Review of Educational Research
http://rer.aera.net

Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students

Patricia A. Lauer, Motoko Akiba, Stephanie B. Wilkerson, Helen S. Apthorp, David Snow and Mya L. Martin-Glenn

REVIEW OF EDUCATIONAL RESEARCH 2006 76: 275
DOI: 10.3102/00346543076002275

The online version of this article can be found at:
http://rer.sagepub.com/content/76/2/275

Published on behalf of

American Educational Research Association

and

http://www.sagepublications.com

Review of Educational Research
Summer 2006, Vol. 76, No. 2, pp. 275–313

Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students

Patricia A. Lauer
Rocky Mountain Center for Health Promotion and Education
Motoko Akiba
University of Missouri, Columbia
Stephanie B. Wilkerson
Magnolia Consulting, LLC
Helen S. Apthorp
Mid-continent Research for Education and Learning
David Snow
Billings Public Schools
Mya L. Martin-Glenn
Mid-continent Research for Education and Learning

Schools and districts are adopting out-of-school-time (OST) programs such as after-school programs and summer schools to supplement the education of low-achieving students. However, research has painted a mixed picture of their effectiveness. To clarify OST impacts, this synthesis examined research on OST programs for assisting at-risk students in reading and/or mathematics. Researchers analyzed 35 OST studies that employed control or comparison groups and met other inclusion criteria. Meta-analyses indicated small but statistically significant positive effects of OST on both reading and mathematics student achievement and larger positive effect sizes for programs with specific characteristics such as tutoring in reading. Whether the OST program took place after school or during the summer did not make a difference in effectiveness.

KEYWORDS: at-risk students, math achievement, meta-analysis, out-of-school-time programs, reading achievement.

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 supplemental educational services. These services must occur outside the school day and be backed by evidence that the services are effective in raising student achievement (No Child Left Behind Act of 2001, section 1116[e]). Our study responds to this need for evidence through a synthesis of research on the effectiveness of OST programs in assisting at-risk students in reading and mathematics, the content areas emphasized by NCLB.

As we and other researchers have found, OST programs 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 questions:

- Based on rigorous research and evaluation studies, what is the effectiveness of OST programs in assisting at-risk students in reading and mathematics?
- How does the effectiveness of OST differ by program and study characteristics?

Out-of-School Time

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 and during the summer.

After-School Programs

According to De Kanter (2001), 6 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 8 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 (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 in the early 1900s 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 starting in the 1940s. Halpern noted that only recently have policymakers suggested after-school programs as a way to improve student achievement, a policy that Halpern opposes because of 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 economically 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 effects 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 have devised creative programming strategies (Grossman, Walker, & Raley, 2001), a result that has contributed to the variation in content among after-school programs.

Summer Schools

A report by Cooper, Charlton, Valentine, and Muhlenbruck (2000) described the history and goals of summer school. As in the case of 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, in comparison with after-school programs, summer school programs tend to be more oriented toward academic improvement and less oriented toward multiple goals.

OST Programs and Low-Income Children

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, in comparison with 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). 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.

Prior Research on OST Programs

Prior reviews related to OST programs informed the current synthesis. Cooper et al. (2000) reported on a 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 than did 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 programs, 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 students who entered programs with higher achievement, and that students who attended the programs more frequently benefited more than those with lower attendance. 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 well-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 academically 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 outcomes 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 types of outcomes measured. The current synthesis examines OST programs with academic and other focuses 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 review was to examine the roles of after-school programs in promoting academic success and positive early adolescent development. Miller described the effects of various 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.

Research on Moderators of OST Effectiveness

Based on the research literature related to OST and student achievement, we identified the following program characteristics as possible moderators of OST effectiveness: timeframe, grade level, program focus, program duration, and student grouping. Timeframe refers to whether the OST program 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 than for 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).

Because of the wide variation in the focuses and goals of OST programs, it is logical to investigate whether the degree to which an OST program focuses on academics might influence its effectiveness 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 economically disadvantaged students. However, others suggest that, to be effective, strategies for economically 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 school is not likely to be beneficial. Miller supports a wide variety of activities for OST learning programs.

We identified the duration of an OST program 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. For this meta- analysis we examined exposure effects through the proxy variable of program duration because of incomplete attendance data in the studies that we examined. 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 (Karweit, 1985; Ascher, 1990).

The final program characteristic that 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 Policy Studies Associates (1995) report 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 programs, we also looked at three characteristics of studies. As mentioned previously, researchers (Scott-Little et al., 2002; Fashola, 1998) have identified the need for higher-quality research of OST programs. 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 quality ratings. Another study moderator in this synthesis was the type of publication, such as peer-reviewed journal article or dissertation. As Cooper (1998) indicated, peer-reviewed journals are more likely to publish research studies that report statistically significant effects than studies that support the null hypothesis. A final study moderator 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, and 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.

To summarize, the current synthesis contributes to the knowledge base about OST programs for at-risk students in several ways. We searched for and examined research on programs delivered in all OST timeframes, including summer school, after school, before school, vacation sessions, and Saturday schools. To lend support to our conclusions, studies were included in the synthesis only if they used a comparison group of students who did not experience the OST program under investigation. In addition, studies were rated for alignment with criteria of research quality, and synthesis results are described in relation to these ratings. There are separate meta-analyses of the influences of OST programs on student achievement in reading and in mathematics, which enhance prior narrative reviews of research on after-school programs and which provide comparisons with prior research on summer schools. Finally, using meta-analytic techniques, we describe the influences of several OST moderators, including timeframe, student grade level, program focus, program duration, and student grouping.

Method

Literature Searches

In May 2003, we conducted 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 PsycINFO database subsequently was searched for "supplementary" (41 citations), "summer school" (57 citations), "after school" (207 citations), and "vacation programs" (3 citations), resulting in a total of 308 citations. We searched the titles of Dissertation Abstracts with parameters of 1985–2003, not college, English language only, and PhD dissertations only, for "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, 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, 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.

Inclusion Criteria

We used the following criteria for including studies:

- Studies had to concern an OST program for K–12 students. We defined an OST program as an education intervention delivered outside the regular school day.
- Studies had to be published or reported in or after 1985 and implemented in the United States.
- Studies had to include some type of direct assessment of students' academic achievement in reading, mathematics, or both. Examples included classroom assessments, standardized tests, and grades in subject areas.
- Studies had to examine the effectiveness of an OST program for students who are at risk for school failure. For purposes of this synthesis, we defined at-risk status as (a) low student performance on standardized tests, classroom assessments, or teacher-assigned grades; or (b) characteristics typically associated with lower student achievement and school dropout in large-scale data collections, including low socioeconomic 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).

- Studies had to include a control/comparison group, which we defined as a group of students who did not participate in the OST program under investigation and whose achievement results were compared with those for students who did participate.
- Studies had to disaggregate student results for specific OST programs. Five studies were excluded because they aggregated data statewide or nationally so that results could not be connected to specific programs.
- Studies were not included if they examined OST programs designed for and delivered only to special populations such as special education students, English language learners, and migrant students.
- Studies could be published or unpublished, including evaluation reports, conference presentations, and dissertations.
- For purposes of meta-analysis, studies had to include sufficient quantitative information for calculation of effect sizes (Cooper, 1998).

We read each article that was ordered and received by July 16, 2003. Thirty-five studies met the criteria for inclusion, 30 with reading outcomes and 22 with mathematics outcomes. Of these, 17 studies addressed outcomes in both subject areas. There were 336 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 at-risk students as defined for this synthesis. There were 18 studies excluded solely because they did not have sufficient information for meta-analysis.

Coding of Studies

Each study was coded for information about the OST program that was implemented, the student sample, the research design, statistical results, and research quality. Program information included the nature of the program (e.g., homework help, one-on-one tutoring), content (e.g., reading, mathematics), focuses (academic, social), timeframe (e.g., after school, summer school), descriptions of specific strategies related to reading or mathematics, qualifications of those implementing the OST programs, and program duration. The last was defined as the total number of program hours made available to students who attended the program. This was calculated as the number of hours of programming each day times the total number of days the program was offered.

Student information included how the study identified them as at risk and student characteristics of gender, ethnicity, and grade level. We coded the grade levels of students using four categories: lower elementary (K–2), upper elementary (3–5), middle school (6–8), and high school (9–12). 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 kindergartners through third graders and was categorized as lower elementary rather than upper elementary.

We described the research design 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. Designs were coded for whether

students were pretested on achievement prior to program implementation and posttested afterward or only posttested. Statistical results were coded for each outcome measure for each student group in the study and included the information needed to conduct a meta-analysis: sample sizes, group means and standard deviations, effect sizes, and inferential test statistics.

To code the research quality of studies, we used Shadish, Cook, and Campbells's (2002) framework on threats to validity and the Study Design and Implementation Assessment Device developed by Valentine and 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 program) 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 study report. 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). Table 1 describes the characteristics of a study with a "medium" quality rating, which is the rating that was assigned to the majority of the studies in this synthesis.

Coding Procedures

Coding procedures incorporated Stock's (1994) recommendations for reducing coding errors. Each of the authors participated in coder training, which involved an overall description of the coding form, explanations for items in each section,

TABLE 1
Characteristics of a medium-quality study

Type of validity	Study characteristics[a]
Construct validity	The description of the intervention is adequate and largely reflects commonly held ideas about its definition.
	Treatment fidelity is discussed, but there is no report of its assessment.
	There is evidence for the alignment of the outcome measure with the intervention and for construct validity of the outcome measure.
Internal validity	The steps taken to make student groups comparable may have been inadequate.
	There were no identified processes or events that could be alternative explanations, but some alternative explanations are plausible.
External validity	Most of the 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 outcome measures.

[a]Study characteristics are on based on Valentine and Cooper's (2003) *What Works Clearinghouse Study Design and Implementation Assessment Device.*

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 four studies that had both reading and mathematics student outcomes. The authors then compared completed forms, identified and resolved discrepancies, and revised the coding form for 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. Based on their content expertise, two different pairs of authors coded the studies for reading and mathematics. Within each pair, authors reached consensus on coding outcomes. Coding results for studies included in both the reading and mathematics meta-analyses were compared, and any discrepancies were resolved. The first author reviewed the studies independently of the author pairs and verified the accuracy of the study codes entered in the meta-analysis databases.

Results

Meta-Analytic Methods

Separate meta-analyses were conducted for studies with reading and mathematics student outcomes and involved four steps: (a) computation of an effect size for each research study, (b) computation of an overall effect size across the research studies, (c) homogeneity analysis, and (d) moderator analysis. To assist with data analysis and presentation, we used Comprehensive Meta-Analysis, a stand-alone software program developed in 1999 by Biostat®. We report Hedges *g* as the measure of effect sizes, which Comprehensive Meta-Analysis calculates to include the adjustment 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 we 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. For all studies, we used the pooled standard deviation from the treatment and control groups (of either the gain scores or the posttest scores, depending on the measure) 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 the latter, the mean of the effect sizes is the single effect size for the study. The number of independent samples in studies varied from one to five.

Data from independent samples were used to compute the overall effect sizes for reading and mathematics studies. The effect size or sizes from each study were 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 those that were more than three interquartile ranges beyond the effect size that was at the 75th

percentile in the distribution (Cooper et al., 2000). Using this method, we identified one outlier for reading (Leslie, 1998) and none for mathematics. The reading outlier was changed to the effect size value at the 75th percentile of the distribution of the reading effect sizes. This change did not influence the meta-analysis results in comparison with the results without the adjustment, so the original meta-analysis is reported here. In computing the overall effect sizes, we employed both fixed-effects and random-effects models (Cooper, 1998). There is debate among meta-analysts about which method provides a more accurate estimate of effect size, so we report lower and upper limits of the 95% confidence interval based on both models.

Homogeneity analyses were conducted to determine whether the effect sizes from the studies varied more than expected by sampling error alone. For both reading and mathematics, the resulting *Q* value was statistically significant, indicating that the effect sizes were not homogenous. We therefore proceeded with moderator analyses to identify factors that might explain the variation in effect sizes across studies. To analyze this variance, we used a fixed-effects model and calculated a *Q* statistic for each moderator. As Lipsey and Wilson (2001) indicate, this method is appropriate for categorical moderator variables such as those in this synthesis and "is best suited to testing a limited set of a priori hypotheses regarding moderator variables" (p. 120). A weighted regression analysis is another approach to moderator analysis that is used when there are multiple moderators measured by continuous variables, which was not the case here. The total number of effect sizes analyzed for each moderator depended on the unit of analysis and whether there was sufficient information to code the study for the moderator. The units of analysis for the moderator of grade level were 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.

Reading Studies

Table 2 describes the characteristics of the 35 studies in the synthesis and indicates whether studies examined student achievement in reading, mathematics, or both. The publication years of the 30 reading studies range from 1986 to 2003; 12 of the studies were published in 2000 or later. Fourteen studies examined OST programs implemented during summer school, 15 examined after-school programs, and there was 1 study of a program that combined summer school and Saturday school. Twenty-one studies looked at OST programs that emphasized academics, and 9 looked at programs that focused on both academic and social skills. The latter group of programs included recreational, cultural, or vocational components in addition to their emphasis on academic and social skills. The program implementers were teachers in 20 of the reading studies, paid college students in 6 of the studies, and a mix of paid and volunteer adults in 3 of the studies. In the remaining study, the implementers were peer tutors (Mooney, 1986). The majority of the studies (23) reported aggregated reading scores from standardized assessments, including seven state tests. Seven studies employed other outcome measures, including teacher grade, end-of-grade tests, and researcher-developed assessments. Nine of the studies randomly assigned students to treatment and control groups. One study matched groups with a pretest, 14 studies matched groups using other criteria such as demographics, and 6 studies did not report

TABLE 2
Studies of out-of-school-time programs

Study	Treatment sample size[a]	Student grade level[b]	Student at-risk description	Out-of-school-time program description	Meta-analysis content area and measure
Baker & Witt (1996)	302	3rd–6th	Low SES, minority	After school; academically oriented activities in the context of recreational experiences; teacher-directed, large- and small-group instruction; activities to promote cultural awareness and positive self-esteem and attitude	Reading and mathematics Texas Assessment of Academic Skills
Bergin, Hudson, Chryst, & Resetar (1992)	10	K–3rd	Low SES, low English proficiency	After school; phonics-based, direct instruction with child-centered, culturally sensitive teaching methods and materials	Reading Metropolitan Readiness Test Metropolitan Achievement Test
Borman, Rachuba, Fairchild, & Kaplan (2002)	438	K–1st	Low SES	Summer school; integrated read-aloud and math activities; recreation, art, foreign language; small class size	Reading Comprehensive Test of Basic Skills
Branch, Milliner, & Bumbaugh (1986)	752	8th–9th	Low SES	Summer school; Summer Training and Education Program (STEP), existing federal work program combined with dropout prevention strategies	Reading and mathematics Metropolitan Achievement Test
Cosden, Morrison, Albanese, & Macias (2001)	90	4th–6th	Low performing	After school; homework time and support	Reading and mathematics Teacher grades

D'Agostino & Hiestand (1995)	1,006	4th	Low performing, low SES	Summer school; academic focus emphasizing higher-order thinking, questioning, and problem-solving skills	Reading and mathematics Iowa Test of Basic Skills
Finch (1997)	35	7th	Low SES	Before and after school; computer-assisted instruction designed to supplement students' mathematics curriculum	Mathematics Michigan Educational Assessment Program
Foley & Eddins (2001)	1,978	4th–5th	Low SES, minority	After school; literacy-based activities; YMCA program addressing socio-emotional behaviors and core values	Reading New York State English Language Arts Assessment
Gentilcore (2002)	114	8th	Low performing	After school; preparation to pass state assessment; workbook practice in reading and writing	Reading New York State English Language Arts Assessment
Harlow & Baenen (2001)	86	8th	Low SES	Summer and Saturday school; intensive enrichment program stressing academic excellence, leadership, creativity, and diversity; small classes	Reading and mathematics District end-of-grade tests
Hausner (2000)	128	K	Low performing, low SES	After school; scaffolded instruction, shared and guided reading, independent learning; teacher-directed, small- and large-group instruction	Reading Unpublished observation survey

(continued)

TABLE 2 (*Continued*)

Study	Treatment sample size[a]	Student grade level[b]	Student at-risk description	Out-of-school-time program description	Meta-analysis content area and measure
Hink (1986)	48	1st–9th	Low performing	Summer school; teacher-directed, remedial, large-group instruction.	Reading and mathematics Comprehensive Test of Basic Skills
Howes (1989)	22	1st	Low performing, low SES	Summer school; remedial instruction to groups of 10–15 students; focus on developing phonics, comprehension and writing skills	Reading Gates-McGintie Reading Test
Kociemba (1995)	192	2nd and 5th	Low performing	Summer school; compensatory academic programming in preparation for re-take of state reading and mathematics tests	Reading and mathematics California Achievement Test
LeBoff (1995)	40	3rd	Low performing	Summer school; remedial Title 1 program for urban children	Reading and mathematics Random House Achievement Program in Comprehension Researcher-developed mathematics test
Legro (1990)	49	1st–2nd	Low SES	After school; one-on-one homework tutoring; emphasis on social and communication skills; parent involvement; partnership program	Reading and mathematics Teacher grades
Leslie (1998)	73	6th–8th	Low performing	After school; one-on-one tutoring; homework support; computer-assisted instruction; student incentives	Reading and mathematics Iowa Test of Basic Skills

Levinson & Taira (2002)	1,289	3rd and 5th	Low performing	Summer school; homework support; computer-assisted instruction; teacher-directed large-group instruction; leveled trade books; word study, reading, vocabulary, writing	Reading Iowa Test of Basic Skills
Luftig (2003)	34	K	Low performing	Summer school; small-group tutoring; phonics instruction aligned with district curriculum	Reading Summer Success Reading Test
McKinney (1995)	47	1st–2nd	Low performing, minority	After school; one-on-one tutoring; self-concept and nonacademic enrichment	Reading and mathematics Stanford Achievement Test
McMillan & Snyder (2002)	90	9th	Low performing	Summer school; remedial program designed to help students pass the state assessment	Mathematics Virginia Standards of Learning Test
Mooney (1986)	15	4th	Low performing	After school; tutoring of 4th grade students by 8th grade students on understanding and completing reading homework	Reading Gates-McGintie Reading Test
Morris, Shaw, & Perney (1990)	30	2nd–3rd	Low performing	After school; one-on-one tutoring; shared reading, word study, story writing; basal sets and trade books	Reading Word recognition, basal passages, spelling
Prenovost (2001)	271	6th–8th	Low performing	After school; homework support; enrichment with art, life skills, field trips, sports	Reading and mathematics Stanford Achievement Test

(continued)

TABLE 2 (*Continued*)

Study	Treatment sample size[a]	Student grade level[b]	Student at-risk description	Out-of-school-time program description	Meta-analysis content area and measure
Raivetz & Bousquet (1987)	141	9th	Low performing	Summer school; one-on-one tutoring; teacher-directed large-group instruction	Reading and mathematics New Jersey High School Proficiency Test
Reed (2001)	30	1st	Low performing	Summer school; individualized instruction addressing language development, phonics, and reading fluency	Reading McGraw-Hill Terra Nova
Rembert, Calvert, & Watson (1986)	87	10th–12th	Low SES, minority	Summer school; remedial classroom instruction on a college campus, mentoring; computer-assisted instruction	Reading and mathematics Comprehensive Test of Basic Skills
Riley (1997)	78	9th–12th	Low SES	Summer school; remedial program on a college campus	Mathematics Researcher-developed mathematics test
Ross, Lewis, Smith, & Sterbin (1996)	328	2nd–4th	Low performing	After school; small-group tutoring; cooperative learning; teacher-directed instruction; focus on reading, writing, and computer skills	Reading Tennessee Comprehensive Assessment Program
Schacter (2001)	21	1st	Low performing	Summer school; one-on-one tutoring; teacher-directed instruction; word study, phonics instruction, reading, journal writing; computer-assisted instruction; activities to promote social and emotional growth	Reading Gates-McGintie Reading Test

Smeallie (1997)	31	6th–8th	Low performing	After school; tutoring; homework assistance; teacher-directed instruction on study skills; counseling; parent seminars on homework issues	Reading and mathematics Teacher grades
Ward (1989)	385	3rd and 6th	Low performing	Summer school; teacher-directed instruction; no basal texts, hands-on activities	Reading and mathematics California Achievement Test
Weber (1996)	29	3rd–6th	Low performing	Summer school; remedial program in rural location	Mathematics Riverside Tests of Achievement and Proficiency
Welsh, Russell, Williams, Reisner, & White (2002)	3,780	K–8th	Low SES	After school; large-scale urban program designed by After-School Corporation	Reading and mathematics New York City standardized tests
Zia, Larson, & Mostow (1999)	1,863	3rd–5th	Low performing	Summer school; remedial program designed to build student confidence in mathematics	Mathematics Instructional Systems in Mathematics

Note. SES = socioeconomic status.

[a]The sample size for the meta-analysis could be smaller based on the data available to calculate effect sizes.

[b]The grade levels of student data in this table are those that were reported in the study. The grade levels used to calculate effect sizes in Tables 3, 4, 5, and 6 may be different because they reflect the available data.

any matching. We computed effect sizes based on 16 studies that reported gain scores or pretest-posttest difference scores and 14 studies that reported only posttest scores. All of the studies concerned at-risk students, although each study defined "at-risk" on the basis of different characteristics. As Table 2 indicates, in the majority of both reading and mathematics studies, the defining characteristic was either low performance or low SES. The grade level of the students in the reading studies ranged from kindergarten to 12th grade, with the majority of studies focused on the elementary grades (K–5). The duration of OST programs that were described in the reading studies ranged from 4 weeks to the entire school year over a period of 1, 2, or 3 consecutive years. The total number of hours offered by each program ranged from 9 to 525 hours with a median of 84 hours. (There were 4 reading studies for which program duration could not be measured.)

With regard to research quality, 3 studies (Branch, Milliner, & Bumbaugh, 1986; Schacter, 2001; Borman, Rachuba, Fairchild, & Kaplan, 2002) received "high" ratings of 24, 25, and 28 points, respectively, on the quality scale. These studies 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 20 studies rated as "medium" (15–22 quality points, mean = 18) addressed most of the quality indicators, but with less sufficiency or clarity. The 7 studies with a "low" rating (9–14 quality points, mean = 12) 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, 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 intervention effectiveness within subgroups. The mean quality rating for all the reading studies was 17 points or "medium."

Meta-Analysis of Reading Studies

We calculated effect sizes for each of 42 independent samples yielded from 30 studies. Table 3 presents information on each independent sample, including the number of treatment students (those who received the OST program); defining characteristics of the independent sample, such as grade level or gender; the effect size for the study; and the lower and upper limits of the 95% confidence interval for the effect size. For the reading studies, Figure 1 provides a graphic display of each of the 42 effect sizes and their confidence intervals as reported in Table 3.

The display in Figure 1 indicates an overall tendency toward positive effects of OST programs for improving the reading achievement of at-risk students. The average effect size overall based on a fixed-effects model is .05, and the average effect size overall based on a random-effects model is .13. The 95% confidence intervals around these effect sizes do not include zero, which supports the conclusion that the OST programs that were examined through this meta-analysis had a significantly positive effect on the reading achievement of at-risk students ($p < .05$). The homogeneity analysis for the reading studies resulted in a Q value of 127.49 ($p < .0001$), indicating statistically significant variation among the effect sizes.

TABLE 3
Effect sizes and 95% confidence intervals for studies of out-of-school-time programs for improving reading achievement

Citation	Treatment *n*	Grade level	Effect size	95% Confidence interval	
				Lower bound	Upper bound
Baker & Witt (1996)	236	3rd–6th	.30	.02	.58
Bergin et al. (1992)	10	K–3rd	.34	−.56	1.24
Borman et al. (2002)	145	K–1st (2000 Cohort)	−.03	−.28	.22
Borman et al. (2002)	293	K–1st (1999 Cohort)	.07	−.13	.27
Branch et al. (1986)	752	8th–9th	.11	.01	.21
Cosden et al. (2001)	12	4th–6th	.38	−.37	1.14
D'Agostino & Hiestand (1995)	1,006	4th	−.14	−.25	−.03
Foley & Eddins (2001)	376	4th	−.03	−.14	.07
Foley & Eddins (2001)	255	5th	−.04	−.16	.08
Gentilcore (2002)	114	8th	.00	−.35	.35
Harlow & Baenen (2001)	65	8th	.17	−.19	.53
Hausner (2000)	128	K	.43	.19	.67
Hink (1986)	38	1st–9th	.40	−.06	.86
Howes (1989)	12	1st (Treatment A)	.02	−.69	.72
Howes (1989)	10	1st (Treatment B)	.02	−.88	.91
Kociemba (1995)	79	5th	.04	−.25	.32
Kociemba (1995)	113	2nd	.71	.37	1.05
LeBoff (1995)	20	3rd (male)	−.03	−.67	.61
LeBoff (1995)	19	3rd (female)	.42	−.25	1.08
Legro (1990)	19	2nd	.06	−.59	.71
Legro (1990)	30	1st	.92	.23	1.60
Leslie (1998)	11	6th	.90	−.07	1.87
Leslie (1998)	18	8th	.88	.03	1.73
Leslie (1998)	10	7th	2.35	.75	3.95
Levinson & Taira (2002)	76	5th	−.12	−.42	.18
Levinson & Taira (2002)	71	3rd	−.03	−.33	.27
Luftig (2003)	16	K	1.28	.53	2.04
McKinney (1995)	20	1st–2nd	.09	−.52	.70
Mooney (1986)	15	4th	.67	−.10	1.44
Morris et al. (1990)	30	2nd–3rd	.50	−.02	1.03
Prenovost (2001)	155	6th–8th (male)	.02	−.16	.20
Prenovost (2001)	116	6th–8th (female)	.12	−.09	.34
Raivetz & Bousquet (1987)	141	9th	.21	.03	.39
Reed (2001)	17	1st (male)	−.17	−.80	.46
Reed (2001)	13	1st (female)	.26	−.49	1.01
Rembert et al. (1986)	87	10th–12th	.38	.04	.73
Ross et al. (1996)	117	2nd–4th	.18	−.08	.44
Schacter (2001)	21	1st	.73	.14	1.32
Smeallie (1997)	31	6th–8th	−.76	−1.29	−.23
Ward (1989)	73	6th	−.36	−.67	−.05
Ward (1989)	136	3rd	−.17	−.40	.05
Welsh et al. (2002)	146	K–8th	−.02	−.23	.19

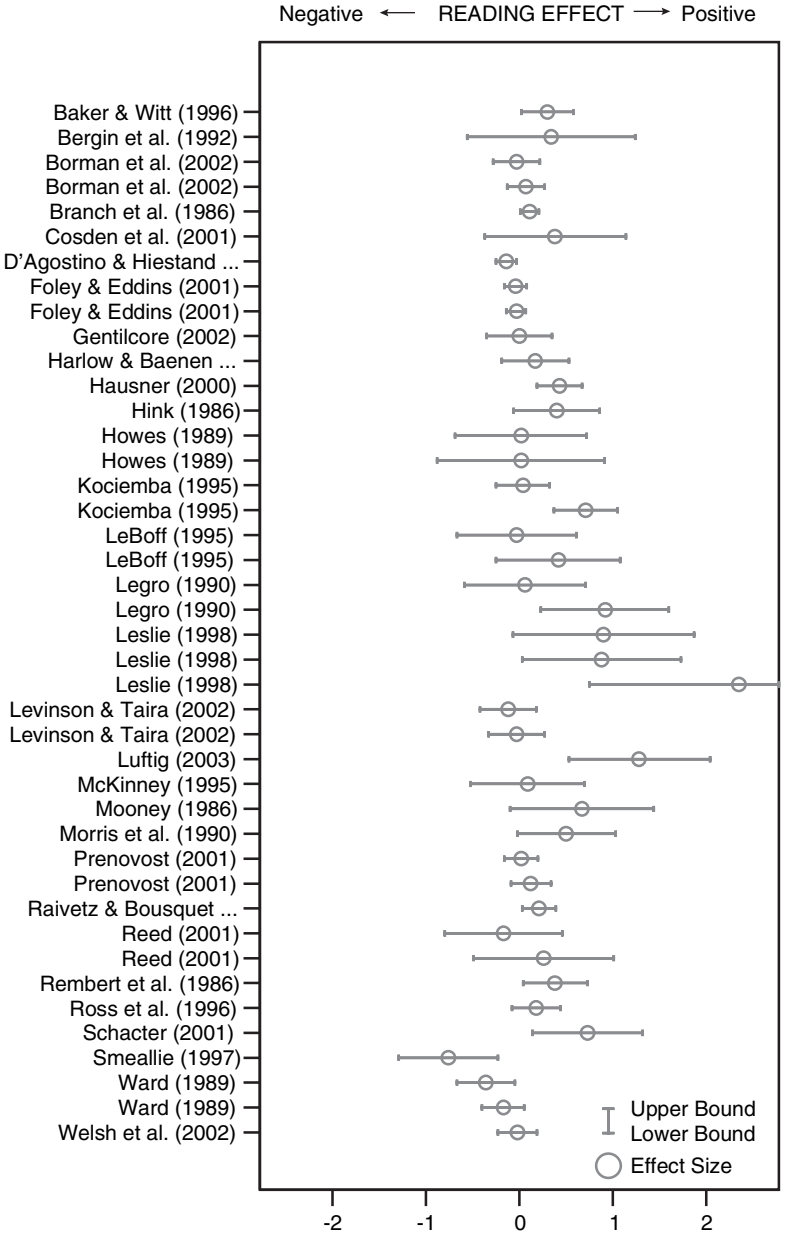

FIGURE 1. *Distribution of effect sizes from studies of out-of-school-time programs for improving reading achievement. This figure graphically displays the effect sizes and corresponding confidence intervals reported in Table 3. As reported in Table 3, the upper bound for the Leslie (1998) effect size that extends off the scale in this figure is 3.95.*

Moderator Analyses of Reading Studies

Table 4 presents the mean effect sizes for each level of the different moderators, including the program characteristics of timeframe, grade level, activity focus, duration, and student grouping, and the study characteristics of quality, publication type, and score type. In Table 4, when the 95% confidence interval does not include zero, the effect size of the moderator is significantly different from zero. Table 4 also includes *Q* values for homogeneity analyses among the effect sizes for each moderator. A statistically significant *Q* value indicates that the moderator influences the variation among the effect sizes for the studies.

As indicated in Table 4, the effect sizes of after-school and summer school programs were significantly greater than zero (.07 and .05, respectively), but based on the *Q* statistic, the effect sizes of OST programs in reading were not significantly influenced by timeframe. For the analysis of student grade level, two studies were excluded because of overlapping grade levels (Hink, 1986; Welsh et al., 2002). The effect sizes for lower elementary, middle, and high school students were significantly greater than zero (.22, .09, and .25, respectively). The homogeneity analysis for grade level yielded a statistically significant *Q* value indicating that grade level accounted for some of the variation among the effect sizes. Only the OST programs that had a primarily academic focus had an effect size that was significantly different from zero (.07); and, based on the *Q* value, focus of activities in OST programs did not significantly influence effect sizes. The durations of the OST reading programs were divided into quartiles for the moderator analysis, and the *Q* value for duration was statistically significant. The programs with durations of 44 to 84 hours and 85 to 210 hours had effect sizes that were significantly greater than zero (.28 and .15, respectively); those for programs longer than 210 hours and shorter than 44 hours were not significantly different from zero. Twenty-six studies reported a student grouping structure used by the OST program, and the *Q* value indicated significant variation among the effect sizes for this moderator. Working with students one-on-one had the largest effect size (.50), and a combination of student grouping structures had the next largest effect size (.15); both effect sizes were significantly greater than zero. Large-group structures and placing students in small groups of 10 or fewer had smaller and nonsignificant effect sizes.

For the moderator of study quality, the effect sizes of high- and medium-quality studies were significantly different from zero, with average effect sizes of .11 and .09, respectively. The *Q* value was statistically significant, indicating that effect sizes were influenced by study quality. As indicated in Table 3, only the studies published in peer-reviewed journals had an average effect size that was significantly greater than zero (.41), and the *Q* value for publication type was statistically significant. For the moderator of score type, only the effect size for posttest scores (.07) was statistically different from zero; but, based on the *Q* value, score type did not have a statistically significant influence on the effect sizes for reading.

To examine the reading studies for possible relationships among moderators, we constructed correlation matrices for program and study characteristics of the reading studies (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. Summer schools, more than after-school programs, focused primarily on academics. 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.

TABLE 4
Moderators of effect sizes for studies of out-of-school-time programs for improving reading achievement

Moderator	k[a, b]	Q	Effect size[c]	95% Confidence interval: Lower bound	95% Confidence interval: Upper bound
OST timeframe		.55			
After school	15		.07	.01	.14
Summer school	14		.05	.01	.11
Summer school and Saturday school	1		.17	−.19	.53
Grade level[d]		24.80**			
Lower elementary (K–2)	14		.22	.11	.33
Upper elementary (3–5)	15		−.03	−.08	.02
Middle school (6–8)	9		.09	.02	.17
High school (9–12)	2		.25	.09	.41
Focus		2.30			
Academic	21		.07	.01	.13
Academic + social	9		.05	−.01	.11
Duration		18.51**			
≤ 43 hrs	7		.18	−.02	.38
44–84 hrs	7		.28	.15	.41
85–210 hrs	8		.15	.08	.23
> 210 hrs	4		−.04	−.12	.05
Grouping		13.03**			
Large group (11 or more)	7		.07	−.03	.18
Small group (10 or fewer)	6		.01	−.07	.10
One-on-one tutoring	5		.50	.21	.79
Mixed	8		.15	.06	.23
Study quality		6.35*			
High	3		.11	.02	.20
Medium	20		.09	.02	.15
Low	7		−.03	−.11	.05
Publication type		7.86*			
Conference paper/ technical report	15		.04	−.01	.09
Dissertation	11		.08	−.05	.20
Peer-reviewed journal	4		.41	.16	.66
Score type		.12			
Gain score	16		.05	−.01	.11
Posttest score	14		.07	.01	.13

[a]Number of effect sizes included in the analysis.
[b]The unit of analysis (k) for grade level as a moderator is the independent sample, within-study effect sizes (one to three per study).
[c]Fixed-effects model.
[d]We coded grade levels of students by using four categories: lower elementary (K–2), upper elementary (3–4), middle school (6–8), and high school (9–12). When an independent sample overlapped two categories, we chose the category that included the majority of grade levels. For example, the Bergin et al. (1992) study included K–3, so it was categorized as K–2 rather than 3–4.
*$p < .05$. **$p < .01$.

Other Influences on Effective OST Programs for Reading

The 30 studies included in the reading meta-analysis examined OST programs that varied in their approaches to improving students' reading skills. In this section we describe features that program implementers highlighted as important aspects of effective OST programs, but which we were unable to include in the moderator analyses.

Baker and Witt (1996) evaluated two after-school programs in Texas and concluded that the programs had greater impact on the students who participated more often. The programs 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.

Most of the synthesis studies did not report the qualifications of those implementing the program beyond basic descriptions such as "teacher" or "college student." However, 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. Volunteer tutors implemented this OST program 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)

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). Schacter (2001) studied the impact of a systematic reading curriculum, with one-on-one tutoring and recreational activities, which was implemented at an 8-week summer day camp for promoting social and emotional growth. The purpose of the program, which was designed for economically disadvantaged children, was to turn first graders' summer reading losses into gains. The author identified the summer camp context as instrumental to the success of the program.

A well-defined reading curriculum is another feature emphasized by implementers of effective OST programs. The structure of the curriculum in Hausner's (2000) study of Project Accelerated Literacy (PAL) included eight components of literacy instruction based on a constructivist approach and scaffolded learning: reading aloud to children, shared reading, guided reading, independent reading, modeled writing, shared writing, guided writing, and independent writing. Features of the PAL program included (a) small class size; (b) a variety of learning centers that integrated literacy tools and tasks (e.g., play office, art center, cooking, and book corner); (c) a 2-hour block of time for literacy instruction through large-group, small-group, and individual instruction; (d) teaching practices based on each student's performance on standards; (e) scaffolded teaching that followed a pattern of modeling, guiding, observing, and practicing skills for students; and (f) a thematic curriculum (e.g., foods,

sea life, and community helpers) reflected in each activity center. As a result of this 30-week after-school program, at-risk kindergarten participants experienced gains in literacy learning as compared with their peers in the control group.

There was also evidence for the importance of a well-defined curriculum in summer schools. Rembert, Calvert, and Watson (1986) evaluated a summer school for at-risk students in Grades 10 through 12. The program provided college preparation classes that focused on skill mastery in basic academics and simulated college instruction. In comparison with the control group, participants in the summer school scored significantly higher on the reading portion of the Comprehensive Test of Basic Skills.

Mathematics Studies

Table 2 describes the characteristics of the 22 studies in the synthesis that examined the influence of OST programs on mathematics achievement. The studies were published from 1986 to 2002, with 5 published in 2000 or later. Twelve studies examined OST programs implemented during summer school, 8 examined after-school programs, there was 1 study of a before-and-after-school program, and 1 study of a combined summer school and Saturday school program. Seventeen of the OST programs studied focused primarily on academics, and 5 OST programs combined academics with other emphases such as recreation, arts programming, life skills, and mentoring. The program implementers were teachers in 16 of the mathematics studies, paid college students in 4, and the 2 remaining studies provided information on implementers.

The majority of the studies (17) reported aggregated mathematics scores from standardized assessments, including four state tests. Five studies employed other outcome measures, including teacher grades, end-of-grade tests, and researcher developed assessments. Seven of the mathematics studies randomly assigned students to treatment and control groups, 10 studies matched student groups using other criteria such as demographics, and 5 studies did not report any matching of groups. For 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. The grade level of the students in the mathematics studies ranged from kindergarten to 12th grade, with 14 studies addressing elementary grades (K–5), and 16 studies addressing secondary grades (6–12). The OST mathematics programs ranged in total time from a 6-week after-school program that had 12 hours duration to 525 hours in a longitudinal study of an after-school intervention. The median program duration was 82 hours. (There were 7 mathematics studies for which program duration could not be coded.)

Regarding research quality, 1 study (Branch et al., 1986) was rated as "high" (24 quality points), 12 as "medium" (15–20 quality points, mean = 17), and 9 as "low" (9–14 quality points, mean = 12). The reasons for these ratings were similar to those given for the ratings of the reading studies. The mean quality rating for all the mathematics studies was 16 points or "medium."

Meta-Analysis of Mathematics Studies

Table 5 shows the effect sizes and the confidence intervals of the effect sizes for mathematics OST studies, along with the sizes and characteristics of the treatment samples. All of these 33 effect sizes and their corresponding confidence intervals

TABLE 5
Effect sizes and 95% confidence intervals for studies of out-of-school-time programs for improving mathematics achievement

Citation	Treatment *n*	Grade level	Effect size	95% Confidence interval	
				Lower bound	Upper bound
Baker & Witt (1996)	236	3rd–6th	.307	.027	.587
Branch et al. (1986)	752	8th–9th	.227	.126	.329
Cosden et al. (2001)	12	4th	.837	.058	1.617
D'Agostino & Hiestand (1995)	1,006	4th	−.156	−.264	−.048
Finch (1997)	12	7th (male)	.375	−.395	1.146
Finch (1997)	23	7th (female)	−.008	−.656	.639
Harlow & Baenen (2001)	64	8th	.162	−.201	.520
Hink (1986)	28	1st–9th	−.028	−.564	.508
Kociemba (1995)	42	5th	.391	.036	.746
Kociemba (1995)	79	2nd	.078	−.206	.363
LeBoff (1995)	20	3rd (male)	.379	−.268	1.025
LeBoff (1995)	19	3rd (female)	.736	.053	1.418
Legro (1990)	19	2nd	.366	−.289	1.022
Legro (1990)	30	1st	.515	−.148	1.179
Leslie (1998)	11	6th	.185	−.415	.786
Leslie (1998)	10	7th	.346	−.848	1.540
Leslie (1998)	18	8th	.621	−.241	1.482
McKinney (1995)	23	1st–2nd	−.138	−.726	.451
McMillan & Snyder (2002)	90	9th	1.331	.818	1.844
Prenovost (2001)	155	6th–8th (male)	.005	−.208	.218
Prenovost (2001)	116	6th–8th (female)	.081	−.188	.351
Ravietz & Bousquet (1987)	136	9th	.219	.034	.404
Rembert et al. (1986)	87	10th–12th	.340	−.003	.683
Riley (1997)	23	9th–12th (male)	.827	.290	1.364
Riley (1997)	55	9th–12th (female)	.990	.535	1.446
Smeallie (1997)	31	6th–8th	−.102	−.610	.407
Ward (1989)	67	6th	−.055	−.374	.265
Ward (1989)	108	3rd	−.101	−.344	.143
Weber (1996)	29	3rd–6th	−.316	−.768	.136
Welsh et al. (2002)	183	K–8th	.240	.041	.438
Zia et al. (1999)	947	4th	.074	.007	.141
Zia et al. (1999)	916	3rd	.061	−.007	.129
Zia et al. (1999)	809	5th	.061	−.011	.133

are displayed graphically in Figure 2. The graph in Figure 2 illustrates a tendency toward positive effects of OST programs for improving the mathematics achievement of at-risk students. The average overall effect size based on a fixed-effects model was .09, and the average overall effect size based on a random-effects model was .17, both statistically greater than zero. The homogeneity analysis resulted in a Q value of 102.72, which was statistically significant ($p < .0001$) and indicated variation among the effect sizes greater than expected by sampling error alone.

Moderator Analyses of Mathematics Studies

Table 6 presents mean effect sizes for the program and study moderator variables in the mathematics studies. The effect sizes of both after-school programs and summer schools were significantly greater than zero (.16 and .09, respectively). However, the Q value was not statistically significant, indicating that the overall effect size of OST programs for mathematics was not influenced by timeframe. For the analysis of student grade level, 2 studies were excluded because of overlapping grade levels (Hink, 1986; Welsh et al., 2002). Among the remaining studies, the largest effect size was observed for high school students (.44), followed by the effect size for middle school students (.16) and that for upper elementary students (.05), all significantly greater than zero. The 3 studies of lower elementary grades had an average effect size that was not significantly different from zero. The statistically significant Q value indicated that grade level accounted for some of the variance in the overall effect size. Regarding activity focus in the OST programs, the effect sizes for studies with primarily academic or combined academic and social focuses were .07 and .19, respectively, and both were significantly greater than zero. The Q value indicated a statistically significant influence of program focus on effect sizes. The durations of the OST mathematics programs were divided into quartiles, resulting in a slightly different distribution than that for the reading studies. Programs with a duration of 46–75 hours had the largest effect size (.23), followed by those with 76–100 hours (.22) and those with more than 100 hours (.16). Only the effect sizes of the programs with durations of more than 45 hours were significantly greater than zero. There was statistically significant variation among different program durations based on the Q value. There were 17 mathematics studies that reported a student grouping structure used by the OST program, and the Q value indicated significant variation among the effect sizes for this moderator. The effect size for mixed student grouping structures (.25) was positive and significantly different from zero, as were the effect sizes for small and large student groupings (.18 and .08, respectively). The effect sizes for studies that involved tutoring in mathematics were not significantly different from zero.

In the moderator analysis of study quality, the one mathematics study that was coded as high in research quality produced the largest effect size (.23), followed by the effect sizes for medium-quality studies (.10); both were statistically greater than zero. The effect size for low-quality studies did not differ from zero, and the Q value indicated that study quality was a statistically significant moderator of effect size. The effect sizes for conference papers/technical reports, dissertations, and peer-reviewed journal articles were all significantly greater than zero (.11, .13, and .08, respectively), but the Q value for publication type was not statistically significant. For the moderator of score type, only the mean effect size for gain scores was significantly greater than zero (.13), and the Q value indicated that score type did not have a statistical influence on the effect sizes.

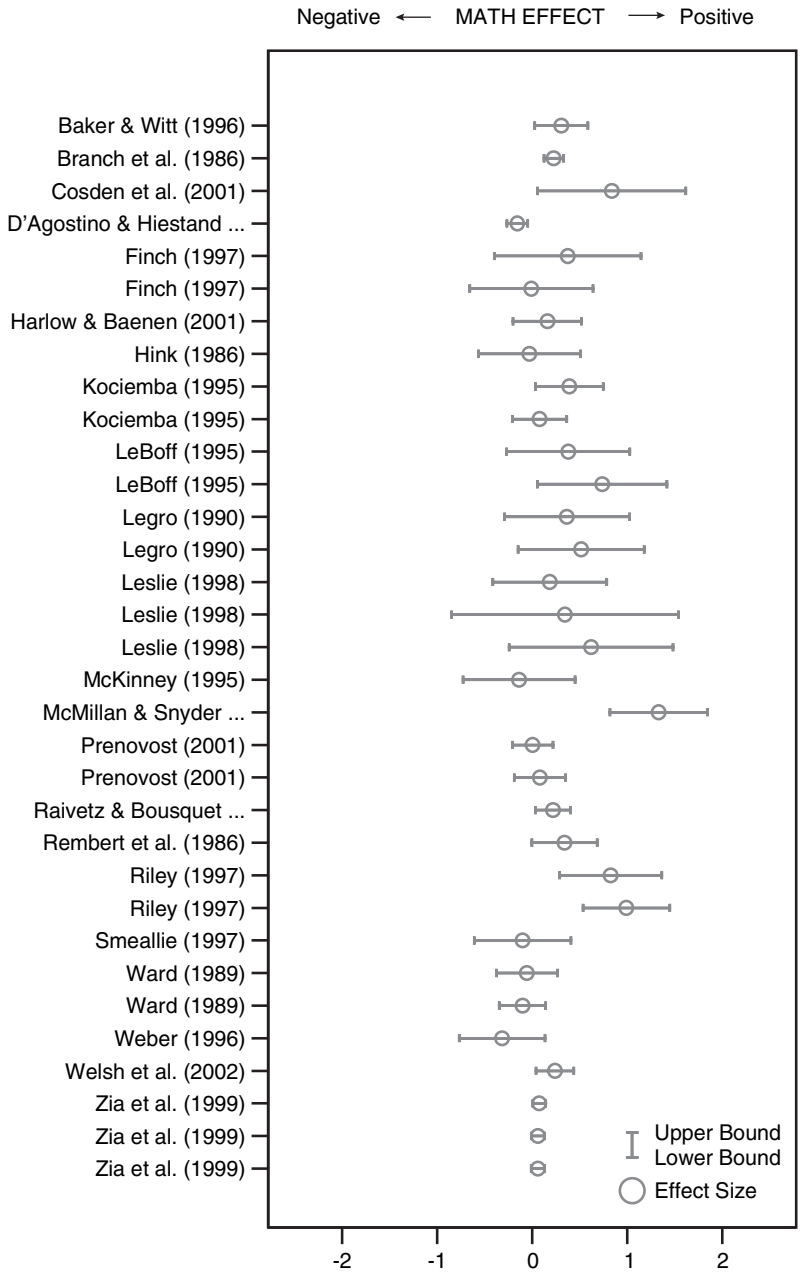

FIGURE 2. *Distribution of effect sizes from studies of out-of-school-time programs for improving mathematics achievement. This figure graphically displays the effect sizes and corresponding confidence intervals reported in Table 5.*

TABLE 6
Moderators of effect sizes for studies of out-of-school time programs for improving mathematics achievement

Moderator	k[a, b]	Q	Effect size[c]	95% Confidence interval: Lower bound	95% Confidence interval: Upper bound
OST timeframe		1.42			
After school	9		.16	.05	.27
Summer school	12		.09	.04	.14
Summer school and Saturday school	1		.16	−.20	.52
Grade level[d]		33.29**			
Lower elementary (K–2)	3		.22	−.14	.58
Upper elementary (3–5)	11		.05	.01	.08
Middle school (6–8)	11		.16	.08	.23
High school (9–12)	5		.44	.30	.59
Focus		6.04*			
Academic	17		.07	.02	.12
Academic + social	5		.19	.11	.28
Duration		9.93*			
≤ 45 hrs	4		.06	−.01	.13
46–75 hrs	4		.23	.09	.38
76–100 hrs	4		.22	.13	.32
> 100 hrs	3		.16	.04	.29
Grouping		21.21**			
Large group (11 or more)	5		.08	.02	.15
Small group (10 or fewer)	3		.18	.02	.35
One-on-one tutoring	3		.22	−.12	.56
Mixed	6		.25	.16	.34
Study quality		10.26**			
High	1		.23	.13	.33
Medium	12		.10	.05	.16
Low	9		.02	−.06	.10
Publication type		.83			
Conference paper/ technical report	8		.11	.05	.17
Dissertation	11		.08	.01	.25
Peer-reviewed journal	3		.08	.01	.14
Score type		3.63			
Gain score	10		.13	.08	.18
Posttest score	12		.04	−.03	.12

[a]Number of effect sizes included in the analysis.
[b]The unit of analysis (k) for grade level as a moderator is the independent sample, within-study effect sizes (one to three per study).
[c]Fixed-effects model.
[d]We coded grade levels of students by using four categories: lower elementary (K–2), upper elementary (3–4), middle school (6–8), and high school (9–12). When an independent sample overlapped two categories, we chose the category that included the majority of grade levels. For example, the Bergin et al. (1992) study included K–3, so it was categorized as K–2 rather than 3–4.
*$p < .05$. **$p < .01$.

We constructed correlation matrices for program and study characteristics of the mathematics studies. Studies of students in Grades 3–12 reported primarily programs that had an academic focus, while the two studies of students in Grades K–2 reported programs that had both an academic and a social focus. Studies of programs with shorter durations (under 75 hours) focused solely on academic goals, while studies of programs with longer durations reported both academic and social goals. More summer schools than after-school programs focused primarily on academics, and the programs that were using large group instruction focused only on academics. Regarding study characteristics, most of the studies that were rated as low quality reported only posttest scores; 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.)

Other Influences on Effective OST Programs for Mathematics

The 22 studies included in the mathematics meta-analysis described a wide variety of programs. In this section we describe some features of effective OST programs for mathematics that we were unable to include in the moderator analyses.

A feature of effective programs serving high school participants was the combination of academic instruction with career or college skills. A summer school program studied by Riley (1997) brought high school students to a college campus. The students were taught high school mathematics in large classes and were provided with individual and small-group tutoring. In comparison with a matched group of students from low-SES families, there were positive effects on mathematics achievement for both male and female program participants. The program studied by Rembert et al. (1986) was a residential summer school camp designed to introduce at-risk students to a collegiate atmosphere including access to academic classes, laboratories, computers, and recreational facilities. The study demonstrated positive effects on both mathematics and reading achievement.

Counseling is another component found in some of the studies with positive effects on mathematics achievement. The Summer Training and Educational Program (STEP) was designed to promote high school graduation and successful transition to careers with a federal summer jobs program (Branch et al., 1986). The students were exposed to academic classes and life and career counseling, resulting in positive effects on the mathematics and reading achievement of treatment participants.

Discussion

For both reading and mathematics outcomes and fixed- and random-effects models, the overall effect sizes were statistically greater than zero. In answer to our first research question, based on rigorous research studies (as defined by the use of control or comparison groups), OST programs can have positive effects on the achievement of at-risk students.

Interpretation of Effect Sizes

Several factors influence the interpretation of the overall effect sizes. OST programs supplement the regular school day, so the interpretation of effect sizes for typical education interventions might not apply (see, e.g., Cohen's [1988] description of an effect size of .20 as small). As Cooper et al. (2000) discuss, previous meta-analyses suggest that effect sizes of .10 to .20 are not trivial (Lipsey & Wilson, 1993)

and may be typical for remedial programs. For example, Borman and D'Agostino (1996) reported an average effect size of .11 for year-long Title 1 programs. 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. However, Cooper et al. included studies that used one-group pretest and posttest designs that they cited as possibly inflating the effect size estimates. The effect size they computed for studies that used random assignment was .14 for both fixed-effects and random-effects models. These results are more consistent with our findings using a random-effects model for studies of OST programs, all of which included control or comparison groups.

Second, the students who participated in OST programs in the studies 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 OST programs. Granted, the comparisons are of at-risk treatment students with at-risk control students, which means that the effects are positive in relation to at-risk students who do not participate in an OST intervention. This implies that OST programs are unlikely to close the achievement gap between at-risk and more advantaged students. Nonetheless, our results suggest that at-risk students who participate in OST programs improve learning outcomes more than at-risk students who do not participate.

Third, some moderators of the OST programs that are reviewed in this synthesis resulted in larger positive effects on student achievement as compared with the overall effect sizes. Results from moderator analyses help policymakers and those who fund OST programs to identify intervention characteristics that can result in larger positive influences on the achievement of at-risk students.

Moderator Influences

In answer to our second research question, the effectiveness of OST programs differs by various program and study characteristics, depending on whether the student outcome is reading or mathematics achievement. The timeframe for delivery of OST programs was not a statistically significant moderator for reading or mathematics studies. Whether OST programs were implemented in an after-school setting or during summer school did not influence the impact of OST on student achievement. However, one of the strongest effects in the synthesis was the positive impact of tutoring on reading, and this type of instruction occurred only in the studies of after-school programs. In general, the studies of summer schools reported more use of large groups for instruction than did the studies of after-school programs.

Grade level was a statistically significant moderator of effect size for both reading and mathematics outcomes. For reading, significant positive effect sizes were highest in the lower elementary and high school grades; for mathematics, significant positive effect sizes were highest for students in middle and high school. The results for reading confirm the importance of early-grades intervention for students who are underachieving in reading, but these results also suggest that reading/language arts improvement is possible for high school students in OST programs (e.g., Branch et al., 1986). 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. There were fewer

mathematics studies than reading studies that involved elementary students; therefore, additional research is needed on the influence of OST in relation to grade level and content area. In addition, the lack of positive findings in reading for upper elementary students needs further investigation to determine whether elements of OST programs contribute to this result.

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, programs that were both academic and social had a higher mean effect size than those that were mainly academic. This suggests that OST programs 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 to motivate students to attend, particularly in the upper grades (Miller, 2003; De Kanter, 2001; Huang, Gribbons, Kim, Lee, & Baker, 2000). However, it should be noted that the four OST mathematics programs with fewer than 45 hours had an academic focus. It is possible that the short duration and not the academic focus led to fewer positive findings for these programs in comparison with programs that were both academic and social.

Only 5 of the 35 studies in this synthesis described student attendance in OST programs, so it was not feasible to analyze the moderating influences of student attendance on effect sizes. Instead we analyzed program duration, which measures the potential for student exposure to OST programming. For both reading and mathematics, statistically significant effect sizes were larger for OST programs that were more than 45 hours in duration, but the reading programs with the longest durations (more than 210 hours) had an average effect size that was not significantly different from zero. For mathematics, the programs longest in duration (more than 100 hours) had a slightly smaller mean effect size than programs with medium duration. Although the data are not available to confirm this, it is probably more difficult for longer programs than for shorter programs to keep students motivated and attending on a regular basis. Some programs reviewed in this synthesis focused on improving student engagement in learning in hopes that their attendance in the OST programs would increase. Baker and Witt (1996) evaluated an after-school program that engaged students by conducting academic activities in the context of a recreational experience. Other programs used incentives for attending and participating in OST programs, such as paid wages (Branch et al., 1986) and a token-based economy (Leslie, 1998).

Three of the studies in the synthesis linked attendance to student achievement. Prenovost (2001) reported the mean number of days that students attended summer school and analyzed for high- and low-dosage participants in comparison with nonparticipant student matches. High-dosage students generally improved more than low-dosage students in relation to nonparticipating students. Welsh et al. (2002) found larger effects sizes for students who were "highly active" (attended 60 days each year) in 2 years of after-school programming as compared with those who were highly active for 1 year. In a longitudinal study of a summer school, Borman et al. (2002) reported that the effect sizes for 1 year of attendance were small but increased for students who attended 3 years. One problem with interpreting a link between attendance and achievement is that attendance at OST programs is voluntary and not mandated. If the students with higher attendance are more motivated

academically than those who drop out, program effects might be due to higher student motivation more than to the OST intervention (Fashola, 1998). Complicating the issue is the fact that very few studies document the number of students who dropped out of OST programs and the reasons they dropped out. Clearly, the issue of student attendance in OST programs is complex. Ideally, evaluations of OST should document all the variables related to time and student exposure, including student attendance and dropout, program duration, and the distribution of programming over time (e.g., a 100-hour program distributed over 5 weeks).

Regarding student grouping, the largest average positive effect size in the synthesis occurred for the studies of reading that used one-on-one tutoring. This result confirms other research that demonstrates the positive influence of tutoring and individualized help for at-risk students in reading (Elbaum, Vaughn, Hughes, & Moody, 2000). For mathematics, OST programs using small group instruction or a combination of student grouping structures had higher effect sizes than programs with large group instruction and tutoring. This finding for mathematics aligns with positive effects on student achievement from small-group instruction in classrooms during the school day (Lou et al., 1996).

We examined three study characteristics for possible moderating influences on effect sizes. The moderator of study quality significantly influenced the effect sizes for both content areas. Only the high- and medium-quality studies had average effect sizes for reading and mathematics that were significantly greater than zero. These studies had the characteristics of scientifically based research as described in NCLB—that is, an experimental research design with random assignment or a quasi-experimental design with matching. The positive effects for the more rigorous studies in this synthesis lend support to the conclusion that OST can have positive influences on student achievement.

Type of publication was a statistically significant moderator of effectiveness of OST for reading achievement but not for mathematics. The average effect size for reading studies reported in peer-reviewed journals was larger than for presentations/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 those with nonsignificant 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 did not significantly influence the effect sizes. For reading outcomes, only the average effect size based on posttest scores was significantly greater than zero, while for mathematics outcomes only the average effect size for gain scores was significantly greater than zero.

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." A related issue is the lack of documentation for student attendance and dropout, which makes it difficult to describe the

treatment population. Another problem with research on OST programs 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 programs 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 on and evaluation of OST programs 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.

1. OST programs can have positive effects on the achievement of 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). Our results contrast with the first-year evaluation of the 21st Century Community Learning Centers (U.S. Department of Education, 2003), which found no statistically significant effects of after-school programs on reading or mathematics achievement for elementary or middle school students. However, the 21st Century 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.[1] 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 results indicate that program duration and student grouping influence program effectiveness. Aggregating results across programs that vary in these characteristics can mask positive outcomes.

2. Our findings suggest that the timeframes of OST programs do not influence their effectiveness. 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. The feasibility of implementing effective instruction in one-on-one or small-group settings also should be considered.

3. Students in both elementary and secondary grades can benefit from OST programs for improved reading; in contrast, there are indications that benefits for mathematics achievement occur primarily in the secondary grades. These findings are encouraging, but additional research is needed, given the greater difficulty in recruiting older students into OST programs (Grossman et al., 2001).

4. OST programs 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 content area.

5. Administrators of OST programs should monitor program implementation and student learning 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.

6. OST programs 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.

7. Research syntheses of OST programs should examine both published and unpublished research and evaluation reports. Estimates of the true effect of OST programs on student achievement will be inaccurate if only published studies are examined, because statistically nonsignificant 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.

8. Future research and evaluation studies should document the characteristics of OST programs 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 more evidence on the characteristics of effective OST programs.

Notes

Mid-continent Research for Education and Learning (McREL), where all of the authors were employed while working on this article, supported the work at least in part with federal funds from the U.S. Department of Education through regional educational laboratory contract No. ED-01-CO-0006. The content of the article does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. government. We wish to express our appreciation to Zoe Barley, at McREL, for her helpful advice and review comments, and to Rebecca Van Buhler, at McREL, for helping us to conduct literature searches and screen research articles. We also thank Pamela Blair, at McREL, for producing the final figures for this article.

[1]The evaluation of the 21st Century Community Learning Centers (U.S. Department of Education, 2003) 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. This evaluation and four other studies were excluded because data were aggregated across the sites, although the interventions differed from site to site.

References

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

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*

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)

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.

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: University of Wisconsin. 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)

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

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), 1–118.

*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)

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

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.

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. (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*

*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, Department of Evaluation and Research. (ERIC Document Reproduction Service No. ED 466 489)

*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. (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.

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

*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.

*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 and Information Studies.

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

Karweit, N. (1985). Should we lengthen the school term? *Educational Researcher, 14*(6), 9–15.

*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.

*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 annual meeting of the American Educational Research Association, New Orleans.

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks: Sage Publications.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458.

*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 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 annual meeting of the American Educational Research Association, New Orleans.

Miller, B. M. (2003). *Critical hours: Afterschool programs and educational success.* Quincy, MA: 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.

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

No Child Left Behind Act of 2001. Pub. L. No. 107-110, 115 Stat.1425 (2002).

Policy Studies Associates. (1995). *Extending the learning time for disadvantaged students: An idea book: Vol. 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.

*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.

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

*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*

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.

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

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.

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 Undersecretary.

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.

*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)

*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/papers/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)

Authors

PATRICIA A. LAUER is a former Principal Researcher at Mid-continent Research for Education and Learning. Currently, she is the Evaluation Director for the Rocky Mountain Center for Health Promotion and Education, 7525 West 10th Avenue, Lakewood, CO 80214; e-mail *patl@rmc.org.* Her research interests include preservice and inservice teacher education and methods for delivering them, including online learning.

MOTOKO AKIBA is a former Senior Researcher at Mid-continent Research for Education and Learning. Currently, she is an Assistant Professor of Educational Policy in the Department of Educational Leadership and Policy Analysis, University of Missouri, Columbia, 202 Hill Hall, Columbia, MO 65211; e-mail *akibam@missouri.edu.* Her research interests include multicultural teacher education, school safety, and comparative and international education.

STEPHANIE B. WILKERSON is a former Principal Evaluator at Mid-continent Research for Education and Learning. Currently, she is the President of Magnolia Consulting, LLC, a small, woman-owned research and evaluation group with headquarters at 3171 Courthouse Road, Louisa, VA 23093; e-mail *stephanie@magnoliaconsulting.org.* Her research and evaluation interests include curriculum efficacy, education and public outreach, organizational change, and fidelity of implementation.

HELEN S. APTHORP is a Principal Researcher at Mid-continent Research for Education and Learning, 4601 DTC Boulevard, Suite 500, Denver, CO 80237; e-mail *hapthorp@mcrel.org.* Her research interests are in classroom instruction and the academic success of economically disadvantaged students.

DAVID SNOW is a former Research Consultant for Mid-continent Research for Education and Learning. He currently works for Billings Public Schools and independently as an Education Consultant. He can be reached at *davesnowmt@hotmail.com.* His research interests are centered on improving teaching practice, with an emphasis on working with at-risk students.

MYA L. MARTIN-GLENN is a Research Associate at Mid-continent Research for Education and Learning, 4601 DTC Boulevard, Suite 500, Denver, CO 80237; e-mail *mmartin-glenn@mcrel.org.* Her research interests include school effectiveness, teacher education, and urban education.