significantly greater than zero. Large-group structures revealed a significant effect size of .16, and placing students in small groups of 10 or less had the smallest effect size of .04, which was not significantly different from zero. Based on the homogeneity analysis, there was significant variation among the effect sizes related to the grouping structures used by OST programs (Q=12.30, $p < .001$).

Study Characteristics as Moderators of Effect Sizes of OST Strategies for Reading

This section presents results from the moderator analysis of study characteristics of study quality, publication type, and score type. The study is the unit of analysis for each moderator. Table 2.5 reports the total number of studies for each moderator category, fixed-effect sizes with confidence intervals, and *Q* values, which indicates the amount of variation among the average effect sizes associated with each moderator. As noted for program characteristics (Table 2.4), when the 95 percent confidence interval does not include zero, the average effect size of the moderator is significantly different from zero.

Table 2.5. Study Characteristics as Moderators of Effect Sizes of OST Strategies for Reading

Moderator	k[a]	Q	Effect Size[b]	95% Confidence Interval	
				Lower Bound	Upper Bound
Study Quality		2.72			
High	2		.11	-.10	.32
Medium	18		.13	.06	.21
Low	7		.05	-.02	.12
Publication Type		10.82**			
Conference paper/report	13		.07	.02	.12
Dissertation	10		.14	.01	.28
Peer-reviewed journal	4		.55	.26	.85
Score Type		4.21			
Gain Score	14		.01	-.08	.10
Posttest Score	13		.12	.06	.18

[a] Number of effect sizes included in the analysis
[b] Fixed-effects model
*$p < .05$ ** $p < .01$

We previously described our approach to reviewing the quality of studies and explained some of the key methodological differences among studies in the three categories of high, medium, and low quality. The only effect size that was significantly different from zero was for medium-quality studies, which had an average effect size of .13. The Q statistic was not statistically significant, indicating that effect sizes were not influenced by study quality.

As indicated in Table 2.5, 13 studies were conference presentations or proprietary reports, 10 were dissertations, and 4 were published in peer-reviewed journals. The studies published in peer-reviewed journals produced the largest effect size of .55, which was significantly greater than zero, as was the much smaller effect size for conference papers (.07). The statistically significant Q value of 10.82 ($p < .01$) indicated that publication type made a difference in the computed effect sizes.

For the moderator of score type, gain scores (or pretest/posttest difference scores) were used to calculate effect sizes for 14 studies, whereas posttest scores were used for the remaining 13 studies. Only the effect size for posttest scores was statistically different from zero, but based on the Q value, score type did not have a statistically significant influence on the effect sizes.

Moderator Relationships

To examine the studies for possible relationships among moderators, we constructed correlation matrices for strategy and study characteristics (Cooper, 1998). Studies of after-school programs reported more one-on-one instruction and mixed-group strategies than studies of summer school, which reported more use of large groups. Studies of after-school programs reported shorter durations (less than 45 hours) than studies of summer schools. The grade level of students in the studies was not related to other moderators. There were no relationships among the studies for the moderators of research quality, publication type, and score type.

NARRATIVE REVIEW OF STUDIES

The 20 studies that were not included in the meta-analysis — because they had insufficient data for calculating effect sizes or because they were qualitative — are discussed in this section. Three of these studies employed a qualitative methodology, and 17 used a quantitative design. Table 2.6 presents characteristics of these 20 studies, including the treatment sample size, grade level(s), timeframe (i.e., summer school, after-school, or Saturday school), focus (i.e., academic only or academic and social), grouping (e.g., large group, small group, one-on-one, or a combination), and student outcome results (i.e., all positive, mostly positive, even, mostly negative, or all negative).

The publication years of the studies presented in Table 2.6 ranged from 1985 to 2003 and included dissertation studies (4), proprietary project evaluations (14), and studies published in refereed journals (2). Treatment sample sizes among these 20 studies ranged from 3 to 147,894. The majority of studies (10) used a variety of student grouping, such as a combination of one-on-one tutoring with large- or small-group instruction. Five interventions used small-group instruction, one used large-group instruction, and four studies did not report student grouping characteristics.

Six studies included participants in elementary school (i.e., K–5), four targeted middle school students (i.e., 6–8), one included high school students, and nine studies focused on students across school levels. Of the six interventions studied for elementary students, three were reported to have mostly positive or all positive results for student learning in reading. Of the four programs targeting middle school students, two were reported to have mostly positive or all positive results for student learning in reading. Of the nine interventions that included students across more than one school level (e.g. elementary and middle school), six were reported to have mostly positive or all positive results for student learning in reading.

Table 2.6. Study Characteristics of Narrative Review Reading Studies

Author(s) and (Year)	Treatment Sample Size	Grade Level(s)	Time Frame	Program Focus	Student Grouping	Results[a]
Branch, Milliner, & Bumbaugh (1986)	752	6th–8th	summer school	academic & social	individualized; self-paced	mp
Duffy (2001)[b]	10	2nd	summer school	academic	small & large group	ap
Grimm (1997)	19	6th–8th	summer school & after school	academic & social	one-on-one mentoring & small group	mn
Hansen, Yagi, & Williams (1986)	871	3rd–7th	summer school	academic & social	missing	mp
Holdzkom (2002)	3,043	3rd–8th	summer school	missing	missing	mp
Huang, Gribbons, Kim, Lee, & Baker (2000)	4,312	2nd–5th	after school	academic & social	large group; one-on-one	mp
Jacob & Lefgren (2001)	147,894	3rd & 6th	summer school	academic	large group; cooperative learning	mp
King & Kobak[b] (2000)	13	7th	summer school	academic &social	large group; cooperative learning	ap
Kushmuk & Yagi (1985)	67	3rd–7th	summer school	academic & social	small group	e
Leboff (1995)	40	3rd	summer school	academic	missing	e
Lodestar Mgmt. Research (2003)	160	2nd–8th	after school	academic & social	varies by site	an
Ortiz (1993)[b]	3	1st	after school	academic	small group	mp
Paeplow, Baenen, & Banks (2002)	116	2nd–8th	summer school	academic	small group; one-on-one tutoring	mn
Phelan (1987)	17	7th & 8th	Saturday school	academic	small group	e
Pyant (1999)	30	K–4th	after school	academic & social	small group	e
Rachal (1986)	9,675	2nd–5th	summer school	missing	missing	mn
Roderick, Engel & Nagaoka (2003)	21,000	3rd, 6th, & 8th	summer school	academic	one-on-one tutoring; small group	ap
Ronacher, Tullis, & Sanchez (1990)	1,072	9th–12th	Saturday school	academic & social	large group	e
Schinke, Cole, & Poulin (2000)	283	5th–8th	after school	academic	one-on-one; large group	mp
Sipe, Grossman, & Milliner (1988)	1,272	8th–10th	summer school	academic & social	small group	mp

[a] Indicates whether the comparisons in the study were all positive (ap), mostly positive (mp), even (e), mostly negative (mn), or all negative (an) (Cooper et al. 2000)
[b] Qualitative study

Twelve studies examined summer school programs, five researched after-school programs, two involved Saturday school, and one studied a combination (summer school and after school). Of the five *summer school* programs that focused *solely on academics*, three found mostly positive or all positive results of the intervention on

student reading. Of the six *summer school* programs that focused on *academics and social skills*, four found mostly positive or all positive results of the intervention on student reading. The two *after-school* programs that focused *solely on academics* found mostly positive or all positive results of the intervention on student reading. Of the three *after-school* programs that focused on *academics and social skills*, only one found mostly positive or all positive results of the intervention on student reading. The Saturday school programs, one focusing on solely academics and one emphasizing academics and social skills, reported even results; that is, there were about the same number of significant and non-significant on student learning for treatment groups in comparison to control groups. Two studies of summer school programs did not report enough information to determine the program focus, although one study reported mostly positive results and the other reported mostly negative results.

COMMON FEATURES HIGHLIGHTED IN STUDIES

The 47 studies included in this synthesis examined OST strategies that vary in their approaches to improving students' reading skills. However, after reading and rereading the studies, we found that many of them shared features that program implementers highlight as critical components of their OST strategies. This section supplements our meta-analysis results with summaries of studies that best exemplify some of these common features. These studies also give examples of the OST programs that informed our results.[14]

Linking Attendance to Achievement

Some of the programs in this synthesis emphasized the theory that more time on task will result in higher student performance. As a result, these programs focused on improving student engagement in learning in hopes that their attendance in school and in the OST programs will increase. Incentives for attending and participating in OST programs included paid wages (Branch, Milliner, & Bumbaugh, 1986), game-like cooperative learning activities (King & Kobak, 2000), and token-based economies (Leslie, 1998).

Baker and Witt (1996) evaluated two after-school programs in Austin, Texas, and concluded that the after-school program had greater impact on those students who participated more often. The OST strategies employed by this after-school program

[14] Unless otherwise noted, the effect sizes reported in this section are from the meta-analysis described in Table 2.3.

were aimed at increasing student interest and engagement in learning by presenting academically oriented activities in the context of a goal-oriented, fun, recreational experience. According to the authors, through quality contact time with students, program staff provided students with a positive use of their free time after school and increased engagement in learning activities. (The study had an effect size of .30.)

The LA's Best after-school enrichment program, evaluated by Huang, Gribbons, Kim, Lee, and Baker (2000), was based on the theory that attendance predicts performance and that more time on learning tasks results in higher levels of performance. In order to encourage student attendance, the program integrated homework assistance and recreational, social, and motivational activities in a safe environment for second through fifth graders. Based on a sample of cohorts, the authors found that over time, the students with the highest level of participation in LA's Best continued to demonstrate increased school attendance and increased standardized test scores.

Ensuring Staff Quality

Many of the synthesis studies did not report the qualifications of those implementing the program, although some of the programs included a training component, especially when volunteers were used as tutors. In their study of the Howard Street Tutoring Program for low-achieving second and third graders, Morris, Shaw, and Perney (1990) noted that a critical component of the program was the quality of the supervisor. This OST strategy is implemented by volunteer tutors using specific reading strategies including shared reading, word study, reading books, and writing stories. The researchers stated that for effective implementation, the supervisor of tutors must possess the following:

> (1) theoretical knowledge of the beginning reading process, (2) experience in teaching beginners how to read, (3) confidence . . . that almost all children can learn to read and write, and (4) an ability to work constructively with adults in a mentor/apprentice relationship. (p. 148)

Tutored children experienced learning gains as a result of the program (d = .50), but the researchers emphasized that learning gains did not occur until 50 hours of "well-planned, closely supervised one-to-one tutoring" (p. 147). At the middle school level, King and Kobak's (2000) study found that supervision from content-specific lead teachers was key to ensuring instructional quality in the summer academy program.

Duffy (2001) evaluated a summer school program for underachieving second graders that used a balanced, accelerated, and responsive approach to literacy instruction. Duffy's evaluation employed qualitative methodology and was rated as high in research quality for this synthesis. Duffy emphasized that both the teacher and the reading program are key to ensuring that all children learn to read well. Responsive teaching involves teachers making modifications to program components according to the assessed needs of their students on a daily basis. Duffy's research showed that responsive teaching included not only meeting students' cognitive needs but also their behavioral and emotional needs based on the premise that when students feel safe and valued, they are more willing to take risks in literacy learning. The researcher found that many of the students in the program made significant progress in the areas of word identification and fluency.

Developing Academic and Social Skills

The National Institute on Out-of-School Time suggested that interventions that focus on social and behavioral skills also provide expanded opportunities in which literacy skills can develop (Hynes, O'Connor, & Chung, 1999). Some of the studies included in this synthesis recognized the link between academic and human development and therefore addressed the social and emotional needs of students in addition to providing academic instruction (Foley & Eddins, 2001; Schacter, 2001; Legro, 1990; Pyant, 1999). Schacter (2001) evaluated the impact of an eight-week, summer day camp that promoted social and emotional growth implementing a systematic reading curriculum with one-on-one tutoring and recreational activities. The purpose of the camp, which was designed for disadvantaged children, was to turn first graders' reading losses into gains. The treatment group showed significant reading improvement compared to control students (d = .73). The author identified the summer camp context as instrumental to the success of the program.

Implementing a Well-Defined Reading Curriculum

The structure of the curriculum in Hausner's (2000) study of the Project Accelerated Literacy (PAL) included eight major components of literacy instruction based on a constructivist approach and scaffolded learning: read aloud to children, shared reading, guided reading, independent reading, modeled writing, shared writing, guided writing, and independent writing. Features of the PAL program included (1) a small class size, (2) a variety of learning centers that integrate literacy tools and tasks (e.g., play office, art center, cooking, and book corner); (3) a two-hour block of time for literacy instruction through large-group, small-group, and individual instruction; (4) teaching practices based on each student's performance on standards; (5) scaffolded teaching that follows a pattern of modeling, guiding, observing, and

practicing skills for students; and (6) a thematic curriculum (e.g., foods, sea life, and community helpers) reflected in each activity center. As a result of this 30-week, half-day program, at-risk kindergarten participants experienced gains in literacy learning compared to their peers in the control group (d = .43).

Ross, Lewis, Smith, and Sterbin (1996) evaluated The After-School Tutoring Program for second through fourth graders in 13 Title I schools in Memphis, Tennessee. This OST intervention used a curriculum modeled on strategies from the Success For All program and was offered three days a week, one hour a day throughout the school year. Components included Story Telling and Retelling (StaR), Listening Comprehension, reading and follow-up activities with tradebooks from the *Scott Foresman Book Festival* kits, writing, book club, computer skills, and test-taking strategies. Participants showed gains in reading achievement compared to a matched control group (d = .18).

Bergin et al. (1992) evaluated the Hilltop Emergent Literacy Project (HELP), an after-school intervention program for educationally disadvantaged students in kindergarten through the third grade. Serving mostly African-American participants, the program used culturally sensitive teaching methods and materials to implement a phonics-based curriculum. Features of the HELP program included (1) a favorable teacher student ratio with volunteers from a local university teacher preparation program; (2) an emphasis on promoting social connectedness by providing students with extra attention and emotional support; (3) stimulating intrinsic interest by using a curriculum that gives students learning choices; (4) a mix of independent, small-group, and large-group literacy activities; and (5) using the Sing, Spell, Read & Write curriculum, which encourages singing and movement as part of the learning process. As a result of participating in the HELP program for 16 months and six hours a week, students performed better in reading than their peers in control groups (d = .34).

We found evidence of the effectiveness of a well-defined curriculum and structured approach for both elementary and secondary grade levels. Hink (1986) evaluated a structured summer school program for students in grades 1 to 9 (d =.40). Summer school began with placement tests to give teachers direction in their instruction; learning objectives were identified for each student and progress was evaluated at the end of the summer school through posttesting. Rembert, Calvert, and Watson (1986) evaluated a summer school for 10^{th}- 11^{th}-, and 12^{th}-grade students with "evidence of college level academic potential, but low motivation or intention toward postsecondary education" (p. 376). The summer school provided college preparation classes that focused on skill mastery in basic academics and simulated college instruction. Compared to the control group, participants in this summer school scored significantly higher on the reading portion of the Comprehensive Test of Basic Skills (d = .51).

Preventing Learning Loss and Sustaining Gains

Some studies aimed at closing the achievement gap indicated that OST strategies can be effective at preventing learning loss, especially during the summer months (e.g., Branch et al., 1986; Borman, Rachuba, Fairchild, & Kaplan, 2002; Sipe, Grossman, & Milliner, 1988). In particular, Borman et al. (2002) reported evidence of a cumulative impact on learning of students participating in summer school over a period of two and three years, although in some cases, poor multi-year attendance rates might have accounted for declines in treatment effects. The authors suggested that cumulative benefits of summer school programs over time prevent low-achieving students from experiencing the "summer slide," whereby they fall behind their peers in reading ability.

In contrast, other studies found that students did not sustain learning gains over time. For example, Hausner's (2000) evaluation of an after-school kindergarten literacy program reported that low-performing students' literacy scores increased significantly (d = .40) but that these students did not show sustained improvement in the second grade. The author suggested that at-risk students need more than one literacy intervention to retain the gains made as a result of the early intervention program.

DISCUSSION OF FINDINGS

Based on the overall effect sizes of .06 for the fixed-effects model and .13 for the random-effects model, and given that these are significantly greater than zero, the OST strategies studied in the meta-analysis significantly increased the reading achievement of low-achieving or at-risk students. The results suggest that the positive effect of OST strategies is about one-tenth of a standard deviation. The homogeneity analysis demonstrated a large variation among effect sizes reported by the 27 studies in the meta-analysis. The moderator analysis showed that three program characteristics — grade level of the sample students, program duration, and student grouping — contributed to this variation. Neither program timeframe nor program focus contributed to the variation in effectiveness. In the narrative review, 11 of the 20 studies reported mostly positive or all positive results of OST strategies on student learning in reading. Five studies reported even results, and four studies reported negative results.

As in Cooper et al.'s (2000) meta-analysis of summer school programs, the youngest and oldest students benefited the most from participating in OST strategies in reading. Based on our meta-analysis, the positive effect of OST strategies for low-achieving or at-risk kindergarten through second-grade students was about one-fourth

of a standard deviation (.26). In comparison to lower elementary students, it is interesting to note that upper elementary students (third through fifth grades) experienced the smallest effects, including slightly negative effects. This supports research showing that interventions focused on the prevention of reading disabilities in elementary students are most effective when they are delivered to children very early and before reading problems become intractable and self-esteem issues complicate the learning process (Mathes, 2003). Findings from the narrative review indicated that at least half of the interventions targeting elementary students, middle school students, or a combination of both levels resulted in mostly positive or all positive effects on student learning.

Program duration was a statistically significant moderator. The data indicated that OST strategies had significantly positive effects when implemented for at least 45 hours but less than 210 hours. A program that lasts fewer than 45 hours might not be long enough to influence student achievement in reading, and it might be difficult to sustain the conditions that promote student learning over a longer period of time, as indicated by the negative effect size found for programs longer than 210 hours. This finding is consistent with research included in this synthesis as well as with past research that suggests that positive OST effects on student learning can diminish over time (Cooper et al., 2000; Duffy, 2001; Hausner, 2000; Walker & Vilella-Velez, 1992).

With regard to the program characteristic of student grouping, the use of one-on-one tutoring in OST programs had positive impacts on students' reading performance with an effect size of .50. Of the five studies that used a one-on-one grouping structure as part of the intervention, three studies reported mostly positive or all positive results for student learning in reading. Using a one-on-one grouping structure, tutors or teachers have the best opportunities for assessing individual learning needs and responding to those needs appropriately, which is critical for helping children learn to read. This is consistent with other research that has shown that tutoring, when structured, individualized, and supervised by professional educators, is effective in improving reading (Elbaum, Vaughn, Hughes, & Moody, 2000; Barley et al., 2002).

In addition to these program characteristics, the data on study characteristics showed that publication type explained some of the variation in effect size. The largest effect sizes occurred for articles published in peer-reviewed journals, which is consistent with the notion that journals tend to include studies that report significant rather than non-significant findings (Cooper, 1998).

The meta-analysis of summer school programs conducted by Cooper et al. (2000) reported an effect size of .24 (fixed-effects model) for the effectiveness of remedial

summer programs based on reading outcomes. Although Cooper et al.'s effect sizes provide a context for interpreting the results of our meta-analysis, there are distinct methodological differences between the syntheses. The Cooper et al. meta-analysis included studies that used single group pre- and posttest designs that they cited as possibly inflating the effect size estimates as a result of the unknown influences of history, maturation, and regression to the mean effects. Due to the potential for bias from various study designs, Cooper et al. computed an overall effect size for studies that used random assignment and found that students participating in summer school scored about one-seventh of a standard deviation higher than control group students on outcome measures (an effect size of .14 for both fixed-effects and random-effects models). These results are more consistent with our findings for OST studies on reading achievement, all of which included control or comparison groups.

IMPLICATIONS FOR POLICY AND PRACTICE

Our findings from the 27 studies included in the meta-analysis revealed an overall tendency for positive impacts in reading for low-achieving or at-risk students who participate in OST strategies. This suggests that policymakers and practitioners should consider the use of OST strategies as potentially effective ways of providing students with instruction and related experiences that can help them advance their reading achievement. Based on our review of all the studies in the synthesis that examined reading achievement, some conclusions can be made related to effective practice.

An effective OST strategy for improving the reading of low-achieving students is the use of tutoring and individualized instruction. Reports by Morris et al. (1990) and Leslie (1998) described the characteristics of after-school tutoring programs that had positive effects on reading. OST strategies for reading improvement are particularly helpful for students in the early elementary grades (e.g., Kociemba, 1995; Schacter, 2001). There are other characteristics of successful OST strategies described by researchers of successful programs. OST programs for reading achievement should employ methods to ensure staff quality (Morris et al., 1990) and implement a well-defined reading curriculum, such as the one used by HELP, which Bergin et al. (1992) evaluated. Program implementers also should deliver OST activities in environments that will appeal to at-risk students (Schacter, 2001). Finally, when considering the use of OST to improve reading achievement, policymakers and practitioners should examine other features of programs that this synthesis documented as successful.

CHAPTER 3

Studies of the Effectiveness of Out-of-School-Time Strategies for Mathematics Achievement

The nation's schools are struggling to address the needs of students who are performing below academic standards as well as those who are at risk for failure. In many cases, these students and their specific needs are identified by in-school staff, but teachers are finding that the deficiencies cannot be effectively addressed in the course of the traditional classroom day. One option being leveraged is the use of out-of-school time (OST). Educators see potential in using OST strategies to help their students reach or exceed standards. In essence, OST is being used to provide low-performing students with an opportunity to catch up to their peers.

The OST research encompasses a variety of programs designed primarily for recreation, homework help, mentoring, or programs that infuse teaching with play. These programs take place during a variety of out-of-school timeframes (before or after school, summer school, and Saturday school). The wide variety of OST programs is the result of programmatic creativity in the hands of educators who take advantage of the relative freedom offered outside the traditional school-day schedule. For this reason, the OST program studies are interesting and unique. And, given the potential of OST strategies to meet the needs of low-performing students, careful examination is important.

In this chapter we examine evidence of the effectiveness of OST strategies in assisting low-achieving or at-risk students to meet mathematics standards. As noted in Chapter 1, the complementary concern — the effectiveness of in-school strategies — was addressed by McREL in 2002 (Barley et. al.). In that study, the authors concluded that school-time tutoring and computer-aided instruction strategies were effective in raising the mathematics achievement levels of low-performing students. In this synthesis, however, we analyze OST strategies for mathematics by addressing these research questions:

1. What is the effectiveness of OST strategies in assisting low-achieving or at-risk students in mathematics?

2. How does the effectiveness of OST strategies differ by *program* characteristics such as timeframe, grade level of students, activity focus, program duration, and student grouping?

3. How does the effectiveness of OST strategies differ by *study* characteristics such as research quality, publication type, and score type?

BACKGROUND

The fact that the nation's public schools have not been meeting the needs of at-risk students has been apparent for many years. The widely distributed Coleman Report (Coleman et al., 1966) drew clear comparisons between the low performance of and the lack of appropriate educational experiences provided for at-risk students. The mathematics classroom of the at-risk student in the 1960s was characterized by inadequate curricula and under-prepared teachers.

This inequality continues. In a government study published 26 years after the Coleman Report, Howe and Kasten (1992) identified a list of "variables related to problems of at-risk students in mathematics" (section 2, page 3). The list is strikingly similar to the characteristics revealed by the Coleman Report, including "inappropriate curriculum," "small amount of homework assigned," and "low school academic expectations" (section 2, page 4). In its 1992 *Handbook of Research on Mathematics Teaching and Learning* (Secada, 1992), the National Council of Teachers of Mathematics recognized this continuing disparity and noted that the "...American educational system is differentially effective for students depending on their social class, race, ethnicity, language background, gender, and other demographic characteristics" (p. 623).

In 2000, only a minority of students in the United States achieved at a middle level of performance in mathematics on the National Assessment of Educational Progress (NAEP). The percentages of students who performed at or above a proficient level were 26 percent at grade 4, 27 percent at grade 8, and 17 percent at grade 12 (Braswell et al., 2001). At every grade level, students who were from low-income families, and therefore eligible for free or reduced-price lunch, scored significantly lower in mathematics than students who did not receive this benefit. These statistics indicate the need to improve achievement in mathematics for all students and especially at-risk students.

One step in erasing this inequality can be taken by introducing at-risk students to effective instructional strategies. A number of researchers have been interested in identifying such practices, particularly those that address the needs of at-risk students (see Cooper et al., 2000; Slavin & Madden, 1989). This chapter joins this effort

through a synthesis of recent studies of OST programs to assist at-risk students. The goal of this chapter is to collect, synthesize, and present resulting evidence for the use of effective OST mathematics strategies.

METHODOLOGY

As was the case in Chapter 2, we relied on both meta-analysis and narrative descriptions of studies to address our research questions. (Appendix B describes the meta-analysis methodology.) We addressed our first research question — the effectiveness of OST strategies in mathematics —with the computation of overall effect sizes based on fixed- and random-effects models, which are presented with 95 percent confidence intervals. We addressed the second and third research questions by computing average effect sizes for each moderator characteristic. We conducted homogeneity analyses to determine whether the average effect sizes differed significantly by moderator characteristics more than would be expected by sampling error alone. Finally, we reviewed studies that examined unique features of OST strategies or employed special conditions.

Study Selection

Based on the literature searches described in Chapter 1, we collected studies that reported the effectiveness of OST strategies in improving the mathematics achievement of low-performing or at-risk students. There were 33 studies that met the inclusion criteria described in Chapter 1.

All 33 studies employed quantitative methods to examine the effects of OST strategies. Of these, 22 provided enough information to compute effect sizes for a meta-analysis. The other studies were examined in a narrative review and are included in the current report in a narrative description, along with studies in the meta-analysis that had important characteristics.

Data Analysis

As described in Chapter 1, the studies were reviewed by two to four researchers, and the coding reliability was examined for several studies to check for consistency. (The coding instrument is provided in Appendix A.) To address our first research question on the effectiveness of OST strategies in mathematics, we computed the overall effect size using both fixed- and random-effects models. The overall effect sizes were based on 33 independent samples from 22 studies that reported enough information to calculate effect size estimates. The effect size weighted by sample size (weighted

d) was calculated for each independent sample (see Appendix B). The number of independent samples from a single study varied from one to three. We used a 95 percent confidence interval around the overall effect size of each sample to determine if the effects of the OST mathematics strategies were significantly greater than zero.

Four program characteristics were examined as moderators in order to address the second research question: (1) strategy timeframe, (2) grade level of students, (3) program duration, and (4) activity focus. In contrast to the studies in Chapter 2 on OST strategies for reading, the studies addressing OST for mathematics did not include sufficient information to examine student grouping as a strategy characteristic. In an effort to answer the third research question, we analyzed three study characteristics: (1) research quality, (2) publication type, and (3) type of score.

The moderator analysis of timeframe was conducted with studies as the unit of analysis. To examine how the timeframe of OST strategies might explain differences in effect sizes among the different studies, we computed the average effect sizes for two main types of OST timeframes: after-school programs and summer schools. There are other timeframes in the studies that we address through narrative review.

Because several of the studies examined OST effects on children in different grades, the moderator analysis of student grade level was conducted with independent samples as the unit of analysis. We coded grade level of students using four categories: lower elementary (K–2), upper elementary (3–5), middle school (6–8), and high school (9–12). The grade levels of two independent samples overlapped all four categories, so these data were excluded from this analysis. When an independent sample overlapped two categories, we chose a category where the majority of the students' grade levels were applicable. For example, the Baker and Witt (1996) study included students in grades 3 through 6, so the study was assigned to upper elementary (3–5) rather than middle school (6–8).

The analysis of activity focus was conducted with the study as the unit of analysis, as were the moderator analyses of the remaining characteristics. For each study, the activity focus was coded as "academic" or "academic and social." Those studies in which the OST strategy focused almost solely on academic enrichment in mathematics, including homework assistance, study skills, and remedial lessons, were coded as "academic." The studies in which the reported OST strategy focused not only on academic enrichment, but also on social enrichment including music, art, social skills, and recreational activities, were coded as "academic and social."

We determined the total hours of treatment in a review of each study; this value was in turn coded as the strategy duration. We then assigned studies to one of four categories: 45 hours or less, 46 to 75 hours, 76 to 100 hours, and more than 100

hours. Seven of the studies did not report sufficient information to compute the total hours; thus they were excluded from this analysis.

We analyzed two publication characteristics: study quality and publication type. Study quality was coded as high, medium, or low based on the quality indicators the project team developed. The studies also were categorized by their publication types: conference paper/report, dissertation, or peer-reviewed journal. An additional study characteristic coded was the type of score reported — gain score or posttest score.

OVERVIEW OF STUDIES

Table 3.1 describes the 33 studies that composed the body of research on OST strategies to assist low-achieving or at-risk students in mathematics. Similar to the research on OST strategies for reading presented in Chapter 2, the studies that addressed mathematics achievement represented a variety of programs. Study completion dates were from 1985 to 2003; seven of the 33 studies were completed in 2001 or later. The treatment samples ranged in size from small to large, and all the studies used a quantitative approach. The programs were implemented using various timeframes: 17 were implemented only during the summer, 12 only after school, 1 only on Saturdays, and 3 used a combination of times, including one program that was implemented before and after school (Finch, 1997). Nearly half of the programs studied appeared to focus solely on academics, but some authors omitted intervention descriptions, making it difficult to accurately count these programs. Recreation, arts programming, life skills, and mentoring were components of the programs that combined academics with other emphases.

Eleven of the studies presented in Table 3.1 do not report data sufficient to calculate effect sizes, so they are not included in the meta-analysis results in this chapter. Descriptions of the studies excluded from the meta-analysis are provided in the narrative review section. It is important to note here that these two groups of studies did not differ greatly on study characteristics such as grade level(s), timeframe, program focus, or grouping strategies.

Table 3.1. Studies of Out-of-School-Time (OST) Mathematics Strategies

Author(s) and (Year)	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Baker & Witt (1996)	302	3rd–6th	low SES[a]	After-school recreation programs in which certified teachers facilitate a variety of activities from recreation to academics	after school
*Branch, Milliner, & Bumbaugh (1986)	752	8th–9th	low SES	STEP (Summer Training and Education Program) combined an existing federal work program with drop-out prevention strategies	summer school
Cosden, Morrison, Albanese, & Macias (2001)	90	4th–6th	low performing	Homework time and support	after school
D'Agostino & Hiestand (1995)	1,006	4th	low performing	Academic focus emphasizing higher order thinking, questioning, and problem-solving skills	summer school
Finch (1997)	35	7th	low SES	Computer-assisted instruction sessions designed to supplement students' mathematics curriculum	before and after school
Grimm (1997)	19	5th–8th	low SES	Residential summer program with follow-up mentoring from shipyard workers; summer school and follow-up activities included academic classes to support or remediate skills, dinners with mentors, and field trips	summer and after school
Hansen, Yagi, & Williams (1986)	871	3rd–7th	not promoted	Arts-and-crafts and academic remediation program for public school students in Portland, Oregon (see also Kushmuk & Yagi, 1985)	summer school
Harlow & Baenen (2001)	86	7th	low SES	North Carolina program stressing academics and life skills - students are taught in small groups by exemplary high school and college students	summer and Saturday school
Hink (1986)	48	1st–9th	educator identified	Program providing remedial classes in reading and math, teacher- directed, large-group instruction	summer school
Huang, Gribbons, Kim, Lee, & Baker (2000)	4,312	2nd–5th	low SES and low performing	Program providing homework assistance as well as field trips and other enrichment to students in Los Angeles, California	after school
Kociemba (1995)	192	2nd & 5th	low performing	Compensatory programming in preparation for re-take of Minnesota State reading and math tests	summer school
Kushmuk & Yagi (1985)	67	3rd–7th	not promoted	Arts-and-crafts and academic remediation program for public school students in Portland, Oregon (see also Hansen, Yagi, & Williams, 1986)	summer school
LeBoff (1995)	40	3rd	low performing	Remedial Chapter 1 program for urban youth - no specific program description was provided	summer school

Author(s) and (Year)	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Legro (1990)	49	1st–2nd	low SES	One-on-one homework tutoring; parent involvement, partnership program; social & communication skills component	after school
Leslie (1998)	39	6th–8th	low performing	Program combining tutoring and computer-assisted instruction	after school
Lodestar Mgmt. Research (2003)	160	2nd–8th	low performing	Program designed to fill after-school time with constructive activity including reading, writing, and recreation	after school
McKinney (1995)	47	1st–2nd	low performing	One-on-one tutoring program; self-concept and non-academic enrichment component	after school
McMillan & Snyder (2002)	90	9th	low performing	Remedial program aimed to assist students in passing Virginia State tests	summer school
Paeplow, Baenen, & Banks (2002)	116	2nd–8th	low performing	Leadership program utilizing tutoring and cooperative learning	summer school
Prenovost (2001)	271	6th–8th	low performing	Homework support, enrichment, field trips, and sports	after school
Pyant (1999)	30	K–4th	low SES	Tutoring and social skills instruction program	after school
Rachal (1986)	9,675	2nd–5th	low performing	Compensatory/remedial program in Louisiana	summer school
Raivetz & Bousquet (1987)	136	9th	low SES	Tutoring program and also large group instruction	summer school
Rembert, Calvert, & Watson (1986)	87	10th–12th	educator identified	Remedial program for high school students on a college campus with computer-assisted instruction	summer school
Riley (1997)	78	9th–12th	low SES	Remedial program for high school students on a college campus	summer school
Ronacher, Tullis, & Sanchez (1990)	1,072	9th–12th	low performing	Study and life skills program	Saturday school
Schinke, Cole, & Poulin (2000)	283	5th–8th	low SES	Compensatory program with a variety of components including mentoring, writing activities, reading for enjoyment, and cognitive games	after school
Sipe, Grossman, & Milliner (1988)	1,272	5th–7th	low SES and low performing	A work-study program providing basic skills remediation (including silent sustained reading and computer-assisted instruction) and life skills instruction; includes data from five urban demonstration sites	summer school
Smeallie (1997)	31	6th–8th	low performing and educator identifier	Tutoring program encouraging homework completion	after school
Ward (1989)	175	3rd & 6th	low performing	Teacher-directed instruction with an emphasis on minimal skill achievement; no basals allowed, hands-on activities instead	summer school

Author(s) and (Year)	Treatment Sample Size[b]	Grade Level(s)	Student Description	Strategy Description	Time Frame
Weber (1996)	29	3rd–6th	low performing	Rural program - no specific intervention description was provided	summer school
Welsh, Russell, Williams, Reisner, & White (2002)	3,780	K–8th	low SES	Large-scale New York City program	after school
Zia, Larson, & Mostow (1999)	1,863	3rd–5th	low performing	Math Power program designed to remediate and build confidence in mathematics students	summer school

[a] SES: socio-economic status
[b] The n for the meta-analysis could be smaller based on the data available to calculate effect sizes.
*Studies rated as "high" based on quality indicators used for this synthesis.

All the studies included a measure of student learning in mathematics, as required for inclusion in the synthesis. Of the 33 studies with mathematics outcomes, 23 reported aggregated mathematics scores from standardized assessments, and 10 employed other outcome measures including teacher grades, end-of-grade tests, and researcher-designed assessments.

Eight of the 33 mathematics studies used random assignment to treatment and control groups. None matched groups with a pretest, but 16 of the studies matched groups using other criteria such as demographics, and nine studies did not report any matching. For the 22 studies included in the meta-analysis, we computed effect sizes based on 10 studies that reported pretest-posttest differences or gain scores and 12 studies that reported only posttest scores.

Table 3.1 illustrates the variety of students targeted by the programs studied. The body of research covers the complete range from kindergarten through 12th grade. The distribution of targeted grades included a considerable number of studies at each of the levels: lower elementary (n = 10), upper elementary (n = 20), middle school (n = 19), and high school (n = 7). There is, however, a notable concentration of research in the lower grades (8th and below). The student descriptions provided by research authors also varied; although, in each case, the students were in some way identified as being at risk for academic failure. As described in Chapter 1, this was indicated through some measure of low performance or through identification of the student participants as members of low-SES families.

The nature of the OST programming varied greatly among the studies. For example, there were after-school programs of short duration and day-long programs that filled the summer months. In fact, the large differences among the time spans for interventions encouraged us to examine the total amount of program time (program duration) as a program moderator. The mathematics programs studied here ranged in total time from a six-week after-school program that had 12 total hours duration

(Smeallie, 1997) to 525 hours in a longitudinal study of an extended after-school intervention (Welsh, Russell, Williams, Reisner, & White, 2002). Twenty-two of the 33 studies provided enough information to determine the strategy duration statistic for the program studied. The median strategy duration was 82 hours for these 22 studies.

Research Quality Review

As described previously, we coded 33 studies for their quality. The results are presented in Table 3.2. It should be noted that the inclusion criteria are sufficiently rigorous such that all 33 of these studies can be considered informative. As noted in Chapter 2, we had hoped to find reports that included thorough descriptions of the interventions, discussion of fidelity measures, use of comparable treatment and control groups, concern for potential effects caused by concurrent events, appropriate target participants, and accurately estimated and reported effect sizes. However, only one of the 33 studies that addressed mathematics outcomes did all of these things, while others omitted treatment descriptions, neglected to report important statistics, or in some other way made it difficult for us to determine the nature of the relationship between the reported intervention and performance results.

Table 3.2. Ratings of Mathematics Studies Based on Quality Indicators

Methodology	Rating	Number of Meta-analysis Studies	Number of Narrative Review Studies	Total Number of Studies
Quantitative	High	1	-	1
	Medium	12	5	17
	Low	9	6	15
Total		22	11	33

META-ANALYSIS RESULTS

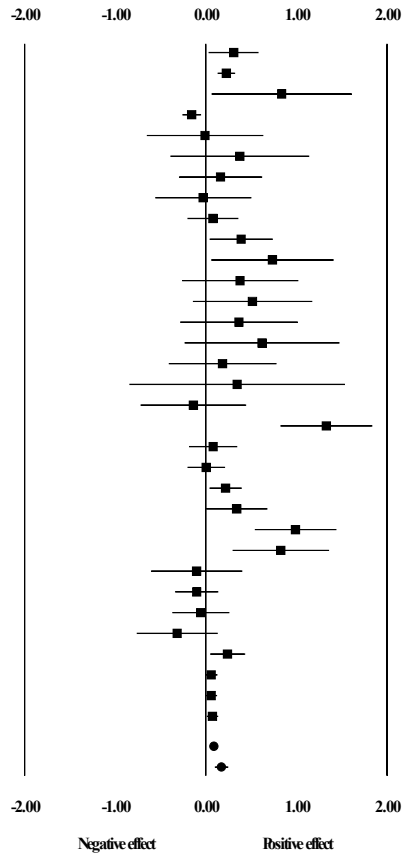
The following section describes the results of our meta-analysis. The process, introduced previously and described at length in Appendix B, includes both a meta-analysis for overall effect size and an examination of moderator effects. The meta-analysis was conducted on a subset (22 studies) of the available research because in the other studies, authors did not include enough information to compute effect sizes.

Overall Effect Size of OST Strategies in Mathematics and Homogeneity Analysis

Our first research question concerns the effectiveness of OST strategies in assisting low-achieving or at-risk students in mathematics. To answer the question, we started with 33 effect sizes based on 33 independent samples from 22 studies. These effect sizes are presented in Table 3.3, which shows the graphic distribution of the effect sizes along with the size of the sample and sample characteristics. The graph illustrates a tendency toward positive effects of OST strategies for improving the mathematics achievement of at-risk students. The overall effect size based on a fixed-effects model was .09 and the overall effect size based on a random-effects model was .17[15]. The confidence intervals around these effect sizes do not include zero,

Table 3.3. Effectiveness of OST Strategies for Improving Student Achievement in Mathematics

	Citation	Treatment N	Sample	Lower	Effect	Upper
	Baker & Witt (1996)	236	G3-6	.027	.307	.587
	Branch (1986)	752	G8-9	.126	.227	.329
	Cosden et al. (2001)	12	G4	.058	.837	1.617
	D-Agostino & Hiestand (1995)	1006	G4	-.264	-.156	-.048
	Finch (1997)	23	G7 (F)	-.656	-.008	.639
	Finch (1997)	12	G7 (M)	-.395	.375	1.146
	Harlow & Baenen (2001)	41	G7	-.301	.162	.625
	Hink (1986)	28	G1-9	-.564	-.028	.508
	Kociemba (1995)	79	G2	-.206	.078	.363
	Kociemba (1995)	42	G5	.036	.391	.746
	Leboff (1995)	19	G3 (F)	.053	.736	1.418
	Leboff (1995)	20	G3 (M)	-.268	.379	1.025
	Legro (1990)	30	G1	-.148	.515	1.179
	Legro (1990)	19	G2	-.289	.366	1.022
	Leslie (1998)	18	G8	-.241	.621	1.482
	Leslie (1998)	11	G6	-.415	.185	.786
	Leslie (1998)	10	G7	-.848	.346	1.540
	McKinney (1995)	23	G1&2	-.726	-.138	.451
	McMillan & Snyder (2002)	90	G9	.818	1.331	1.844
	Prenovost (2001)	116	G6-8 (F)	-.188	.081	.351
	Prenovost (2001)	155	G6-8 (M)	-.208	.005	.218
	Ravietz & Bousquet (1987)	136	G9	.034	.219	.404
	Rembert et al. (1986)	87	G10-12	-.003	.340	.683
	Riley (1997)	55	G9-12 (F)	.535	.990	1.446
	Riley (1997)	23	G9-12 (M)	.290	.827	1.364
	Smeallie (1997)	31	G6-8	-.610	-.102	.407
	Ward (1989)	108	G3	-.344	-.101	.143
	Ward (1989)	67	G6	-.374	-.055	.265
	Weber (1996)	29	G3-6	-.768	-.316	.136
	Welsh et al. (2002)	183	K-9	.041	.240	.438
	Zia et al (1999)	809	G5	-.011	.061	.133
	Zia et al (1999)	916	G3	-.007	.061	.129
	Zia et al (1999)	947	G4	.007	.074	.141
fixed	**Combined (33)**			**.059**	**.090**	**.121**
Random	**Combined (33)**			**.095**	**.174**	**.253**



[15] The two effect sizes are different because weighting by sample size has less impact in the random-effects model compared to the fixed-effects model (Cooper et al., 2000).

which indicates that the effectiveness of OST strategies on mathematics outcomes is statistically greater than zero. No statistical outliers were identified among the effect sizes. (See Appendix B for a description of the outlier analysis.)

The homogeneity analysis resulted in a Q value of 107.4, which was statistically significant (p <.0001). This indicated that the variation among the effect sizes was significantly more than expected by sampling error alone. Therefore, we proceeded with additional analyses to identify moderators that might explain the variation.

Program Characteristics as Moderators of Effect Sizes of OST Strategies for Mathematics

Table 3.4 presents average effect sizes weighted by the sample sizes within each level of four moderator variables: timeframe, grade level, program duration, and activity focus. The table reports the total number of effect sizes analyzed for each moderator, which depended on the unit of analysis and whether there was sufficient information to code the study for the moderator. The unit of analysis for the moderator of grade level was the effect sizes of independent samples of students at the different grade levels. For all other moderators, the unit of analysis was the overall effect size of the study. In Table 3.4, when the 95 percent confidence interval does not include zero, the average effect size of the moderator is significantly different from zero. The table also includes Q values for homogeneity analyses among the effect sizes for each moderator.[16]

The average effect sizes of both after-school programs and summer school programs were significantly greater than zero. However, the Q statistic was not statistically significant, indicating that the overall effect size of OST strategies for mathematics was not influenced by timeframe. This might indicate the greater importance of program features other than the timeframe in which OST strategies were delivered.

Regarding the analysis of student grade level, two studies were excluded due to overlapping grade levels (Hink, 1986; Welsh et al., 2002). Among the remaining studies, three analyzed programs that served lower elementary grade students, eight studies reported programs implemented for upper elementary grade students, another seven studies were of middle school students, and the remaining four studies involved high school students. The effect sizes varied from .05 to .44. The largest effect size was observed for high school students, followed by the effect size for middle school interventions, and then lower elementary school interventions. The

[16] We conducted homogeneity analyses of effect sizes based on the fixed-effects model only.

upper elementary interventions reported the smallest overall effect size. The 95 percent confidence intervals indicated that the effect sizes for middle school and high school students were significantly greater than zero, whereas the effect sizes for lower and upper elementary grade students were not significantly greater than zero. Thus, the results suggest a possible tendency for OST strategies to be more effective for students in the higher grades. The *Q* value of 32.79 was statistically significant ($p < .0001$), which indicated that the grade level accounts for some of the variance in the overall effect size.

Table 3.4 Program Characteristics as Moderators of Effect Sizes of OST Strategies for Mathematics

Moderator	*k*[a]	*Q*	Effect Size[b]	95% Confidence Interval	
				Lower Bound	Upper Bound
OST Timeframe		.52			
After school	8		.13	.01	.25
Summer school	12		.09	.04	.13
Summer and Saturday school	1		.16	-.30	.63
Grade Level[c]		32.79**			
Lower elementary (K–2)	4		.13	-.09	.35
Upper elementary (3–5)	11		.05	.01	.08
Middle (6–8)	11		.16	.08	.24
High (9–12)	5		.44	.30	.59
Focus		10.36**			
Academic	18		.06	.01	.11
Academic + social	4		.23	.14	.33
Duration		10.73*			
≤ 45 hrs	4		.06	-.01	.13
46-75 hrs	4		.26	.11	.41
76-100 hrs	4		.22	.13	.32
>100 hrs	3		.11	-.02	.25

[a] Number of effect sizes included in the analysis
[b] Weighted *d*, fixed-effects model
[c] The unit of analysis (*k*) for this moderator is within-study effect sizes — one or more per study.
* $p < .05$ ** $p < .01$

When we looked at the activity focus in OST programs, the OST strategies reported by 18 studies were primarily academic, and four studies reported OST strategies focused on both academics and social enrichment. The average effect sizes for studies with academic focus or academic and social were .06 and .23, respectively

and both were significantly greater than zero. The Q value of 10.36 ($p < .01$) indicated a statistically significant influence of strategy focus on effect size.

Regarding strategy duration, programs with a duration of 46–75 hours had the largest effect size (.26), followed by 76–100 hours (.22), and a duration of more than 100 hours (.11). The smallest effect was produced from programs that lasted for 45 hours or less (.06). Interestingly, the effects of the programs with durations of 45 hours or less and more than 100 hours did not significantly differ from zero. However, there was statistically significant variation among different strategy durations based on the Q value of 10.73 ($p < .05$). The data indicate that OST strategies were effective for mathematics when implemented for at least 46 hours but less than 100 hours. A duration of 45 hours or less might not be long enough to have a significant effect on student achievement. The small effect from implementations of more than 100 hours might be due to lower student attendance, although there are no data to confirm this.

Study Characteristics as Moderators of Effect Sizes of OST Strategies for Mathematics

The previous analyses revealed three program moderators that explained variation across effect sizes. We also conducted analyses to determine whether study characteristics influenced effect sizes, as indicated in Table 3.5. As noted for program characteristics (Table 3.4), when the 95 percent confidence interval does not include zero, the average effect size of the moderator is significantly different from zero.

The one study in our body of research that was coded as high in quality produced the largest effect size of OST strategies on mathematics achievement (.23), followed by the effect sizes for medium-quality (.10) and low-quality (.01) studies. Although the high- and medium-quality studies reported significantly positive effects, the effect sizes reported by low-quality studies were not significantly greater from zero. Study quality was a statistically significant moderator of effect size as indicated by the Q value of 10.77 ($p < .01$). This result confirms the positive effects of OST strategies for mathematics achievement as evidenced by the higher quality studies in our review.

The average effect sizes reported for conference papers and other reports were significantly positive compared to the effect sizes reported for dissertations and peer-reviewed journal publications, which were not significantly different from zero. However, publication type did not statistically influence the overall effect size as indicated by the Q value for this moderator. For the moderator of score type, gain scores (or pretest/posttest difference scores) were used to calculate effect sizes for 10 studies, while posttest scores were used for the remaining 12 studies. Only the effect

size for gain scores was significantly different from zero, and the *Q* value indicated that score type had a statistically significant influence ($p < .05$) on the effect sizes.

Table 3.5. Study Characteristics as Moderators of Effect Sizes of OST Strategies for Mathematics

Moderator	*k*[a]	*Q*	Effect Size[b]	95% Confidence Interval	
				Lower Bound	Upper Bound
Study Quality		10.77**			
High	1		.23	.13	.33
Medium	12		.10	.04	.15
Low	9		.01	-.07	.09
Moderator	***k***[a]	***Q***	**Effect Size**[b]	**95% Confidence Interval**	
				Lower Bound	**Upper Bound**
Publication Type		.45			
Conference paper/ report	8		.11	.05	.17
Dissertation	11		.08	-.05	.21
Peer-reviewed journal	3		.08	.01	.15
Score Type		4.89*			
Gain score	10		.13	.08	.18
Posttest score	12		.03	-.04	.10

[a] Number of effect sizes included in the analysis
[b] Weighted *d*, fixed-effects model
* $p < .05$ ** $p < .01$

Moderator Relationships

To examine the studies for possible relationships among moderators, we constructed correlation matrices for strategy and study characteristics (Cooper, 1998). Studies of students in grades 3–12 reported primarily program strategies that had an academic focus, while studies of students in grades K–2 reported only foci that were academic with social activities. There were a similar number of studies of after-school programs and summer schools for each level of strategy duration, except for the longest duration (more than 100 hours), which was reported only by studies of after-school programs. Studies of programs with shorter durations (less than 75 hours) had strategies that were only academic, while studies of programs with longer durations reported both academic and academic with social foci. Regarding study characteristics, most of the studies rated as low quality reported only posttest scores, and the studies rated as medium quality reported both gain scores and posttest scores. (The one study with a rating of high quality reported gain scores.) There were no

apparent relationships among the studies between score type and the other moderators that we examined.

NARRATIVE REVIEW OF STUDIES

The following discussion is provided to communicate the varied characteristics of the studied programs as well as a profile of the body of research included in this chapter. It is important to note that the 11 studies that were not included in the meta-analysis due to insufficient data for calculating effect sizes are included in this discussion. All of the 11 studies excluded from the meta-analysis used quantitative methodology and designs, and most included activity descriptions. Table 3.6 presents characteristics of these 11 studies, including the treatment sample size, grade level(s), timeframe (i.e., summer school, or after-school), focus (i.e., academic only or academic and social), grouping (e.g., large group, small group, one-on-one, or a combination), and student outcome results (i.e., all positive, mostly positive, even, mostly negative, or all negative).

Table 3.6. Study Characteristics of Narrative Review Mathematics Studies

Author(s) and (Year)	Treatment Sample Size	Grade Level(s)	Time Frame	Program Focus	Student Grouping	Results[a]
Grimm (1997)	19	6th–8th	summer school & after school	academic & social	one-on-one mentoring & small group	mn
Hansen, Yagi, and Williams (1986)	871	3rd–7th	summer school	academic & social	missing	mp
Huang, Gribbons, Kim, Lee, & Baker (2000)	4,312	2nd–5th	after school	academic & social	large group; one-on-one	mp
Kushmuk & Yagi (1985)	67	3rd–7th	summer school	academic & social	small group	e
Lodestar Mgmt. Research (2003)	160	2nd–8th	after school	academic & social	varies by site	e
Paeplow, Baenen, & Banks (2002)	116	2nd–8th	summer school	academic	small group; one-on-one tutoring	mn
Pyant (1999)	30	K–4th	after school	academic & social	small group	e
Rachal (1986)	9,675	2nd–5th	summer school	missing	missing	mn
Ronacher, Tullis, & Sanchez (1990)	1,072	9th–12th	Saturday school	academic & social	large group	e
Schinke, Cole, & Poulin (2000)	283	5th–8th	after school	academic	one-on-one & large group	mp
Sipe, Grossman, & Milliner (1988)	1,272	8th–10th	summer school	academic & social	small group	mp

A comparison between these 11 studies and those included in the meta-analysis (see Table 3.3) revealed a number of similarities. As was seen in the meta-analysis group, these studies too represented a variety of publication years (from 1985 to 2002) and a variety of treatment sample sizes (from 19 to 9,675), and describe a variety of interventions (from tutoring to mixed interventions to large-group sessions). The number of studies at the different grade levels and intervention timeframes are also quite similar to the results presented in Table 3.4. Table 3.6 indicates no apparent relationship between the study or program characteristics and the intervention results. The studies that reported mostly negative results (n = 3) and even results (n = 4) represented a variety of publication years, sample sizes, age groups, and interventions.

COMMON FEATURES HIGHLIGHTED IN STUDIES

The 33 studies included in this chapter described a wide variety of programs. Of course they all involved mathematics instruction, ranging from homework assistance to the administration of a carefully designed curriculum. But these programs have other varying characteristics as well. A number of them were designed to provide counseling or mentoring, some had large recreational components, and some used OST to provide tutoring and small-group instruction. It is clear that OST provides more time for student learning, and there was a group of studies specifically designed to tie this additional time to the performance of participants. We also identified a group of studies that described life-skill programs, of which mathematics instruction was a primary component.

In the next section, some of the programs studied in the research are described in an effort to illustrate not only the variety within the body of available research, but also to provide specific examples of the programs that informed our results.[17]

More Time for Remediation

In a general sense, all OST programming is an effort to affect performance by scheduling more time for instruction. As demonstrated by our meta-analysis, however, strategy duration does not necessarily translate directly into increased student achievement.

[17] Unless otherwise noted, the effect sizes reported in this section are from the meta-analysis described in Table 3.3.

Given this finding, we returned with renewed interest to a small group of studies included in this chapter that were designed specifically to determine whether or not OST has been effective in producing gains in mathematics achievement. This group of studies described in each case an effort to evaluate the effectiveness of a large-scale program, efforts that often involved a number of program sites. For example, a study of a New York City program examined 96 of its sites, research that included data on 3,780 elementary and middle school students (Welsh et al., 2002). There were positive effects on mathematics achievement for 183 students who actively participated for two years (d = .24). The authors noted that the academic gains were particularly strong for their low-achieving students.

Rachal (1986) studied summer school programs across Louisiana in another large-scale evaluation. Among the most significant finding of this report was that the state's summer school program did not result in an increase in state-level test scores or a decline in retention rates as expected. A similar result was reported by Prenovost (2001) after the author completed a survey and records examination of students attending four California after-school programs. The study was designed to determine the effects that these programs might be having on middle school participants, but no statistically significant results were identified (d = .08 for the girls in the study, and d = .00 for the boys).

The Summer Training and Educational Program (STEP), mentioned in Chapter 2, is another large-scale program addressed in the mathematics research (Branch et al., 1986; Sipe et al., 1988). STEP was designed to promote high school graduation and successful transition to careers with what previously had been merely a federal summer jobs program. Thousands of students participated in the five urban programs during the summers of 1986, 1987, and 1988. These students were exposed to academic classes, and life and career counseling. These interventions had measurable academic effects on the treatment participants (d = .23 for Branch et al., 1986).

The last study in this set, the Math Power summer program of Montgomery County Public Schools in Maryland, had been in operation for six years when Zia, Larson, and Mostow (1999) published an analysis of the program's effectiveness. The authors collected data for third- through fifth-grade students and found only small significant mathematics achievement growth for treatment students (d = .06, d = .07, and d = .06, respectively for the three grades).

In each of the studies mentioned above, specific implementation descriptions were omitted that would aid us in synthesizing a set of effective strategies. This is a particularly important omission given the large number of subjects in these studies, and the strong resulting influence that four of these studies (Branch et al., 1986; Prenovost, 2001; Welsh et al., 2002; Zia et al., 1999) had on the meta-analytic

results. It is this set of studies that supports the conclusions that can be drawn in terms of program characteristics as described in the meta-analytic results. Beyond our moderator analysis of program duration, the results from four of these studies informed the moderator analyses of timeframe, participant grade level, and activity focus.

Tutoring

Tutoring has been shown to have a positive effect on the academic performance of low-achieving students (Elbaum et al., 2000; Barley et al., 2002), and the additional flexibility of OST programming makes one-on-one interactions more feasible, so it is no surprise that tutoring programs appear in the OST research.

Pyant (1999), for example, described the El-Shaddai after-school program in Queens, New York. The program, supported by a local church and parent fees, was designed for early elementary students. High school and adult tutors assisted the students through homework review, social skills lessons, and academic lessons including reading, writing, math, and spelling. Another tutoring program supported by a local church was studied by McKinney (1995). The Leap Frog Program of Oxford, Mississippi, combined remedial tutoring with enrichment classes in an effort to meet the needs of the program's first and second graders, although academic effects were not demonstrated by the study results (d = -.14).

Other tutoring programs are described in the research. One leadership program is described by Paeplow, Baenen, and Banks (2002) as utilizing both tutoring and cooperative learning components. Another program, one that added parent classes to its tutoring and class schedule, was described by Smeallie (1997), but neither of these programs reported positive results for participants (d = -.10 in Smeallie, 1997). A program that combined tutoring with computer-assisted instruction was reported by Leslie (1998) to have positive results (d = .19, d = .35, and d = .62, respectively for grades 6 through 8). It is important to note that the Leslie study combined tutoring with computer-assisted instruction, a strategy that Barley et al. (2002) found was effective for increasing mathematics achievement. Thus, these data do not support the use of tutoring as a sole or primary strategy in OST programs designed to address mathematics achievement.

Mathematics Instruction with Life Skills

Several of the programs serving high school participants worked to combine academic instruction with life skills or, more specifically, career or college skills. Rembert et al. (1986) studied an intensive three- to four-week residential summer

school camp conducted between 1982 and 1984. The South Carolina high school participants were identified by their school counselors as being capable yet unmotivated, particularly with respect to college application. The program was designed to introduce these students to a collegiate atmosphere including access to academic classes, laboratories, computers, and recreational facilities. The authors of the study reported positive effects on mathematics achievement (d = .34).

A similar program, the Twenty-first Century Mathematics Center for Urban High Schools, was studied by Riley (1997). This program brought high school students to the Temple University campus. The students were taught high school mathematics in large classes and required to complete worksheets. The program was complemented by an individual and small-group tutoring component. Unlike the other tutoring research presented in the previous section, Riley reported positive effects in mathematics achievement for student participants as compared to a matched group of students from low socio-economic families (d = .83 for the male participants, and d = .99 for the females).

Affecting Performance through Counseling and Mentoring

Harlow and Baenen (2001) conducted an evaluation of the Wake County, North Carolina Summerbridge Program. The program had an academic summer school component followed by school year programming with academic counseling, mentoring, Saturday school, and community service. The seventh graders involved in the program demonstrated performance gains (d = .16) and reduced dropout when compared to a group of similar students who had not attended summer school. Another similar program, the Pride Program in Newport News, Virginia, was studied by Grimm (1997). The Pride Program had a residential summer school and school year components. During the school year, the participating middle school students attended academic classes, and field trips and were mentored by public school staff as well as staff of Newport News Shipbuilding, a partnership business. However, the standardized test results for these participants showed no gains in mathematics performance.

Combining Recreation with Mathematics Instruction

Positive effects (d = .31) were recorded for the third through sixth graders who participated in two Austin, Texas after-school programs (Baker & Witt, 1996). In each of the programs, certified teachers provided the students with a wide variety of activities and classes ranging from recreation to academics. Topics included natural science field trips, gardening, sports, and cultural activities, as well as academic classes. In a more recent study of a similar program, Lodestar Management/Research

(2003) evaluated the Woodcraft Rangers After-school Program. The different classes offered to the second through eighth graders in this Los Angeles program were designed to enhance academic, physical, and social development. The authors reported that the intervention had a limited effect on achievement and the grades of 43 percent of the participants fell.

DISCUSSION OF FINDINGS

The effectiveness of OST strategies in mathematics reported in 33 studies was reviewed and effect sizes from 22 studies were computed and synthesized through meta-analysis. Our analysis provided evidence that OST strategies in mathematics can improve the mathematics achievement of low-achieving or at-risk students. The effect size based on a fixed-effects model was .09, and the effect size based on a random-effects model was .17. This indicates that the mean achievement of the students who received OST programs was .09 to .17 standard deviations higher than those students in the study who did not receive OST programs. With respect to the 11 studies that were excluded from the meta-analysis, it is worth noting that four of these reported mostly positive results, while four reported even results and three more reported mostly negative results.

Cooper et al. (2000) reported an effect size of .27 (fixed-effects model) for the effectiveness of summer school on outcome measures related to mathematics. As indicated in Chapter 2, their meta-analysis included studies that used single group pre- and posttest designs, which were excluded from the current synthesis. The effect size for studies that used random assignment in the Cooper et al. synthesis was .14 under both fixed-effects and random-effects models (p. 90). These results are more consistent with our findings for OST studies with mathematics outcomes, all of which included control or comparison groups.

Although the overall effect sizes demonstrated a positive effect of OST strategies in mathematics, homogeneity analysis indicated large variation across the effect sizes reported by the 22 studies. According to our moderator analysis, three program characteristics were associated with this variation: grade level of the students, activity focus, and strategy duration. In addition, the study characteristic of research quality had a statistically significant influence.

Our data showed that the OST programs implemented for middle school and high school students tended to be more successful for helping low achievers improve mathematics achievement than those implemented for elementary school students. We examined the interaction of grade level with other moderators and did not

identify relationships that might explain the variation in effect sizes by the grade level of students.

Given the strong focus on secondary level in mathematics reform initiatives compared to those on the elementary school level, it might be that the OST strategies in mathematics are more developed at the secondary level. The data from the Third International Mathematics and Science Study (TIMMS) conducted during 1995 and 1999 found that mathematics achievement of 8th and 12th graders in the United States was lower than in most other industrial nations, while our 4th graders' achievement exceeded that of most nations (Mullis et al., 2000). Reformers and educators' attempts to improve the mathematics achievement of secondary grade students over the past decade might be reflected in the development of successful OST strategies to assist low achievers in high school.

Strategy focus and duration were the other two program characteristics that explained the variation of effect sizes across the studies we examined. Both OST strategies that focused on academic enrichment and on social enrichment (e.g., music, art, social and life skills, and recreational or vocational activities), and OST strategies with a purely academic focus had significantly positive average effect sizes. As some researchers have advocated (Heath, 1994; Miller, 2003), low-achieving or at-risk students who are not successful in regular school hours might need a different learning environment in order to improve their achievement. However, the five studies that provided achievement results for their high school participants produced the largest positive effect size (d = .44); these studies were of programs that were academic in emphasis.

The moderator analysis of strategy duration provided an interesting finding that OST strategies in mathematics were more effective when implemented more than 45 hours but less than 100 hours. The smaller effect size of OST strategies implemented more than 100 hours might be due to changes in program implementers or financial resources. It also might be related to "contamination" of internal validity: when program implementation prolongs, the control group students are more likely to be involved in concurrent events or processes, which affects the isolation of effectiveness of the OST strategies.

Although researchers presume there are potential differences in effect sizes by the type of OST strategies such as summer school program and after-school programs, we did not find a statistically significant difference in effectiveness based on the timeframe of OST strategies. Thus, what matters is not *when* the programs are implemented, but *how* they are implemented.

In addition to the program characteristics that explained the effect size variation in mathematics achievement, we also observed that effect sizes differed by our ratings of the research quality of studies. The one study that was coded as high quality had the highest effect size compared to the average effect sizes reported by medium- and low-quality studies. Although we cannot be conclusive about quality based on a single study, the fact that the 12 studies rated as medium quality had larger effect sizes than the 9 rated as low quality supports our confidence about the positive effects of OST strategies in assisting at-risk students to improve their mathematics achievement.

The publication type did not influence the effect size of OST strategies, but the type of score reported in studies had a statistically significant influence on the effect sizes. The average effect size for gain scores was significantly different from zero, although this was not true for the average effect size of studies that reported posttest scores. Most of the studies rated as low quality reported only posttest scores, and the studies rated as medium quality reported both gain scores and posttest scores. The one high-quality study reported gain scores. Studies that report gain scores also give more attention to group differences that might influence results, which leads to higher ratings on criteria related to internal validity, resulting in a higher quality rating.

IMPLICATIONS FOR POLICY AND PRACTICE

Our findings from the meta-analysis and narrative review provided evidence that OST strategies in mathematics can be effective strategies for helping low-achieving or at-risk students. Our ability to make specific strategy recommendations is limited by the lack of details on implementation reported in the available research. However, the research does support some conclusions that can inform effective practice.

OST strategies in mathematics can be particularly effective when they are implemented for secondary students. Programs described by McMillan and Snyder (2002) and Riley (1997) are exemplary resources for implementers of OST programs for high school students. Programs that add social enrichment to an academic focus can have positive effects on mathematics achievement (Branch et al., 1986). As a program moderator, OST tutoring did not improve the mathematics performance of at-risk students in the available research. The exception was Leslie's (1998) study, which is a resource for implementers who are considering an OST tutoring intervention that utilizes computer-aided designs in mathematics instruction. Finally, the studies in our review that documented successful interventions suggest that careful program design and program fidelity are important elements of effective OST strategies for addressing the needs of low-achieving or at-risk students in mathematics.

CHAPTER 4

Summary and Conclusions

This chapter begins with a summary and interpretation of findings on the impact of OST strategies on student achievement in reading and mathematics. This section is followed by a discussion of research issues related to studies of OST. The final section presents conclusions and implications of the research synthesis.

SUMMARY AND INTERPRETATION OF RESULTS

We synthesized research on the effectiveness of OST strategies in assisting low-achieving or at-risk students. We conducted meta-analyses of outcomes in reading achievement from 27 studies and of outcomes in mathematics achievement from 22 studies. An additional 20 studies with insufficient information for meta-analysis informed the findings for reading, and an additional 11 studies informed the results for mathematics. The 53 different studies in the synthesis (27 studies were used for both reading and mathematics) each used a control or comparison group to reach conclusions. Over 40 percent of the studies (23) in the synthesis were published in the year 2000 or later.

Overall Effect Sizes of OST Strategies

For reading outcomes, the overall effect size was .06 for the fixed-effects model and .13 for the random-effects model. For mathematics outcomes, the overall effect size was .09 for the fixed-effects model and .17 for the random-effects model. All four of the effect sizes were statistically greater than zero. In answer to the research problem posed in Chapter 1, the results indicate that based on rigorous research studies (as defined by the use of control or comparison groups), OST strategies can have positive effects on the achievement of low-performing or at-risk students.

Three factors influence the interpretation of the overall effect sizes. First, OST strategies supplement the regular school day, so the interpretation of effect sizes for typical education interventions might not apply (see e.g., Cohen's [1988] statement that an effect size of .20 is small). Second, the students who participated in OST strategies were at risk for school failure. Researchers have referred to resilience and the prevention of learning loss as indicators of positive outcomes for such students (Miller, 2003). Thus, the finding of a positive effect size that is statistically greater than zero is an encouraging result for the use of OST strategies to assist low-

achieving or at-risk students. Third, certain moderators resulted in larger positive effects on student achievement as compared to the overall effect sizes.

Influence of Moderators on Effect Sizes

We examined five characteristics of OST strategies for possible moderating influences on effect sizes. The timeframe for delivery of OST strategies was not a statistically significant moderator. The OST strategies in most of the studies in the synthesis were implemented in either an after-school setting or during summer school. Our results indicate that timeframe per se is not an influence on the impact of OST on student achievement. However, as indicated in Chapter 2, more studies of after-school reading programs were reported to be short in duration (less than 45 hours) compared to studies of summer school reading programs Although short durations were associated with lower effect sizes, studies of after-school programs also reported more one-on-one and mixed-group strategies than studies of summer school, which reported more use of large groups. Because small groups and one-on-one instruction were associated with more positive effects compared to large groups, the benefits of summer schools of longer duration might be offset by the use of large-group instruction.

Grade level was a statistically significant moderator of effect size for both reading and mathematics outcomes. For reading, the largest positive effect size (.26) occurred for students in the lower elementary grades (K–2); for mathematics, the largest positive effect size (.44) was for students in high school (9–12). The result for reading confirms the importance of early intervention for students who are underachieving in reading. The results for mathematics suggest that OST programs might be effective in addressing the achievement deficiencies that can prevent at-risk students from being accepted into postsecondary education programs.

The findings were mixed regarding the activity focus of OST, that is, whether it was primarily academic or academic plus social. For reading outcomes, activity focus was not a statistically significant moderator of effect size; whereas for mathematics outcomes, strategies that were both academic and social had a slightly higher mean effect size than those that were mainly academic. This indicates that OST need not focus only on academics in order to produce positive effects. In fact, some researchers of OST have stressed the need for variety in programming in order to motivate students to attend, particularly in the upper grades (Miller, 2003; De Kanter, 2001; Huang et al., 2000).

For both reading and mathematics, statistically significant effect sizes were larger for OST programs that were more than 45 hours in duration, but the programs with the

longest durations (more than 210 hours for reading and more than 100 hours for mathematics) had effect sizes that were not significantly different from zero.

Although the data are not available to confirm this, it is probably more difficult for longer programs compared to shorter programs to keep students motivated and attending on a regular basis. However, it is interesting that program durations of up to 210 hours were associated with positive effects on reading outcomes, while durations of longer than 100 hours were associated with less positive outcomes in mathematics. This suggests there are differences in the optimal durations for OST strategies that address the two content areas. More research is needed on OST strategies for different content outcomes. The "one size fits all" nature of many OST programs might work against program effectiveness.

Only the reading studies had sufficient information to analyze the statistical influence of the way that students are grouped in OST programs. The largest positive effect (.50) occurred for the studies that used one-on-one tutoring (e.g., Leslie, 1998). This result confirms other research that demonstrates the positive influence of tutoring and individualized help for low-achieving or at-risk students, especially in reading (Elbaum et al., 2000).

We examined three other characteristics for possible moderating influences on effect size. The results for study quality were mixed. In the meta-analyses, there were two high-quality studies with reading outcomes and one high-quality study with mathematics outcomes. Most of the studies were rated as medium in research quality. For mathematics, there was a statistically significant result in favor of higher quality studies, but quality ratings did not significantly influence effect sizes for reading. Thus the overall findings across the two content areas were too varied to support conclusions related to research quality.

Type of publication was a statistically significant moderator of effectiveness of OST for reading achievement but not for mathematics. The effect size for reading studies reported in peer-reviewed journals was larger than for unpublished reports and dissertations. This supports the notion that studies with statistically significant results favoring an intervention are more likely to be published in journals than are non-significant or negative findings. It also emphasizes the importance of locating unpublished program evaluations so that conclusions about intervention effectiveness are based on the complete body of available research.

Finally, the type of score had a significant influence on the effect sizes for mathematics but not for reading. For mathematics outcomes, the average effect size for gain scores (or pretest-posttest difference scores) was significantly greater than zero, although this was not true for the average effect size based on posttest scores.

The distribution of moderators among the mathematics studies indicated that the studies with low-quality ratings reported primarily posttest scores, and studies with medium- or high-quality studies reported both gain scores and posttest scores. It is possible that the reliance on posttest scores instead of gain scores is one reason that the low-quality mathematics studies had lower effect sizes than the medium- or high-quality studies.

RESEARCH ISSUES

Those who research and evaluate OST programs face difficult challenges. In this synthesis, we examined only studies that had a control or comparison group, and we rated the quality of studies higher if they used comparable groups or random assignment of students to groups. But as Miller (2003) observed, "When it comes to out-of-school time, there is no such thing as a 'no treatment' group" (p. 88). The reason is that children are always doing something after school, and the "something" becomes the comparison "intervention." Another issue stems from the fact that attendance at OST programs is voluntary and not mandated. Some studies point to the relationship between attendance and OST effects (Baker & Witt, 1996), yet if the students who attend more are more academically motivated than those who drop out, program effects might be due more to higher student motivation than to the OST intervention (Fashola, 1998). Complicating the issue is that very few studies have documented the number of students who dropped out of OST programs and the reasons they dropped out.

Another problem with research on OST strategies is the failure to describe program details and to assess treatment fidelity. It is difficult to make specific recommendations from the body of research on OST strategies when research and evaluation reports give only vague references to the intervention, such as "homework help," and provide no measures of the degree to which the intervention was implemented. Until research and evaluation of OST strategies become more systematic in measurement and reporting, recommendations for specific practices can be based only on minimal evidence.

CONCLUSIONS AND IMPLICATIONS

The results of this synthesis lead to several conclusions and implications for practice and policy related to OST and its evaluation:

OST strategies can have positive effects on the achievement of low-achieving or at-risk students in reading and mathematics. This finding supports Cooper et al.'s (2000) meta-analytic results for

summer school and previous narrative reviews of research on after-school programs (e.g., Fashola, 1998). With regard to the recent evaluation of the 21st Century Community Learning Centers program (U.S. Department of Education, 2003), our results suggest that after-school programs can influence student learning. Conclusions about the ineffectiveness of that program might be due to the aggregation of interventions that have different characteristics in the evaluation study. Our synthesis indicated that program duration and student grouping influence program effectiveness. Aggregating results across programs that vary in these characteristics can mask positive outcomes.

The timeframes of OST programs do not influence the effectiveness of OST strategies. In deciding whether to fund OST programs, policymakers should look at other factors such as program duration, cost, and implementation issues (e.g., staff recruitment, program location) when choosing between after-school and summer school programs .

Students in early elementary grades are more likely than older elementary and middle school students to benefit from OST strategies for improved reading, while there are indications that the opposite is true for mathematics. The findings for reading achievement support prior research on the importance of early reading skills, while the results for mathematics are encouraging. However, additional research is needed given the greater difficulty in recruiting older students into OST programs (Grossman et al., 2001).

OST strategies need not focus solely on academic activities to have positive effects on student achievement. Study results indicate that OST programs in which activities are both academic and social can have positive influences on student achievement. This finding supports the belief that OST programs should address the developmental needs of the whole child (Halpern, 2002) and offer a variety of activities (Miller, 2003). However, our results also suggest that effectiveness related to program focus might vary depending on grade level and content area.

Administrators of OST programs should monitor program implementation and student learning in order to determine the appropriate investment of time for specific strategies and activities. Although OST programs need to deliver strategies for a minimum

amount of time to be effective (i.e., more than 45 hours), longer OST programs do not necessarily have more positive outcomes. Optimal duration may depend on the content area. This result supports other findings that extending the time for learning does not mean that students will be engaged in learning during that additional time (WestEd, 2002).

OST strategies that provide one-on-one tutoring for at-risk students have positive effects on student achievement in reading. This was one of the strongest findings from the meta-analysis and is supported by other research on tutoring of at-risk students during the school day (Barley et al., 2002; Elbaum et al., 2000). OST programs that have reading improvement as a goal should provide individual tutoring of students.

Research syntheses of OST programs should examine both published and unpublished research and evaluation reports. Estimates of the true effect of OST strategies on student achievement will be inaccurate if only published studies are examined because statistically non-significant findings tend not be published or even submitted for publication. To balance the breadth of inclusion, researchers should examine the methodological quality of unpublished studies.

Future research and evaluation studies should document the characteristics of strategies and their implementation. Researchers and evaluators have proposed guidelines for OST programs, such as the need for structure and trained staff (Fashola, 1998), but systematic documentation through research and evaluation is lacking. Policymakers, administrators, and educators need evidence on the characteristics of effective OST strategies.

References

(References marked with an asterisk indicate studies included in the meta-analyses.)

The After-School Corporation. (1999). *After-school programs: An analysis of need, current research, and public opinion.* Retrieved August 13, 2003, from http://www.tascorp.org

After-school grants yield few results, ED study says. (2003). *What Works in Teaching and Learning, 35*(2), 3, 9.

Apthorp, H. S., Dean, C. B., Florian, J. E., Lauer, P. A., Reichardt, R., & Snow-Renner, R. (2001). *Standards in classroom practice: Research synthesis* (REL Deliverable #2001-17). Aurora, CO: Mid-continent Research for Education and Learning.

Ascher, C. (1990). *Summer school, extended school year, and year-round schooling for disadvantaged students. ERIC/CUE Digest Number 42.* New York: ERIC Clearinghouse on Urban Education. (ERIC Document Reproduction Service No. ED 298 213)

*Baker, D., & Witt, P. A. (1996). Evaluation of the impact of two after-school programs. *Journal of Park and Recreation Administration, 14*(3), 23–44.

Barley, Z., Lauer, P. A., Arens, S. A., Apthorp, H. S., Englert, K. S., Snow, D., & Akiba, M. (2002). *Helping at-risk students meet standards: A synthesis of evidence-based classroom practices* (REL Deliverable #2002-20). Aurora, CO: Mid-continent Research for Education and Learning.

*Bergin, D. A., Hudson, L. M., Chryst, C. F., & Resetar, M. (1992). An after-school intervention program for educationally disadvantaged young children. *Urban Review, 24*(3), 203–217. (ERIC Document Reproduction Service No. EJ 450 987)

Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis, 18*(4), 309–326.

*Borman, G. D., Rachuba, L. T., Fairchild, R., & Kaplan, J. (2002). *Randomized evaluation of a multi-year summer program: Teach Baltimore. Year 3 report. Draft*. Madison, WI: University of Wisconsin, Madison. Retrieved April 7, 2003, from http://www.jhu.edu/teachbaltimore/resources/docs/Year_III_Report.pdf

*Branch, A. Y., Milliner, J., & Bumbaugh, J. (1986). *Summer Training and Education Program (STEP): Report on the 1985 summer experience*. Philadelphia: Public/Private Ventures. (ERIC Document Reproduction Service No. ED 283 028)

Braswell, J. S., Lutkus, A. D., Grigg, W. S., Santapau, S. L., Tay-Lim, B., & Johnson, M. (2001). *The nation's report card: Mathematics 2000.* Washington, DC: National Center for Education Statistics. Retrieved August 13, 2003, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001517

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: Government Printing Office. (ERIC Document Reproduction Service No. ED 012 275)

Committee for Economic Development. (2000, December). *Measuring what matters: Using assessment and accountability to improve student learning.* NewYork: Author.

Cooper, H. (1998). *Synthesizing research* (3rd ed). Thousand Oaks, CA: SAGE Publications.

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,* Serial No. 260, 65(1).

*Cosden, M., Morrison, G., Albanese, A. L., & Macias, S. (2001). When homework is not home work: After-school programs for homework assistance. *Educational Psychologist, 36*(3), 211–221.

*D'Agostino, J., & Hiestand, N. (1995, April). *Advanced-skill instruction in chapter 1 summer programs and student achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 390 929)

Davidson, J., & Koppenhaver, D. (1993). *Adolescent literacy: What works and why* (2nd ed.). New York: Garland Publishing.

De Kanter, A. (2001). After-school programs for adolescents. *NAASP Bulletin, 85*(626), 12–21.

Duffy, A. M. (2001). Balance, literacy acceleration, and responsive teaching in a summer school literacy program for elementary school struggling readers.

Reading Research and Instruction, 40(2), 67–100. (ERIC Document Reproduction Service No. EJ 624 633)

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-on-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology,* 92(4), 605–619.

Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness* (Report No. 24). Baltimore, MD: Center for Research on the Education of Students Placed at Risk. Retrieved July 2, 2003, from http://www.gse.uci.edu/asp/aspeval/resources/Report_24_Fashola.pdf

Fashola, O. S. (2002). *Building effective afterschool programs.* Thousand Oaks, CA: Corwin Press.

*Finch, C. E., Jr. (1997). The effect of supplementary computer-assisted instruction upon rural seventh-grade students to improve math scores as measured by the Michigan Educational Assessment Program test. (Doctoral dissertation, Walden University, 1997). *Dissertation Abstracts International, 58,* 08A.

*Foley, E. M., & Eddins, G. (2001). *Preliminary analysis of virtual after-school program participants' patterns of school attendance and academic performance*. New York: National Center for Schools and Communities.

*Gentilcore, J. C. (2002). *The effect of an after-school academic intervention service on a New York State eighth grade English language arts assessment: A case study.* (Doctoral dissertation, Hofstra University, 2002). *Dissertation Abstracts International, 63,* 06A.

Grigg, W. S., Daane, M. C., Jin, J. & Campbell, J. R. (2003). *The nation's report card*: *Reading 2002*. Retrieved Aug 2, 2003, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003521

Grimm, M. T. A. (1997). A study of the "pride" program: A residential, summer-school program for at-risk middle-school students in an urban community (intervention, mentorship, school business partnerships). (Doctoral dissertation, Old Dominion University, 1997). *Dissertation Abstracts International, 58*, 09A.

Grossman, J. B., Walker, K., & Raley, R. (2001, April). *Challenges and opportunities in after-school programs: Lessons for policymakers and funders.* Philadelphia: Public/Private Ventures.

Halpern, R. (1999). After-school programs for low-income children: Promise and challenges. *Future of Children, 9*(2), 81–95. (ERIC Document Reproduction Service No. EJ 600 564)

Halpern, R. (2000). The promise of after-school programs for low-income children. *Early Childhood Research Quarterly, 15*(2), 185–214. (ERIC Document Reproduction Service No. EJ 633 367)

Halpern, R. (2002). A different kind of child development institution: The history of after-school programs for low-income children. *Teachers College Record, 104*(2), 178–211. Retrieved May 28, 2003, from http://www.tcrecord.org/pdf/10823.pdf

Hansen, J. B., Yagi, K., & Williams, I. (1986). *Elementary summer school, 1985 program in the Portland Public Schools*. Portland, OR: Portland Public Schools, Oregon Dept. of Research, Evaluation, and Testing. (ERIC Document Reproduction Service No. ED 273 662)

*Harlow, K., & Baenen, N. (2001). *The effectiveness of the Wake Summerbridge Summer Enrichment Program. Eye on evaluation. E&R report*. Raleigh, NC: Wake County Public School System & the Department of Evaluation & Research. (ERIC Document Reproduction Service No. ED 466 489)

Harvard Family Research Project. (2003, spring). The Evaluation Exchange Special Report on the 21st Century Community Learning Centers Program National Evaluation [Special Report]. Compiled by Priscilla Little. *The Evaluation Exchange, IX*(1). Retrieved August 31, 2003, from http://www.gse.harvard.edu/~hfrp/eval/issue21/special.html

*Hausner, M. E. I. (2000, May). *The impact of kindergarten intervention project accelerated literacy on emerging literacy concepts and second grade reading comprehension.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. (ERIC Document Reproduction Service No. ED 451 951)

Heath, S. B. (1994). The project of learning from the inner-city youth prospective. *New Directions for Child Development, 63*, 25–34.

*Hink, J. J. (1986). A systematic, time-extended study of a remedial reading and math summer school program (self-funding, standardized test, public school, student learning, elementary grades). (Doctoral dissertation, Wayne State University, 1986). *Dissertation Abstracts International, 47*, 04A.

Holdzkom, D. (2002). *Summer academy in the wake county public schools: An evaluation of the 2001 program*. Retrieved April 7, 2003, from http://www.wcpss.net/evaluation-research/index_reports/2002/0209_alp.pdf

Howe, R., & Kasten, M. (1992). *Students at risk in mathematics: Prevention and recovery in elementary schools.* Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education. (ERIC Document Reproduction Service No. ED 350 175)

*Howes, M. (1989). *Intervention procedures to enhance summer reading achievement (summer school, library reading program).* (Doctoral dissertation, Northern Illinois University, 1989). *Dissertation Abstracts International, 51*, 01A.

Huang, D., Gribbons, B., Kim, K. S., Lee, C., & Baker, E. L. (2000). *A decade of results: The impact of the LA's Best After School Enrichment Initiative on subsequent student achievement and performance.* Los Angeles: UCLA Center for the Study of Evaluation, Graduate School of Education & Information Studies, University of California.

Hynes, K., O'Connor, S., & Chung, A. M. (1999). *Literacy: Exploring strategies to enhance learning in after-school programs*. Boston: Wellesley Collage M.A. Center for Research on Women. (ERIC Document Reproduction Service No. ED 438 045)

International Reading Association. (1999). *Summary of adolescent literacy, A position statement for the Commission on Adolescent Literacy of the International Reading Association.* Retrieved August 4, 2003, from http://www.reading.org/positions/adol_lit.html

Jacob, B. A., & Lefgren, L. (2001). *Remedial education and student achievement: A regression-discontinuity analysis*. Boston: National Bureau of Economic Research. (ERIC Document Reproduction Service No. ED 465 007)

King, F., & Kobak, J. (2000). *Cim academy summer school: A report of the evaluation of the 2000 summer school*. Portland, OR: Portland Public Schools.

*Kociemba, G. D. (1995). The impact of compensatory summer school on student achievement: Grades 2 and 5 in the Minneapolis Public Schools. (Doctoral dissertation, University of Minnesota, 1995). *Dissertation Abstracts International, 56*, 05A.

Kugler, M. R. (2001). After-school programs are making a difference. *NASSP Bulletin, 85*(626), 3–11.

Kushmuk, J., & Yagi, K. (1985). *Summer school 1984 program in the Portland Public Schools. 1983-84 evaluation report*. Portland, OR: Portland Public Schools, Department of Research and Evaluation.

*Leboff, B. A. (1995). The effectiveness of a six-week summer school program on the achievement of urban, inner-city third-grade children. (Doctoral dissertation, Texas Southern University, 1995). *Dissertation Abstracts International, 56*, 10A.

*Legro, D. L. (1990). An evaluation of an after-school partnership program: The effects on young children's performance. (Doctoral dissertation, University of Houston, 1990). *Dissertation Abstracts International, 52*, 02A.

*Leslie, A. V. L. (1998). The effects of an after-school tutorial program on the reading and mathematics achievement, failure rate, and discipline referral rate of students in a rural middle school (rural education). (Doctoral dissertation, University of Georgia, 1998). *Dissertation Abstracts International, 59*, 06A.

*Levinson, J. L., & Taira, L. (2002, April). *An investigation of summer school for elementary students: Outcomes and implications.* Paper presented at the American Educational Research Association, New Orleans, LA.

Lodestar Management/Research. (2003). *Woodcraft Rangers: Los Angeles Unified School District After School Education and Safety Program. Annual evaluation report 2001–02*. Los Angeles: Lodestar Management/Research, Inc.

*Luftig, R. L. (2003, May). *When a little bit means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers.* Paper presented at the American Educational Research Association, Chicago.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mathes, P.G. (2003, April). *Preventing and remediating reading difficulties*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

McComb, E. M., & Scott-Little, C. (2003). *After-school programs: Evaluations and outcomes*. Greensboro, NC: SERVE.

*McKinney, A. D. (1995). The effects of an after school tutorial and enrichment program on the academic achievement and self-concept of below grade level first and second-grade students. (Doctoral dissertation, University of Mississippi, 1995). *Dissertation Abstracts International, 56*, 06A.

*McMillan, J. H., & Snyder, A. L. (2002, April). *The effectiveness of summer remediation for high-stakes testing.* Paper presented at the American Educational Research Association, New Orleans, LA.

McREL. (n.d.). List of benchmarks for language arts (4th ed.). Retrieved August 15, 2003, from http://www.mcrel.org/compendium/standardDetails.asp? subjectID=7&standardID=5

Miles, M.B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: Sage Publications.

Miller, B. M. (2003). *Critical hours: Afterschool programs and educational success*. Brookline, MA: Miller Midzik Research Associates for the Nellie Mae Education Foundation. Retrieved June 15, 2003, from http://www.nmefdn.org/CriticalHours.htm

Miller, D. E. (1993). The literature project: Using literature to improve the self-concept of at-risk adolescent females. *Journal of Reading, 36*(6), 442–448.

*Mooney, C. (1986). *The effects of peer tutoring on student achievement*. Union, NJ: Kean College of New Jersey. (ERIC Document Reproduction Service No. ED 270 730)

*Morris, D., Shaw, B, & Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. *Elementary School Journal, 91*(2), 133–150.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., et al. (2000). *TIMSS 1999 International Mathematics Report findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Boston: The International Study Center, Boston College, Lynch School of Education.

National Institute on Out-of-School Time. (2003). *Making the case: A fact sheet on children and youth in out-of-school time.* Retrieved August 13, 2003, from http://www.niost.org/

New Standards Primary Literacy Committee. (1999). *Reading and writing grade by grade: Primary literacy standards for kindergarten through third grade.* Pittsburgh, PA: National Center on Education and the Economy and the University of Pittsburgh.

Ortiz, G. K. (1993). *An exploratory study of the effects of an after-school literacy enrichment program on at-risk students and their parents (literacy development)*. (Doctoral dissertation, University of South Carolina, 1993). *Dissertation Abstracts International, 54*, 09A.

Paeplow, C. G., Baenen, N. R., & Banks, K. E. (2002). *Voyager summer academy 2002 results* (No. E&R Report No. 03.02). Raleigh, NC: Department of Evaluation and Research, Wake County Public School System.

Phelan, W. T. (1987). *Evaluator's report on the Lowell achievement program 1986–1987*. Lowell, MA: Lowell University. (ERIC Document Reproduction Service No. ED 286 980)

Policy Studies Associates. (1995). *Extending the learning time for disadvantaged students: An idea book. Volume 1, summary of promising practices.* Washington, DC: U.S. Department of Education. (ERIC Document Reproduction Service No. ED 389 094)

*Prenovost, J. K. E. (2001). A first-year evaluation of after-school learning programs in four urban middle schools in the Santa Ana Unified School District (California). (Doctoral dissertation, University of California, Irvine, 2001). *Dissertation Abstracts International, 62*, 03A.

Pyant, L. S. (1999). The impact of after-school tutoring with a social skills component on children in kindergarten through fourth grade in an inner-city public school (fourth-grade, inner city, urban education). (Doctoral dissertation, Walden University, 1999). *Dissertation Abstracts International, 60*, 08A.

Rachal, J. (1986). *Student placement study: 1985-86 state-funded compensatory/remedial program evaluation. Placement report. Bulletin 1765*. Baton Rouge, LA: Louisiana State Dept of Education Baton Rouge Office of Research Development. (ERIC Document Reproduction Service No. ED 269 445)

*Raivetz, M. J., & Bousquet, R. J. (1987, April). *How they spent their summer vacation: Impact of a tutorial program for students "at-risk" of failing a state mandated high school proficiency test.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 284 909)

Redd, Z., Cochran, S., Hair, E., & Moore, K. (2002). *Academic achievement programs and youth development: A synthesis*. Washington, DC: Child Trends.

*Reed, G. W. (2001). The relationship between participation in a developmental reading summer school program and reading achievement among low-achieving first grade students. (Doctoral dissertation, St. Louis University, 2001). *Dissertation Abstracts International, 62*, 05A.

*Rembert, W. I., Calvert, S. L., & Watson, J. A. (1986). Effects of an academic summer camp experience on black students' high school scholastic performance and subsequent college attendance decisions. *College Student Journal, 20*(4), 374–384.

*Riley, A. H. J. (1997). Student achievement and attitudes in mathematics: An evaluation of the twenty-first century mathematics center for urban high schools (urban education, summer school). (Doctoral dissertation, Temple University, 1997). *Dissertation Abstracts International, 58*, 06A.

Roderick, M., Engel, M., & Nagaoka, J. (2003). *Ending social promotion: Results from summer bridge*. Chicago: Consortium on Chicago School Research.

Ronacher, K., Tullis, R., & Sanchez, K. (1990). *Required academic proficiency (rap) program: Final report*. Houston, TX: Houston Independent School District, TX Dept of Research and Evaluation. (ERIC Document Reproduction Service No. ED 328 641)

*Ross, S. M., Lewis, T., Smith, L., & Sterbin, A. (1996). *Evaluation of the extended-day tutoring program in Memphis county schools: Final report to CRESPAR*. Memphis, TN: Center for Research in Educational Policy, University of Memphis.

*Schacter, J. (2001). *Reducing social inequality in elementary school reading achievement: Establishing summer literacy day camps for disadvantaged children*. Santa Monica, CA: Milken Family Foundation. Retrieved June 4, 2003, from http://www.mff.org/pubs/reading_camp_study2001.pdf

Schinke, S. P., Cole, K., & Poulin, S. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science, 1*(1), 51–60.

Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation, 23*(4), 387–419.

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning,* (pp. 623–660). Reston, VA: National Council of Teachers of Mathematics.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for causal inference*. Boston: Houghton Mifflin.

Shanahan, T. (2000). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. L. Kamil, P .B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, volume III* (pp. 209–226). Mahwah, NJ: Lawrence Erlbaum and Associates.

Sipe, C. L., Grossman, J. B, & Milliner, J. A. (1988). *Summer training and education program (STEP): Report on the 1987 experience*. Philadelphia: Public/Private Ventures. (ERIC Document Reproduction Service No. ED 300 479)

Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.

Slavin, R. E., & Madden, N. A. (1989). What works for students at risk: A research synthesis. *Educational Leadership,* 46(5), 4–13.

*Smeallie, J. E. (1997). An evaluation of an after-school tutorial and study skills program for middle school students at risk of academic failure. (Doctoral dissertation, University of Maryland, College Park, 1997). *Dissertation Abstracts International, 58*, 06A.

Spielberger, J., & Halpern, R. (with Pitale, S., Nelson, E., Mello-Temple, S., Ticer-Wurr, L., & Van Brunt, C.). (2002). *The role of after-school programs in children's literacy development.* Chicago: Chapin Hall Center for Children at the University of Chicago.

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125–138). New York: Russell Sage Foundation.

U. S. Department of Education. (2003). *When schools stay open late: The national evaluation of the 21st-century community learning centers program. First year findings.* (No. PR02-82). Washington, DC: U. S. Department of Education, Office of the Under Secretary.

Valentine, J. C., & Cooper, H. (2003). *What Works Clearinghouse Study Design and Implementation Assessment Device* (Version 0.6). Washington, DC: U. S. Department of Education.

Walker, G., & Vilella-Velez, F. (1992). *Anatomy of a demonstration: The summer training and education program (STEP) from pilot through replication and postprogram impacts*. Philadelphia: Public/Private Ventures.

*Ward, M. S. (1989, March). *North Carolina's summer school program for high-risk students: A two-year follow-up of student achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 307 687)

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.

*Weber, E. L. (1996). An investigation of the long-term results of summer school. (Doctoral dissertation, University of Wyoming, 1996). *Dissertation Abstracts International, 57*, 05A.

*Welsh, M. E., Russell, C. A., Williams, I., Reisner, E. R., & White, R. N. (2002). *Promoting learning and school attendance through after-school programs: Student-level changes in educational performance across TASC's first three years*. Washington, DC: Policy Studies Associates.

WestEd. (2002). *Improving student achievement by extending school: Is it just a matter of time?* Retrieved December 23, 2002, from http://www.wested.org/wested/papaers/timeandlearning/TAL_PV.html

*Zia, B., Larson, J. C., & Mostow, A. (1999). Instruction and student achievement in a summer school mathematics program. *ERS Spectrum, 17*(2), 39–47. (ERIC Document Reproduction Service No. EJ 589 455)

Appendix A: Coding Instrument

Study Number _____ Coder _____

syntax example: 2013 - category 2,13th document coded

McREL Research Synthesis: 2003
Strategies to Assist Low-Achieving Students Outside the School Day – Coding Guide

Codes: NA = Not Applicable M = Missing

Author(s):_____

Title:_____Report Year _____

Source: あ Journal _____

あ Dissertation あ ERIC report あ ERIC eval あ Other _____

Quality Index: Quantitative _____ Qualitative _____

Information for Table (complete after coding): Treatment sample size _____Grade or age _____

Student description _____ Type of comparison _____

Outcome measure _____ Direction of results* _____

*The number of independent samples revealing comparisons that were all positive (ap), mostly positive (mp) even (e) mostly negative (mn), and all negative (an)

1. PROGRAM/INTERVENTION INFORMATION

1.01 Determination of Low-Achieving or At-Risk:_____

1.02 Locale あ Urban あ Suburban あ Rural あ Missing

1.03 Population Characteristics: grade level and number of students:_____

% FRL _____ % LEP _____ %M _____ % F _____ % Caucasian _____

% African American ____ % Latino ____ % Native American ____ % Asian ____ % Other ____

1.04 Population Treated (check all that apply): あAt-Risk あSpecial Ed. あMigrant あELL あBilingual あGifted/Talented あOther _____

1.05 Assignment of Teachers/Implementers to Treatments: あSelf-selected あRandom あNon-random あMissing あOther _____

1.06 Program Implementer Qualifications: あ Yes あ No

If yes, describe _____

1.07 Avg. Daily Exposure _____hrs. Avg. Weekly Exposure _____hrs. Tot # Weeks _____ Duration _____

----------*For 1.08 – 1.12 check all that apply*----------

Focus: あReading あMath あWriting あScience あRecreational あVocational あCultural あMusic/Art あService Learning あOther _____

1.09 Format: あSummer school あAfter school あBefore school あExtended day あSaturday school あOther _____

1.10 Strategy: あ One-on-one tutoring あ Mentoring by staff/adult role model あ Independent learning あ Homework time/support あComputer-assisted instruction あTeacher directed instruction small grp. あTeacher directed instruction large grp. あLearning incentives あTeacher professional development あParent involvement あOther _____

1.11 Specific Reading Strategies: あShared reading あWord study あGuided reading あWriting & reading own stories
あReading & rereading book あComputer-assisted instruction あOther _____

1.12 Specific Math Strategies: あDrill & practice あProblem solving あManipulatives あComputer-assisted instruction
あOther _____

2. RESEARCH DESIGN

Student Sample Characteristics:

2.01 あ Random あPurposive あPopulation あOther_____

2.02 あ Control Group(s) *describe*_____

2.03 あComparison Group(s) *describe* _____

2.04 Total N in study _____ N in Treatment Group _____ N in Control Group(s) _____

N in Comparison Group(s) _____ , _____ , _____ , _____

2.05 Treatment Group Attrition _____% Control/Comparison Group Attrition _____%

Predominant Methodology *(the methodology on which conclusions are based):*

2.09 あQuantitative, quasi-experimental: *Check One:* あOne-group pretest-posttest あNonequivalent groups pretest-posttest あ Other _____

Characteristics used for equating or matching _____

_____

2.10 あQuantitative, experimental *(randomized trials)*: *Check One:* あposttest only あpretest-posttest

あOther _____

2.11 あQualitative: *Check All That Apply:* あCase study あAction research/Field Study あGrounded theory

あEthnography あOther _____

2.12 あSecondary methods:

Describe: _____

_____

3. QUANTITATIVE ANALYSIS

Outcome Measure Analysis

……

3.01a Measure: _____

Reliability reported: あ Yes (*measure & result):*_____ あ No

Group Characteristics	Treatment Group					Control/Comparison Group				
	N	Pretest		Posttest		N	Pretest		Posttest	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD

Unit of analysis: あ student あ class Direction of effect あ positive あ negative

Effect size _____ Test statistic(s) _____

3.01b Measure: _____

Reliability reported: あ Yes (*measure & result):*_____ あ No

Group Characteristics	Treatment Group					Control/Comparison Group				
	N	Pretest		Posttest		N	Pretest		Posttest	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD

Unit of analysis: あ student あ class Direction of effect あ positive あ negative

Effect size _____ Test statistic(s) _____

……

3.01c Measure: _____

Reliability reported: あ Yes (*measure & result):*_____ あ No

Group Characteristics	Treatment Group					Control/Comparison Group				
	N	Pretest		Posttest		N	Pretest		Posttest	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD

Unit of analysis: あ student あ class Direction of effect あ positive あ negative

Effect size _____ Test statistic(s) _____

……

3.01d Measure: _____

Reliability reported: あ Yes (*measure & result):*_____ あ No

Group Characteristics	Treatment Group					Control/Comparison Group				
	N	Pretest		Posttest		N	Pretest		Posttest	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD

Unit of analysis: あ student あ class Direction of effect あ positive あ negative

Effect size _____ Test statistic(s) _____

3.01e Measure: _____

Reliability reported: あ Yes (*measure & result):*_____ あ No

Group Characteristics	Treatment Group					Control/Comparison Group				
	N	Pretest		Posttest		N	Pretest		Posttest	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD

Unit of analysis: あ student あ class Direction of effect あ positive あ negative

Effect size _____ Test statistic(s) _____

……

3.02 Potential for Meta-analysis Synthesis

_____

_____

_____

3.03 Findings/Conclusions

What are the relevant findings/conclusions from this study that support the synthesis?:

_____

_____

_____

_____

_____

_____

_____

3.04 Quality of Quantitative Research

3.04a Construct Validity – Intervention

Was the intervention properly defined?

あ(3) Yes - the intervention was adequately described and it fully reflected commonly-held or theoretically derived ideas about what the intervention should be

あ(2) Maybe yes - the intervention was adequately described, and it at least largely reflected commonly-held or theoretically derived ideas about what the intervention should be

あ(1) Maybe no - there were important missing details in the description of the intervention and/or possible problems with its implementations

あ(0) No - the intervention did not reflect commonly-held or theoretically derived ideas about what it should be and/or there were known problems with its implementation

3.04b Intervention Fidelity

Was fidelity of intervention measured or discussed?

あ(3) Yes - fidelity measure was described and used
[There is no "Maybe yes" answer for this question.]

あ(1) Maybe no - issues of fidelity were discussed but unclear how it was measured

あ(0) No - issues of fidelity were not discussed

3.04c. Construct Validity – Outcome Measures

Was the outcome measure properly defined and aligned to the intervention?

あ(4) Yes - the report presented evidence that the outcome measure was properly defined and aligned to the intervention
[There is no "Maybe yes" answer for this question.]

あ(2) Maybe no - there was evidence of the face validity of the outcome measure, and it appeared to be properly aligned to the intervention, however, no evidence of construct validity was presented

あ(0) No - it is unclear what the outcome was

***NOTE: you will only have one of the following 3 in your study.**

3.04d. Internal Validity – Selection (for randomized experiments)

Were the participants (e.g., students, schools) in the group receiving intervention comparable to the participants in comparison grp?

あ (4) Yes - participants were randomly assigned to conditions and there was no differential attrition or severe overall attrition

あ (3) Maybe yes - random assignment was used but there was severe overall attrition

あ (2) Maybe no - random assignment was used but there was differential attrition

あ (0) No - although random assignment was used, both severe overall attrition and differential attrition probably led to the groups not being comparable

3.04e. Internal Validity – Selection (for quasi-experimental designs)

Were the participants (e.g., students, schools) in the group receiving the intervention comparable to the participants in the comparison group? (Comparison group can be a normed sample.) [There is no "Yes" answer for these types of designs.]

あ (3) Maybe yes - reasonable steps were taken to make the groups comparable and there was no attrition problem OR the groups were demonstrably equivalent* and there was no attrition problem

あ (2) Maybe no - although steps were taken to make the groups comparable, the steps may not have been adequate

あ (0) No - it is unlikely that the participants in the groups were comparable (or there were no comparisons groups)

3.04f. Internal Validity – Selection (for Regression Discontinuity Designs)

Were the participants (e.g., students, schools) in the group receiving the intervention comparable to the participants in the comparison group (that is the slopes of regression lines were similar on the assignment variable)?

あ (4) Yes - an assignment variable with specified cutoff(s) was used to place participants into groups and there was no attrition problem

あ (3) Maybe yes - an assignment variable with specified cutoff(s) was used but severe attrition may have affected study results

あ (2) Maybe no - an assignment variable with specified cutoff(s) was used but differential attrition may have affected study results

あ (0) No - an assignment variable w/specified cutoff(s) wasn't used to place participants into groups

3.04g. Internal Validity – Contamination

Was the study free of events that happened concurrently with the intervention that confused its effect?

あ (3)Yes - concurrent processes and events that might be alternative explanations to the intervention's effect have been ruled out

あ (2) Maybe yes - there were no identified processes or events that could be alternative explanations, but some alternative explanations remain plausible. [There is no "maybe no" answer for this question.]

あ (0) No - identifiable processes happening at the same time as the intervention may have caused the effect

3.04h. External Validity – Sampling

Were targeted participants, settings, outcomes, and occasions included in the study?

あ (3) Yes - the targeted participants, settings, outcomes, and occasions are represented in the sample

あ (2) Maybe yes - most important characteristics of the participants, settings, outcomes, and occasions are represented in the sample

あ (1) Maybe no - although some important characteristics of the participants, settings, outcomes, and occasions* are represented in the sample, many important targets are not

あ (0) No - the sampled participants were not part of the target population

3.04i. External Validity – Testing within Subgroups

Was the intervention tested for its effectiveness within important subgroups of participants, settings, outcomes, occasions, and intervention variations?

あ (3)Yes - the intervention was tested for its effectiveness on targeted participants, settings, outcomes, occasions, and intervention variations

あ (2) Maybe yes - the intervention was tested for its effectiveness within most important subgroups of the participants and settings

あ (1) Maybe no - although the intervention was tested for its effectiveness within some important subgroups of the participants and settings many were left out

あ (0) No - at best the intervention was only tested for its effectiveness within limited important subgroups of the participants, settings, outcomes, occasions, and intervention variations

3.04j. Statistical Validity – Effect Size Estimation and Completeness of Reporting

This was a combination of two of original criteria from the What Works Clearninghouse.

あ (3) Yes - the effect sizes were reported for all outcomes and appear to be accurately estimated

あ (2) Maybe yes - sufficient statistical information was reported to allow precise effect size calculations for most measured outcomes

あ (1) Maybe no - effect sizes can be calculated only for some outcome measures due to insufficient reporting

あ (0) No - no effect sizes can be calculated due to the lack of crucial statistical information or reported effect sizes are inaccurately estimated

3.05 Total Score = ____ QUALITY INDEX: あ Low (0-14) あ Medium (15-21) あ High (22-26) *(enter index on first page)*

4. QUALITATIVE DESIGNS

4.01 Purpose of qualitative approach: あ Theory building あ Interpretive/descriptive あ Other _____

4.02 Data collection methods used (check all that apply):

あ Nonparticipant observations あ Participant observations あ Focus groups あ Interviews

あ Document review あ Questionnaires あ Other _____

4.03 Data Analysis/Analyses (check all that apply):

あ Content analysis あ Constant comparative method あ Inductive

あ Other _____

4.04 Findings/Conclusions:

What are the relevant findings/conclusions from this study that support the synthesis:

_____

_____

_____

_____

_____

4.05 Quality of Qualitative Research:

Confirmability/statistical conclusion validity *(the ability for others to examine all data sources and processes to assure that the findings are grounded in the data)*

4.05a Were any of the following used in the study?* あ Yes (2) あ Partially (1) あ No (0)

***(if 2 or more were used rate as yes, if only 1 was used rate partially)**

あ Member checking あ Audit trail あ Expert/peer review

4.05b Did the researcher control for researcher effects? あ Yes(3) あ Partially (2) あ No (0)

***(if 2 or more were used rate as yes, if only 1 was used rate partially)**

あ Used unobtrusive measures

あ Disclosed purpose of study and intentions to informants

あ Included variety of informants

あ Triangulated data from two or more sources

Dependability/Construct validity *(the use of methods and techniques to assure that the study's results can be trusted)*

4.05c Are the research questions clear, and are the features of the study design congruent with them?

あ Yes (3) あ Partially (2) あ No (0)

4.05d Were data collected across the full range of appropriate settings, time, respondents, and so on suggested by the research questions?

あ Yes (4) あ Partially (3) あ No (0)

4.05e Are basic paradigms and analytic constructs clearly specified?

あ Yes (3) あ Partially (2) あ No (0)

Credibility/Internal Validity *(The findings are credible to the reader and the researcher has used techniques to ensure the credibility of findings.)*

4.05f Are multiple sources of evidence and/or data collection methods used to produce converging conclusions?

あ Yes (4) あ Partially (3) あ No (0)

If no, is there a coherent explanation for this? あ Yes (3) あ Partially (2) あ No (0)

4.05g Were any of the following conducted?

あ Yes(4) あ Partially (3) あ No (0)

***(if 2 or more were used rate as yes, if only 1 was used rate partially)**

あ Search for disconfirming evidence

あ Generation of rival hypotheses/explanations

あ Negative case analysis

あ Other _____

4.05h Do the presented data and measures reflect the constructs or categories of prior or emerging theory?

あ Yes (3) あ Partially (2) あ No (0)

Transferability/External Validity *(the provision of sufficient "thick description" to enable the reader to decide whether the concepts or themes can be transferred to another setting)*

4.05i Are the characteristics of the original sample of persons, settings, processes (etc.) fully described enough for readers to assess the potential transferability, appropriateness for their own settings?

あ Yes (3) あ Partially (2) あ No (0)

4.05j Does the researcher define the scope and the boundaries of reasonable generalization from the study?

あ Yes (2) あ Partially (1) あ No (0)

4.05k Total Score = __ QUALITY INDEX: あ Low (0-9) あ Medium (10-21) あ High (22-31) *(Enter index on first page.)*

- END

Appendix B: Meta-analysis Methods

Meta-analysis is a research method that quantitatively summarizes and analyzes the results of past studies on the effectiveness of a practice or policy (Cooper, 1998; Cooper & Hedges, 1994; Hedges & Olkin, 1985). For the current synthesis, we used meta-analytic techniques to examine the effectiveness of out-of-school-time (OST) strategies for improving the reading and mathematics achievement of low-achieving or at-risk students. To assist with data analysis and presentation, we used Comprehensive Meta Analysis, a stand-alone software program developed in 1999 by Biostat®.

Meta-analysis generally requires four steps: (1) computation of an effect size for each research study in the synthesis, (2) computation of an overall effect size across the research studies, (3) homogeneity analysis, and (4) moderator analysis. The following sections describe each step in the context of the current synthesis.

EFFECT SIZES FOR INDIVIDUAL STUDIES

An effect size is a standardized estimate of the effectiveness of the practice or policy that is investigated in a research or evaluation study. An effect size is measured by a *d*-index, which refers to the standardized mean difference. For example, a study with $d = 1.00$ indicates that the mean achievement of students who experienced the OST strategy under investigation is one standard deviation higher than the mean achievement of students in the control group who did not experience the strategy. The closer the *d*-index is to zero, the less is the effect of the strategy under investigation, and a negative sign indicates that the strategy is associated with lower scores on the outcome measure.

Effect sizes can be computed from various kinds of quantitative information including means with standard deviations, and *t*, *F*, or chi square values from inferential statistical tests (Cooper, 1998). The sample sizes of treatment and control groups also are required for effect size computation. Most of the studies we used for meta-analyses in this synthesis reported means, standard deviations, and the necessary sample sizes. There are formulas for estimating effect sizes (Rosenthal, 1991), and Comprehensive Meta-analysis calculates Cohen's *d* and Hedges *g*, both common measures. We chose to report Hedges *g* because it adjusts for small sample sizes (Rosenthal, 1991).

For studies with pretests and posttests, we computed separate effect sizes for each test and subtracted the pretest effect size from the posttest effect size to estimate the overall effect (Blok, Oostdam, Otter, & Overmaat, 2002). Some studies reported only the gain or difference scores, which were used to calculate the effect size directly. For studies without reported pretest-posttest scores or gain scores, the posttest scores

were used to compute the effect size for the study. We included type of score in the moderator analysis to assess its influence on effect sizes. For all studies, we used the pooled standard deviation from the treatment and control groups to reflect the different standard errors and sample sizes (Hedges & Olkin, 1985).

While some studies reported an outcome based on a single sample, other studies reported results for multiple independent samples. For example, a study might report separate mathematics score gains for 20 fourth-grade and 20 fifth-grade students. In this case, the study has two independent samples, and two effect sizes can be computed. However, another study might report 40 fifth-grade students' gains in computation skills and problem solving. These two outcomes are not independent because they are from the same students. In this case, the mean of the two effect sizes is the single effect size for the study. In the studies we synthesized, the number of independent samples in a study varied from one to five.

Overall Effect Size Across Studies

Data from independent samples were used to compute the overall effect size. The effect size(s) from each study was weighted by sample size based on the general assumption that studies with larger sample sizes produce more reliable estimates of effects. We examined the distribution of effect sizes for statistical outliers by identifying *d*-values that were more than three interquartile ranges beyond the *d*-value that was at the 75th percentile in the distribution (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). Using this method, there was one outlier identified for reading and none for mathematics. The reading outlier was changed to the *d*-value at the 75th percentile of the distribution of the reading effect sizes. This change did not influence the meta-analysis results compared to results without the adjustment.

In computing the overall effect size, we employed both fixed-effects and random-effects models (Cooper, 1998). There is a debate among meta-analysts over which method provides a more accurate estimate of effect size. As Cooper indicates, the fixed-effects model assumes that the only random influence on effect sizes is sampling error (i.e., chance factors related to the students in a study). The random-effects model assumes that effect sizes also are influenced by chance factors related to other influences (e.g., OST program staff, schools, family characteristics, etc.). To be conservative, we reported lower and upper limits of the 95 percent confidence interval based on both the fixed-effects model and the random-effects model. For both models, if the 95 percent confidence interval around the overall effect size did not include zero, the null hypothesis that OST strategies had no effect on student achievement was rejected. In other words, the effect of OST strategies was statistically different from zero.

Homogeneity Analysis

Homogeneity analysis determines whether the effect sizes from the selected studies vary more than expected by sampling error alone. If the resulting *Q* statistic, which is based on a chi square distribution, is statistically significant, it means that the effect sizes are not homogenous, and moderating factors that might explain the variation across studies should be identified. Because our homogeneity analyses were statistically significant for both the reading and mathematics meta-analyses, we proceeded with moderator analyses.

MODERATOR ANALYSIS

Based on the research problem and questions that our synthesis addressed, we examined how effect sizes varied by the following characteristics of OST strategies: timeframe, grade level, activity focus, program duration, and student grouping. We also examined how effect sizes varied by three characteristics of the research studies in the meta-analyses: research quality, type of publication, and type of score.

We conducted homogeneity analyses to examine the amount of variation across average effect sizes based on each moderator variable (e.g., average effect sizes for summer school and after school timeframes).[18] A statistically significant *Q* indicates that the variation across average effect sizes of the different levels of a moderator variable is greater than expected by sampling error alone. In other words, the moderator has a statistically significant influence on the overall effect size for the meta-analysis. When more than one moderator is statistically significant, it is possible that the moderators are correlated (Cooper, 1998). In interpreting our results, we examined correlation matrices of the moderators for possible interrelationships.

Reading and Mathematics Meta-Analyses

There was sufficient quantitative information in the reviewed studies to conduct separate meta-analyses for reading and mathematics outcomes. We chose not to combine them due to our interest in isolating the effects of OST strategies related to the two content areas and the need to discuss and interpret the effects in the context of each content area. As a result of this approach, 14 studies provided data for both meta-analyses. For these cross-chapter studies, separate effect sizes were computed for reading and mathematics from the same sample of students. Because our goal was to describe the effectiveness of OST strategies separately for the two content areas

[18] We conducted homogeneity analyses of effect sizes based on the fixed-effects model only.

and not to compute one overall effect size for all studies, this occurrence did not bias our results. However, we used caution in drawing overall conclusions across the two content areas.

REFERENCES

Blok, H., Oostdam, M. E., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72*(1), 101–130.

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, Serial No. 260, 65*(1).

Cooper H. (1998). *Synthesizing research* (3rd ed). Thousand Oaks, CA: SAGE Publications.

Cooper, H. & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: SAGE Publications.

Appendix C: Annotated Bibliography

The annotated bibliography provides information on studies in the synthesis that describe examples of effective out-of-school time (OST) programs for low-achieving students. References were chosen for annotation based on the following criteria:

1. The study describes the nature of the OST strategy and its implementation.

2. The study describes evidence of positive impact from the OST strategy on student achievement in reading, mathematics, or both.

3. As a body, the annotated studies address the range of students in grades K–12.

Studies in both the meta-analyses and the narrative reviews were considered for annotation. The annotations for meta-analyzed studies report effect sizes as appropriate. The annotations are presented separately for studies from the reading and mathematics chapters of the synthesis.

READING STUDIES

Duffy, A. M. (2001). Balance, literacy acceleration, and responsive teaching in a summer school literacy program for elementary school struggling readers. *Reading Research and Instruction, 40*(2), 67–100. (ERIC Document Reproduction Service No. EJ 624 633)

Ten underachieving, second-grade students participated in this qualitative study of a summer school program that used a balanced, accelerated, and responsive approach to literacy instruction. Students participated in word study, guided reading, book talks, and read-alouds with the teacher, and wrote and read their own stories. As a result of participating in the program, students improved their word identification abilities, became more fluent in oral reading and writing, increased their instructional reading levels, and became more strategic in reading comprehension. Students also developed more positive attitudes toward reading and had more positive perceptions of themselves as readers.

Jacob, B. A., & Lefgren, L. (2001). ***Remedial education and student achievement: A regression-discontinuity analysis*****. Boston: National Bureau of Economic Research. (ERIC Document Reproduction Service No. ED 465 007)**

The researchers analyzed five years of longitudinal data from the Chicago Public Schools that examined the effects of summer school and grade retention on students failing to meet end-of-grade achievement standards. Summer school was mandatory for failing students in the Chicago Public Schools from 1997–1999. If, after summer school, students again failed to meet end-of-grade achievement standards, grade retention was required (this applied to 10 to 20 percent of the summer school attendees). Each teacher taught a small class (15 students) using required highly structured curriculum and resource materials from the district. The findings indicated that summer school, independent of retention, had significant and positive effects on reading achievement for grade 3 students but not for grade 6 students. The average gain for grade 3 students attending summer school was an estimated 12.5 percent of the annual learning gain.

Leslie, A. V. L. (1998). The effects of an after-school tutorial program on the reading and mathematics achievement, failure rate, and discipline referral rate of students in a rural middle school (rural education). (Doctoral dissertation, University of Georgia, 1998). ***Dissertation Abstracts International, 59*****, 06A.**

This quasi-experimental study examined the effectiveness of after-school tutoring for middle school students who performed poorly on achievement tests or classroom assignments and/or had disciplinary problems. The after-school program combined one-on-one tutoring, homework time/support, computer-assisted instruction, learning incentives, and practice with skill-builder worksheets. The tutors met frequently with classroom teachers who directed the content of the tutoring, and in other cases, classroom teachers themselves provided the tutoring to students who were also in their classes during the day. The program had highly positive effects on students' reading achievement (*d*s = .90, .88 and 2.35 for grades 6, 7 and 8 respectively) and also on their achievement in mathematics. However, it is likely that program effectiveness is linked with student motivation. (Treatment group students attended at least 50 days of the after-school program; students in the control group were students who chose not to attend the after-school program.)

Luftig, R. L. (2003, May). ***When a little bit means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers.*** **Paper presented at the American Educational Research Association, Chicago, IL.**

Ninety-two at-risk elementary students participated in one of two types of summer school reading intervention programs over a three-week period. Both were phonics-based programs that used tutoring instruction. One program was designed by a for-profit company that also used computer-assisted instruction, whereas the other was a locally developed program that tied the phonics instruction to the district curriculum. Students in both treatment groups significantly outperformed students in control groups. The study suggests that at-risk students can benefit from reading remediation with a minimal amount of intervention time (e.g., nine hours).

Morris, D., Shaw, B, & Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. ***Elementary School Journal, 91*****(2), 133–150.**

In this study, 60 low-achieving second and third graders were randomly assigned to either after-school tutoring or a control condition of no tutoring. The year-long after-school tutoring was provided by community volunteers who were supervised by a reading specialist. The supervisors designed each tutoring lesson, and the tutors implemented the lessons and recorded observations for the supervisor; the supervisor, in turn, designed subsequent lessons. Tutorial strategies included shared reading, word study, reading books, and writing stories. The overall positive effect of after-school tutoring on reading achievement (d = .50) required 50 hours of "well-planned, closely supervised one-to-one tutoring" (p. 147).

Ortiz, G. K. (1993). ***An exploratory study of the effects of an after-school literacy enrichment program on at-risk students and their parents (literacy development).*** **(Doctoral dissertation, University of South Carolina, 1993).** ***Dissertation Abstracts International, 54*****, 09A.**

This was a qualitative study that used grounded theory to examine the effects of an after-school literacy program for at-risk first-grade students and their parents. Parents participated in instructional sessions focused on improving their abilities to support their children's literacy development at home. Students learned techniques to improve their literacy abilities and had opportunities to practice those techniques during collaborative reading sessions with parents. The findings revealed that students' reading abilities

improved when they read for relevant purposes, were active participants in the reading process in a risk-free environment, and could share in fun reading activities with their parents.

Rembert, W. I., Calvert, S. L., & Watson, J. A. (1986). Effects of an academic summer camp experience on black students' high school scholastic performance and subsequent college attendance decisions. *College Student Journal, 20*(4), 374–384.

The authors evaluated an academic summer camp for 10th–, 11th–, and 12th–grade students with "evidence of college level academic potential, but low motivation or intention toward postsecondary education" (p. 376). For 3–4 weeks, each of 2–3 summers, students lived in dormitories on a college campus, attended classes, used college library facilities, and experienced a college atmosphere. The college preparation classes focused on skill mastery in basic academics and simulated college instruction. Assistance with career planning and study skills instruction was also provided. Compared to the control group, participants in this academic summer camp demonstrated higher reading achievement (d = .51) and were more likely to enter college. There were positive effects on mathematics achievement as well (d = .34).

Roderick, M., Engel, M., & Nagaoka, J. (2003). *Ending social promotion: Results from summer bridge*. Chicago: Consortium on Chicago School Research.

Third-, sixth-, and eighth-grade students in Chicago Public Schools who failed promotion criteria attended summer school. The summer school participants' achievement was examined in relation to that of a comparable group of students over the course of four years. Summer school participants experienced, on average, a seven percent gain in achievement that lasted over the four years. This boost narrowed achievement gaps but did not allow the target groups to catch up to the levels of achievement demonstrated by peers who passed promotion criteria. Teaching competence was reported to have made a difference, and greater achievement gains were found in classes taught by teachers who were more active in teaching and in individualizing instruction. In all summer school classes, a curriculum aligned with the high-stakes assessment test was used. Monitors checked teachers' pacing and implementation of lessons.

Schacter, J. (2001). ***Reducing social inequality in elementary school reading achievement: Establishing summer literacy day camps for disadvantaged children*****. Santa Monica, CA: Milken Family Foundation. Retrieved June 4, 2003, from http://www.mff.org/pubs/reading_camp_study2001.pdf**

Twenty-one, disadvantaged first-grade students participated in an eight-week, summer day camp that promoted social and emotional growth and implemented a systematic reading curriculum with one-on-one tutoring and recreational activities. Students participated in two hours of reading instruction per day with a credentialed reading teacher and participated in one hour of tutoring each week with a tutor. The treatment group showed significant reading improvement compared to control students (d = .73). The author identified the summer camp context as instrumental to the success of the program.

MATHEMATICS STUDIES

Baker, D., & Witt, P. A. (1996). Evaluation of the impact of two after-school programs. ***Journal of Park and Recreation Administration, 14*****(3), 23–44.**

A group of 302 third- through sixth-grade students from low-income communities participated in two Austin, Texas, after-school programs. The teachers were paid a stipend to facilitate a wide variety of program activities including sports skills classes, arts and crafts, drama, computer classes, cooking, cultural activities, and academic classes and field trips. Each after-school program lasted two hours after the end of the school day, Monday through Friday. The authors reported positive effects on mathematics achievement on the state assessment (d = .31) for the student participants as a result of their participation in after-school programming for one school year. (There were also positive effects on reading achievement, with d = .30.)

Cosden, M., Morrison, G., Albanese, A. L., & Macias, S. (2001). When homework is not home work: After-school programs for homework assistance. ***Educational Psychologist, 36*****(3), 211–221.**

The authors conducted a study of the Gevirtz Homework Project over a three-year time span in three elementary schools in the Santa Barbara School District (CA). The goals of this after-school project were to provide students with structured time to complete assignments and to provide the student participants with both implicit and explicit instruction in study skills. The homework/study sessions were facilitated by classroom teachers and were offered three or four times each week. The authors report that the 32 fourth

through sixth graders who attended the sessions demonstrated significant academic gains (d = .84 for mathematics and d = .95 for reading). (For more information, see http://www.education.uscb.edu/grc/homework.html.)

Huang, D., Gribbons, B., Kim, K. S., Lee, C., & Baker, E. L. (2000). ***A decade of results: The impact of the LA's Best After School Enrichment Initiative on subsequent student achievement and performance.*** **Los Angeles: UCLA Center for the Study of Evaluation, Graduate School of Education & Information Studies, University of California.**

This study is a longitudinal evaluation of a large-scale Los Angeles program known as LA's BEST (Better Educated Students for Tomorrow). This after-school program began in 1988 in answer to rising rates of gang affiliation, drug use, and dropouts. The number of original program sites of 10 has grown to more than 100 as the program continues to gather political and community support in terms of funding and volunteers. The study examined the academic performances of 4,312 students who attended LA's BEST in the 1990s. The students were exposed to homework help sessions, academic tutoring, library activities, and in some cases, remedial instruction. The authors reported mostly positive results in both mathematics and reading on standardized tests for those who had regularly attended the three-hour after-school sessions. (For more information, see http://www.lasbest.org.)

McMillan, J. H., & Snyder, A. L. (2002, April). ***The effectiveness of summer remediation for high-stakes testing.*** **Paper presented at the American Educational Research Association, New Orleans, LA.**

The authors evaluated the effectiveness of a summer school program in Virginia by comparing state test results to district survey information. Sixty-three students failed the ninth-grade state test in the spring of 2001 and retook the test at the end of that summer. The authors reported that summer school programming could account for a large academic effect, particularly in the test subjects of Algebra and World History (d = 1.56 and d = 1.29 respectively).

Riley, A. H. J. (1997). Student achievement and attitudes in mathematics: An evaluation of the twenty-first century mathematics center for urban high schools (urban education, summer school). (Doctoral dissertation, Temple University, 1997). ***Dissertation Abstracts International, 58*****, 06A.**

Students from low-income families attended summer school on the campus of Temple University in the Twenty-first Century Mathematics Center for Urban High Schools. The high school student participants were taught mathematics in both large-classes and tutoring groups and were expected to complete a series of assigned worksheets as they progressed through the

curriculum expectations. The author reported positive academic effects for the 78 participants in the study (d = .83 for the male participants and d = .99 for the females).

Schinke, S. P., Cole, K., & Poulin, S. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science, 1*(1), 51–60.

This was the study of Boys and Girls Clubs of America after-school programs in five urban centers. The 283 fifth- through eighth-grade participants who were residents of public housing were exposed to four to five hours of programming after school each day. The program activities included discussion groups, creative writing sessions, homework help, peer tutoring, and recreational activities. The authors reported mostly positive effects on mathematics achievement (and also reading achievement) of the 283 participants based on teacher reported gains and class-score gains.

Sipe, C. L., Grossman, J. B, & Milliner, J. A. (1988). *Summer training and education program (STEP): Report on the 1987 experience.* Philadelphia: Public/Private Ventures. (ERIC Document Reproduction Service No. ED 300 479)

The Summer Training and Educational Program (STEP) was designed to promote high school graduation and successful transition to careers. The STEP program was an addition to a federal summer jobs program in 1986. Thousands of students participated in the five urban programs during the summers of 1986 through 1988. These students were exposed to academic classes, and life and career counseling, interventions. The researchers documented mostly positive academic effects in both mathematics and reading for the 1,272 participants included in the study.

Welsh, M. E., Russell, C. A., Williams, I., Reisner, E. R., & White, R. N. (2002). *Promoting learning and school attendance through after-school programs: Student-level changes in educational performance across TASC's first three years.* Washington, DC: Policy Studies Associates.

The authors studied 96 sites of a large-scale New York City after-school program. The After-School Corporation (TASC) works to increase the availability and quality of programming for New York's most disadvantaged children in terms of poverty, achievement, and minority status. There were positive significant effects on mathematics achievement reported for 183 students in elementary and middle school who actively participated for two years (d=.24). The TASC website describes the intervention:

TASC-supported programs include educational enrichment through activities in language arts, science, mathematics, fine and performing arts, and sports. Curricula include homework help and build upon and enhance the students' school day experience and support the Department of Education's performance standards. In addition, TASC stresses computer education and health and social development, covering subjects such as drug prevention and nutrition. (Retrieved October 21, 2003, from http://www.tascorp.org)