Contract No.: ED-99-CO-0134
MPR Reference No.: 8658-400

MATHEMATICA
Policy Research, Inc.

A Broader View: The National Evaluation of the 21st Century Community Learning Centers Program

Design Report: Volume I

March 2, 2001

Mathematica Policy Research:
Mark Dynarski
Susanne James-Burdumy
Wendy Mansfield
Daniel Mayer
Mary Moore
John Mullens
Tim Silva

Decision Information Resources:
Carol Pistorino
Douglas Hermond

Submitted to:

U.S. Department of Education
Planning and Evaluation Service
400 Maryland Avenue, SW Suite 6W222
Washington, DC 20202

Project Officer:
Elizabeth Warner, Ph.D.

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

Project Director:
Mark Dynarski, Ph.D.

CONTENTS

CONTENTS (CONTINUED)

TABLES

FIGURES

A BROADER VIEW: THE NATIONAL EVALUATION OF THE 21st CENTURY COMMUNITY LEARNING CENTERS PROGRAM

DESIGN REPORT

Executive Summary

Most school-age children have working parents, yet most schools dismiss their students long before parent workdays end. During the intervening hours—"out-of-school time"—children can either develop their academic and personal skills in safe, supervised settings, or engage in unproductive behaviors in unsafe, unsupervised environments.

In establishing the 21st Century Community Learning Centers (21st CCLC) program, Congress and the U.S. Department of Education (ED) embarked on a strategy to encourage schools and community organizations to work together to expand opportunities for after-school, weekend, or summer programs for children, youths, and families. School districts can use 21st CCLC funds to plan, implement, or expand learning centers located within public school buildings.

In September 1999, ED selected Mathematica Policy Research, Inc. and its partner, Decision Information Resources, Inc., to conduct a national evaluation of the 21st CCLC program. The C. S. Mott Foundation, which has long been a leader in supporting school-community partnerships, also is providing support to enhance the national evaluation.

Background on After-School Programs

A national study found that in 1991, 1.7 million children in grades K through 8 participated in before- and after-school programs. The Census Bureau estimated that in 1995, 2.1 million children ages 5 to 14 were in before and after school programs. Both estimates suggest that about five percent of children in grades K through 8 participate in such programs.

Both educational and economic forces appear to be increasing the demand for programs for before and after school programs. On the educational side, efforts to raise academic achievement have increased the push for after-school programs to provide more opportunities for instruction and academic enrichment. And as more women enter the workforce, their children need a safe place to go when school is out. Children in "self-care"—sometimes called latchkey kids—are a concern for parents and for policymakers. The Census Bureau has estimated that in 1995, 6.9 million children ages 5 to 14 were in self-care, for six hours a week on average. Self-care was more common for older children, but its more common nature does not mean it is necessarily a good thing. Researchers have noted that rates of juvenile delinquency are highest between 3 P.M. and 6 P.M., when school is out, and that substance use, smoking, and other undesirable behaviors often occur after school.

The relationship between after-school programs and student outcomes has been a subject of a number of research studies exploring whether such programs improve grades, school

attendance, and standardized test scores. The picture that emerges from the studies is mixed, with some outcomes being positive and others being unchanged or negative. Some studies have both positive and negative findings. Studies of the effects of after-school programs on other outcomes, such as drug and alcohol use, behavior problems, and self-esteem, also yield mixed findings. Whether outcomes improve more when students receive a larger after-school program "dose" of services also is unclear.

The inconsistent nature of the findings is not surprising, considering the many differences in the studies, which use a range of evaluation designs, data collection instruments, and analysis techniques to examine programs with widely differing structures. Proponents of after-school programs can cite evidence that programs improve outcomes. Opponents can cite evidence to the contrary. Almost none of the studies used rigorous evaluation designs, and more persuasive evidence clearly would be useful in the debate.

The 21st Century Community Learning Centers Program

The 21st CCLC program was authorized by the Improving America's Schools Act of 1994 (P.L. 103-382, Title X, Part 1, sec. 10901-10907). In fiscal year 1998, Congress appropriated $40 million for the program. The appropriation increased to $200 million in fiscal year 1999, $453 million in fiscal year 2000, and $846 million in fiscal year 2001. Grants are for three years and can be awarded only to local education agencies (roughly equivalent to school districts). Centers must operate within public school buildings and must offer at least 4 of 13 activities listed in the authorizing legislation: literacy education programs; senior citizen programs; day care services; integrated education; health, social service, recreational, or cultural programs; summer and weekend school programs; nutrition and health programs; expanded library service hours; telecommunications and technology education; parenting skills; education programs; support and training for child day care providers; employment counseling, training, and placement; services for people who leave school before graduating from secondary school; and services for people with disabilities.

The program currently provides support for grantees in 903 communities and 3,600 schools across the country. Tabulations based on performance-reporting data submitted by grantees in April 2000 indicates that schools served by the program had, compared to the average school, much larger proportions of minority students and a far greater likelihood of being considered "high poverty." Nearly all centers were open 10 or more hours a week, and a third were open 20 or more. The vast majority of centers provided reading, math, and science activities. Also common were enrichment and support activities, including art and music, technology, and social studies.

Framework for the National Evaluation

The national evaluation will provide an understanding of how centers supported by 21st CCLC funds operate, how they affect participants, and how they can improve their services and activities to be more effective. The evaluation represents a partnership between ED and the C. S.

Mott Foundation. The ED component of the evaluation focuses on expanding knowledge of the effectiveness of 21st CCLC after-school programs and identifying practices and approaches that foster effectiveness in different contexts and for different student populations. The Mott component builds on the ED component by focusing on quality program elements and adding more grantees to the national evaluation, and by broadening the study components by adding surveys of principals, program staff, program participants, and nonparticipants. Plans also call for extending the follow-up period in selected sites beyond the two-year follow-up period of the ED evaluation. A companion report describes more fully the enhancements to the national evaluation that are being carried out with support from the Mott Foundation.

The national evaluation has two broad areas of inquiry: (1) local implementation of the national program, and (2) the program's impact on participants. Within the first, the goal is to understand how 21st CCLC programs are implemented and structured, as a means to identify ways to increase their effectiveness and enhance their sustainability beyond the three-year federal grant. Within the second, the goal is to assess whether 21st CCLC after-school programs improve students' in-school performance and out-of-school experiences and behaviors, for whom these programs work, and under what circumstances they work.

The evaluation is based on a conceptual framework of how after-school programs affect student learning, behavior, and personal growth, and the designs of the implementation and impact studies are guided by the conceptual framework. The framework identifies five study areas: (1) the context in which each after-school program operates; (2) the implementation of the after-school program itself; (3) family, individual, and community moderating factors; (4) intermediate effects of after-school programs; and (5) long-term effects. Programs are assumed to be affected by the local context and by their implementation experiences. Students are assumed to be affected by the local context, by moderating factors, and by the program, both on intermediate outcomes (such as completing homework, having a more positive attitude toward school, and having fewer behavioral incidents) and on long-term outcomes (such as improved attendance, grades, and test scores). The model posits that students will experience intermediate effects before they experience long-term effects.

Some intermediate effects might be considered as stand-alone accomplishments for after-school programs. For example, improved self-concept may be an important effect of after-school programs, regardless of whether it leads to higher grades or pursuit of postsecondary education. In the context of the 21st CCLC program, however, there is a clear expectation that the program will improve school performance and out-of-school behaviors.

Design of the National Evaluation

The desire for rigorous measurement of the effects of after-school programs on students argues for the national evaluation to use random assignment as much as is feasible. The desire that findings be broadly representative of the 21st CCLC program argues for selecting a range of sites that span different kinds of programs and geographic regions. However, many of the nascent programs funded by 21st CCLC grants did not yet enjoy the level of demand for their services that would support using random assignment on a broadly representative basis. The

lack of demand led to the choice of a two-part design. For the first part, the evaluation identified grantees serving elementary school children that had excess demand for their services, and worked with the grantees to set up random assignment of eligible children into treatment and control groups. For the second part, the evaluation randomly sampled grantees that served middle school children, and will use matching methods to create comparison groups of students similar to those who participate in 21st CCLC programs.

Random assignment was implemented successfully at seven grantees at the start of the 2000-2001 school year. To achieve its target precision levels, in spring of 2001 the evaluation will continue its efforts to add grantees, with a goal of having up to 20 grantees conduct random assignment. With the planned total sample size, the minimum detectable effect size will be about 9 percent for an outcome with a mean of 50 percent, which translates roughly into the ability to detect about a 5 percentage-point change in the outcome. (The minimum detectable effect size is the effect at which statistical tests of significance are likely to reject the hypothesis of no effect, expressed as a percentage of the outcome's standard deviation.) Levels of statistical precision for the first cohort of seven grantees indicates that a minimum detectable effect size will be about 18 percent, about double that for the desired full sample.

For the middle school design, grantees are sampled first, and participants and comparison group members are sampled second. For the first stage, all first- through third-cohort 21st CCLC grantees that operated a center serving students in grades six through eight were stratified by region and type of community (urban or rural), and 35 grantees were sampled from the strata. Grantees that served too few middle school students (fewer than 50 middle school students) were replaced with grantees from the same stratum. In the second stage, participants were sampled from program lists, and comparison students whose characteristics most closely match program participants will be identified from other schools or from within the same school where a center operates. A questionnaire administered as part of the evaluation will be the basis for the matching.

The planned sample size for the middle school design is 5,250 students, with 150 students in each of 35 grantees and 75 program participants and 75 comparison group members per grantee. Levels of statistical precision for this sample indicate that the minimum detectable effect size is between 10 percent and 12 percent, depending on the degree to which local site characteristics affect the variability of outcomes.

Data and Methods

Implementation Study. The primary source of data for the implementation study will be two rounds of site visits, during which staff at centers, schools, districts, and partner organizations will be interviewed or will participate in focus groups. Additional information about service delivery will be gathered through observations of activities at the center. Reviews of documents related to center operations (for example, needs assessments, handbooks, and evaluators' reports) will also add insight about such topics as recruitment, plans for staff development, and sustainability challenges facing grantees.

The implementation analysis approach will vary depending on the stage of evaluation. In the first stage, during the 2000-2001 school year, the focus will be on understanding and describing how the 21st CCLC program is operating in the sites that are participating in the evaluation. In the second stage of the study, during the 2001-2002 school year, the focus will shift to exploring program characteristics and features that may be linked with program impacts.

Impact Study. The primary source of data for the impact study will be questionnaires administered to students at baseline (fall 2000) and at two follow-up points (spring 2001 and 2002). The questionnaires for elementary school students differ from those for middle school students, reflecting lower levels of maturity and development. A standardized reading test (drawn from the SAT-9 battery) also will be administered to students in kindergarten through grade five when questionnaires are administered (fall 2000, spring 2001, and spring 2002). Questionnaires also will be administered to parents of elementary school students at the same baseline and two follow-up points, to parents of middle school students at the two follow-up points, to the main classroom teacher for the elementary school sample, and to the English/language arts teachers of students in the middle school sample. In addition, information from school records will be collected, and after-school programs will provide student participation and attendance information. For elementary school grantees that begin random assignment in fall 2001, the data collection schedule will be advanced by a year.

The impact analysis will rely on well-tested methods for estimating program impacts for treatment and control/comparison designs. The starting point for these methods is the average outcomes of treatment groups and control/comparison groups. Differences in average outcomes between the groups represent impacts of the program, with the estimates being refined further through econometric techniques. The analysis will assess general impacts, as well as impacts for specific groups of students and for specific types of programmatic approaches.

Timeline and Reports

The first report from the evaluation will be issued in fall 2001, and a second report will be issued in fall 2002. A report synthesizing the evaluation's findings will be issued in spring 2003.

I. INTRODUCTION

Most school-age children have working parents, yet most schools dismiss their students long before the workday ends. During the intervening hours—"out-of-school time"—children can either develop their academic and personal skills in safe, supervised settings, or engage in unproductive behaviors in unsafe, unsupervised settings.

In establishing the 21st Century Community Learning Centers (21st CCLC) program, Congress and the U.S. Department of Education (ED) embarked on a strategy to encourage schools and community organizations to work together to expand opportunities for after-school, weekend, or summer programs for children, youths, and families. Schools or school districts can use 21st CCLC funds to plan, implement, or expand community learning centers located within public school buildings. The program made its first grants in 1998, and by 2000, over 3,000 rural and inner-city public schools in more than 900 communities were participating in the program.

In September 1999, ED selected Mathematica Policy Research, Inc. (MPR) and its partner, Decision Information Resources, Inc. (DIR), to conduct a national evaluation of the 21st CCLC program. The C. S. Mott Foundation, which has long been a leader in supporting school-community partnerships, also is providing support to enhance the national evaluation. The design of the national evaluation is the topic of this report. A companion report describes more fully the enhancements to the national evaluation that are being carried out with support from the Mott Foundation.[1]

[1]See "Enhancing the 21st Century Community Learning Centers Evaluation: A Concept Paper." Princeton, NJ: Mathematica Policy Research, Inc., November 20, 2000.

This report surveys the literature on after-school programs, lays out a conceptual framework for the evaluation and its key hypotheses, and describes plans for the evaluation's qualitative and quantitative analyses. It also provides information about the 21st CCLC grantees that have agreed to be part of the national evaluation, explains how they were identified, and lays out the data collection schedule and instruments on which the analyses will be based.

A. THE NEED FOR AFTER-SCHOOL PROGRAMS

Both economic and educational forces have increased the need for programs that serve school-age children when school is out (before and after school, Saturdays, holidays, and summers).[2] On the economic side, the proportion of married mothers who work (and, it is assumed, need some form of care for their children outside school hours) has increased almost 80 percent in the past three decades (the proportion was 38 percent in 1969 and 68 percent in 1996). Working mothers are also working more hours (Council of Economic Advisers 1999). For low-income parents, welfare reform legislation passed in 1996 doubled the number of adults required to work or participate in work-related activities by 2002, from 25 percent to 50 percent (U.S. General Accounting Office 1997), contributing further to the need for after-school programs.

On the educational side, efforts to raise academic achievement have increased the push for after-school programs that provide more instructional time and permit states and local districts to address demands for higher student performance. California, Texas, Oregon, Kentucky, Georgia, and Washington, for example, are among the states that have linked after-school initiatives with state efforts to increase academic achievement.

[2]Throughout the report, "after-school programs" is used as a general term for programs that provide services and activities outside regular school hours. Many after-school programs also provide services on holidays or weekends, but few provide services or activities at those times but not after school.

Whether the supply of after-school programs is adequate is difficult to gauge. Studies during the past decade suggest that participation in after-school programs is increasing but that most students do not participate. One of the few nationwide studies of before- and after-school programs found that in 1991 about 49,500 programs provided such care for 1.7 million children in grades K through 8. The participation rate that year was five percent (about 31.5 million K-8 students attended public and private schools, according the *Digest of Education Statistics*). About 11 percent of public school students participated in extended-day programs in the 1993-1994 school year (National Center for Education Statistics 1997). More recent data suggest that the percentage continues to rise. The National Survey of America's Families found that 21 percent of six- to nine-year-old children with working mothers had before- or after-school programs as their primary care arrangement in 1997 (Capizzano et al. 2000).

However, it is difficult to know how many students not currently enrolled in after-school programs would participate if more schools offered programs. Some children are cared for by relatives or neighbors, and parents may be comfortable with such an arrangement. Other children care for themselves. Children in "self-care"—sometimes called latchkey kids—probably are the greatest concern for parents and for policymakers. Estimates of the number of children in self-care vary considerably, with one source citing "as many as 15 million" being left alone for at least some time after school, another citing 4 million children 6 to 12 years old being left without adult supervision "on a regular basis," and another stating that "only an estimated 12 percent of children age 5 to 11 ever care for themselves, and they do so for only six hours a week on average."[3] Of course, many parents whose children care for themselves even one hour a week

[3]The 15 million figure and the 5 million figure are cited in a policy brief prepared by the Urban Health Initiative [www.urbanhealth.org/aftershool bp brief2.pdf]. The 12 percent figure is cited in a policy brief on after-school programs prepared by the Cato Institute [www.cato.org/pubs/pas/pa-372es.html].

might find other choices more appealing were they available. The evidence is more clear that rates of juvenile delinquency are highest between 3 P.M. and 6 P.M., when school is out, and that substance use, smoking, and other undesirable behaviors often occur after school (Newman et al. 2000).

B. HOW DO AFTER-SCHOOL PROGRAMS AFFECT STUDENTS?

The push for more after-school programs is related presumably to a sense that the programs are better for students than the alternatives, and the relationship between after-school programs and improved outcomes has been a subject of an increasing number of research studies. However, the studies use a range of evaluation designs, data collection instruments, and analysis techniques to examine programs with widely differing structures. Not surprising, the studies arrive at different findings.

Results of the studies can be categorized into four areas: (1) whether after-school programs improve school outcomes, (2) whether the programs improve other (out-of-school) outcomes, (3) whether extensive program participation is related to better outcomes, and (4) whether particular program characteristics are related to better outcomes. When reviewing the results, it is important to note that few of the studies used rigorous evaluation designs. Only three studies (Quantum Opportunities Program, Big Brother/Big Sisters, and Early Intervention Services for Disabled Children) used a random assignment design, and studies using comparison designs have not been sensitive to selection bias, which arises as a result of the voluntary nature of participation in after-school programs. With selection bias, differences in outcomes may be due to differences between treatment group and comparison group students rather than to participation in after-school programs.

1. School Outcomes

Studies have explored whether after-school programs improve grades, school attendance, and standardized test scores. The picture that emerges from the studies is mixed, with some outcomes being positive and others being unchanged or negative. Some studies have positive findings for some outcomes and negative findings for others.

- ***Grades***
  - Participants in the Boys and Girls Clubs' Educational Enhancement Program (EEP) had higher grades than the comparison group in reading, spelling, history, science, and social studies (Schinke et al. 1998).
  - Participants in Coca-Cola's Valued Youth Program had higher reading grades than comparison students (Fashola 1999).
  - Participants in an after-school recreation program had better math, reading, and science scores than nonparticipants (Baker and Witt 1996).
  - Participants in after-school programs in one urban school district had better reading and math grades than children in other types of care (Posner and Vandell 1994).
  - Participants in the LA's BEST program had test scores higher than those of comparison students in reading, math, science, social studies, and composition (Brooks et al. 1995).
  - Participants in the New York City Beacons after-school program felt that the program was "very helpful" or "pretty helpful" in helping them do better in school (Warren et al. 1999).
  - Big Brothers/Big Sisters program participants had higher grades than the control group (Tierney et al. 1995).[4]

However, these and other studies also yield evidence that after-school programs do not improve, and may even reduce, academic achievement.

- Boys and Girls Club EEP participants did no better than comparison students in English, writing, and geography (Schinke et al. 1998).

[4]The Big Brothers/Big Sisters program is not an after-school program per se, but it shares some of the same elements and is often included in discussions of after-school programs.

- Participants in an after-school recreation program had the same language grades as nonparticipants (Baker and Witt 1996).
- Participants in after-school programs in an urban school district had the same test scores in reading as did children in other types of care (Posner and Vandell 1994).
- Verbal and reading scores for children in after-school child care centers in Dallas were no different from those of children in other types of care, and their math scores were lower (Vandell and Corasaniti 1988).
- Participants in LA's BEST scored lower than comparison group students in reading, math, science, social studies, and composition when different measurement techniques were used (Brooks et al. 1995).

Other school-related outcomes also show inconsistent patterns across studies.

- ***Dropout Rates, Absences, and Tardiness***
  - Coca-Cola's Valued Youth Program participants had lower dropout rates than students in the comparison group (Fashola 1998).
  - Big Brother/Big Sister participants had fewer school absences than control group members (Tierney et al. 1995).
  - Participants in an after-school recreation program were absent from and late to school the same as nonparticipants (Baker and Witt 1996).

- ***Conduct and Work Habits***
  - Participants in after-school programs in an urban school district had better conduct ratings and work habits than children in other types of care (Posner and Vandell 1994).
  - Big Brother/Big Sister participants spent the same amount of time on homework and reading as control group members (Tierney et al. 1995).

2. Out-of-School Outcomes

Studies of the effects of after-school programs on out-of-school outcomes, such as risky behavior, drug and alcohol use, behavior problems, and self-esteem, also yield mixed findings.

- ***Drug and Alcohol Use***
  - Big Brother/Big Sister participants were less likely to start using drugs or to drink alcohol (Tierney et al. 1995).

- Housing projects with the Boys and Girls Clubs' Smart Moves substance abuse program had less drug activity than similar housing projects without such a club (Schinke et al. 1991).
- New York City Beacons after-school program participants felt that their participation was "very helpful" or "pretty helpful" in helping them avoid drug use (Warren et al. 1999).

- ***Behavior Problems***
  - Participants in after-school programs in Boston had fewer behavior problems than children who did not attend after-school programs (Marshall et al. 1997).
  - Most New York City Beacons after-school program participants felt that their participation was "very helpful" or "pretty helpful" in helping them avoid fighting (Warren et al. 1999).
  - Participants in an after-school recreation program had similar behavior and measured self-esteem as comparison group students (Baker and Witt 1996).
  - Participants in the Boys and Girls Club EEP had the same number of behavioral incidents as other students, on average (Schinke et al. 1998).

- ***Arrests***
  - Housing projects with the Boys and Girls Clubs' Smart Moves substance abuse program had lower juvenile arrest rates than similar housing projects without such a club (Schinke et al. 1991; Quinn 1999; and U.S. Department of Education 1998).
  - Big Brother/Big Sister participants were as likely to steal or to damage property as control group students (Tierney et al. 1995).

- ***Self-Esteem, Attitudes, and Relations with Peer and Adults***
  - Coca-Cola's Valued Youth Program participants had higher self-esteem and better attitudes toward school (Fashola 1999).
  - New York City 4-H Club participants scored higher than other youth on educational aspiration, motivation, self-esteem, and decision-making (Rodriguez et al. 1999).
  - Students in formal after-school programs in Milwaukee were better adjusted and had better peer relations (Posner and Vandell 1994).
  - Participants in The After-School Corporation (TASC) program in New York City who previously had behavior problems were developing more self-control, and students in general were developing more self-discipline and motivation (Fiester et al. 1999).
  - Students in a 21st CCLC program in San Francisco had the same levels of self-esteem, social support, self-efficacy, and conflict management skills as comparison students (Trousdale 2000).

- After-school program participants in Milwaukee had the same levels of anxiety and hyperactivity as other children (Posner and Vandell 1994).

3. Is More Better?

Whether students do better when they receive a larger "dose" of after-school programs is unclear. Vandell and Pierce (1999) found that students who frequently attended after-school programs had better work habits and attendance. Ross et al. (1996) found that students performed better in school when they attended the after-school program more often, and Rodriguez et al. (1999) found that youth who were in 4-H clubs for longer than one year scored significantly higher on leadership, conflict resolution, communication, and self-confidence.

These results are consistent with those of other studies of additional instructional time. Hattie et al. (1997) found that adventure programs lasting longer than 20 days had greater effects than shorter programs. Fusaro (1997) found that children in full-day kindergarten performed better on achievement tests than children in half-day programs. Evans and Marken (1984) found that children attending full-day kindergarten have a significantly better attitude toward reading than children attending half-day kindergarten. Walberg and Tsai (1984) found that increased time spent on homework and on leisure reading were associated with higher reading scores. Roderick et al. (1999) found evidence that summer programs had a large positive effect on the percentage of students who were able to reach the Chicago Public School minimum promotion standard. Cooper et al. (1999) found that participation in summer programs that focused on remedial or accelerated learning improved the knowledge and skills of participants. From a broader perspective, Card and Krueger (1992) found that a longer school year was associated with higher income as an adult.

However, other studies have found no effects, or negative effects, of additional instruction time. Taylor et al. (1993) found that additional assistance (through visits from trained staff) for disabled children did not improve outcomes, and participants performed lower than control

group students on the Battelle Developmental Inventory. Evans and Marken (1984) found that children in full-day kindergarten performed no better on California Achievement Tests, attitudes toward school, and personal/social behavior than students in half-day kindergarten. Cohen et al. (1982) found that students tutored for longer periods were less positively affected than students tutored for shorter periods.

4. Does Program Structure Matter?

Different program structures could be associated with different effects. For example, ratios of staff to students, the ways in which students choose their activities, and the ways in which staff interact with students could have different effects on students. To date, the literature on these issues contains only a few studies, but they do suggest that program structures matter. Rosenthal and Vandell (1996) report that higher child/staff ratios are associated with more negative child/staff interactions, and that more flexible offerings are associated with more positive child interactions with staff. Pierce et al. (1999) report that programs with more positive staff attitudes had fewer behavior problems among boys, that programs that had more negative child/staff interactions were associated with lower test scores for boys, and that more program flexibility was associated with boys having better social skills (though lower reading and math grades). Cooper et al. (1999) reported that programs with small group or individual instruction and programs requiring parental involvement had larger effects on outcomes.

5. Interpreting the Literature

Seeing the common threads in the studies reviewed here is challenging. Proponents of after-school programs can cite evidence that after-school programs improve outcomes. Opponents can cite evidence to the contrary. More rigorous evidence clearly would be useful in the debate.

One reason for the inconsistent nature of the findings is the small scale of many of the studies. With small sample sizes, estimated impacts are naturally subject to variability.

Another reason for the disparate findings is the wide range of evaluation designs used. Most studies have used pre-post designs, correlation analysis, regression analysis, and comparison designs. As noted previously, only three studies used random assignment. The studies using comparison designs may be affected by selection bias, which makes it difficult to know whether the observed outcome differences are attributable to program effects or to differences between the participant groups and the comparison groups. For example, the LA's BEST evaluation drew its comparison group from students in the same schools as program participants but who did not participate in the program or who participated for less than three months (Brooks et al. 1995). The TASC evaluation compares participants to nonparticipants from the same schools, but it is not clear whether the evaluation adjusts for differences between participants and nonparticipants (Fiester et al. 1999).

Studies that construct their comparison groups in these ways—using nonparticipants from the program schools or students who stay in the program for a short period of time—generally face the analytic problem that students who participate voluntarily in after-school programs (or participate because their parents want them to) may differ in important ways from those who do not participate. For example, participants may be more motivated than nonparticipants. If so, outcome differences may arise from the greater motivation level of participants, rather than from participation in after-school programs. Moreover, participants who are in programs for longer periods of time may differ systematically from those who participate for shorter periods. Once again, if students differ in this way, impact estimates attributed to staying in the program for a longer period of time may be the result of differences between long- and short-stay students. Analyses that are sensitive to these issues of impact measurement will be useful.

C. THE 21ST CCLC PROGRAM

The 21st CCLC program was authorized by the Improving America's Schools Act of 1994 (P.L. 103-382, Title X, Part 1, sec. 10901-10907). In fiscal year 1998, Congress appropriated $40 million for the program. Subsequently, the appropriation increased dramatically, to $200 million in fiscal year 1999, $453 million in fiscal year 2000, and $846 million in fiscal year 2001. The program now supports 21st CCLC grantees in 903 communities and 3,600 schools across the country (U.S. Department of Education 2000b and more grantees will be funded in 2001). Grants are for three years and can be awarded only to local education agencies (roughly equivalent to school districts). Centers must operate within public school buildings and must offer at least 4 of 13 activities listed in the authorizing legislation.[5]

Limited data available about grantees provides some basis for assessing the size and scope of activities currently being carried out by grantees. Data submitted by grantees on Annual Performance Reports (APRs) in April 2000 provide information about the size of typical

[5]The legislation stipulated that grants were to be used to support not less than four of the following activities:

1. Literacy education programs
2. Senior citizen programs
3. Children's day care services
4. Integrated education, health, social service, recreational, or cultural programs
5. Summer and weekend school programs in conjunction with recreation programs
6. Nutrition and health programs
7. Expanded library service hours
8. Telecommunications and technology education programs
9. Parenting skills education programs
10. Support and training for child day care providers
11. Employment counseling, training, and placement
12. Services for people who leave school before graduating from secondary school
13. Services for people with disabilities

programs, the usual services provided, the types of schools in which programs operate, and the contributions of partnership organizations.[6] Tabulations from these data indicate that:

- Grantees reported serving 696 students and 248 adults (who may be parents or other adults from the community).
- On average, centers providing services at least two days per week served 156 students. More than half of students attending these centers programs attended 30 days or more.
- Schools served by 21st CCLCs had much larger proportions of minority students than the average school. The racial/ethnic breakdown of these students was white (43 percent); black or African American (26 percent); Hispanic or Latino (24 percent); and others, including Asian, Hawaiian or Pacific Islander, and Native American (6 percent). National Center for Education Statistics (NCES) data show that the racial/ethnic breakdown of public school students in 1998-1999 was white (63 percent), black (17 percent), Hispanic (15 percent), and others (5 percent).
- Schools served by 21st CCLCs were much more likely than the average school to be considered "high poverty." Of those schools served by 21st CCLCs that provided information on the number of their students eligible for free or reduced-price lunches, more than two-thirds (66 percent) would fall into this category, with more than half their students eligible for free or reduced-price lunch. Nationally, NCES data show that in only 17 percent of schools were more than half the students eligible for free or reduced-price lunch.
- Nearly all centers were open 10 or more hours a week, and a third were open 20 or more.
- The vast majority of centers provide reading, math, and science activities. Also common were enrichment and support activities, including art and music (72 percent), technology (70 percent), and social studies (64 percent).
- Most centers reported that they collaborated with their partners in providing services (79 percent), setting goals and objectives (73 percent), sharing techniques for conducting activities (77 percent), providing volunteer staff (72 percent), and providing paid staff (69 percent). Fewer than half of centers collaborated with their partners to raise funds (39 percent).
- Centers commonly communicated with the regular school program to recruit and refer students (95 percent), to provide feedback on students (93 percent), to set goals and objectives (92 percent), to communicate curricula (89 percent), and to share

[6]APR data are available for grantees in the first and third cohorts. Data for grantees in the second cohort were provided in November 1999, but most of the second-cohort data were provided to ED in a different format and could not be combined with first- and third-cohort data.

instructional practices (90 percent). Ninety-three percent of centers reported that at least one staff member from the regular school program worked at the center.

Information gleaned from other sources also helps to explain how 21st CCLCs are operating. Discussions with 21st CCLC directors as part of preliminary design activities revealed that centers varied widely in the types of students served. Some programs were focusing their efforts on providing academic support to students who were behind grade level or needed help in math or reading. Other centers were serving all students who walked in, regardless of whether they were low achievers or high achievers. Centers also varied in their academic focus. Many program directors indicated that 21st CCLCs were distinctive in their academic focus compared to other after-school programs in their localities, but the range of academic activities varied from providing help with homework to offering structured curriculum linked with the regular school program.

The work of the national evaluation will provide a better understanding of how 21st CCLCs operate, how they affect participants, and how they can improve their services and activities to be more effective. The chapters that follow provide detailed descriptions of how the evaluation will be carried out.

II. FRAMEWORK FOR THE NATIONAL EVALUATION

The goals of the 21st CCLC national evaluation program are to understand how local programs are implemented, whether participation in after-school programs by elementary and middle school students has an effect on in-school and out-of-school outcomes, and whether variation in local program implementation affects those student outcomes. The evaluation, funded by ED's Planning and Evaluation Service, is intended to contribute information about the effectiveness of 21st CCLC after-school programs, and identify practices and approaches that foster effectiveness in different contexts and for different student populations.

To this core study, the C. S. Mott Foundation is funding an enhancement that builds on the ED objectives by creating opportunities to focus on quality elements of after-school programs and to examine the extended effects of after-school programs (that is, those beyond two years). The combined resources of ED and Mott reflect a joint commitment to address the challenges involved in developing and sustaining after-school programs and in obtaining credible evidence about the effects of after-school programs on student performance and behavior.

One of the design challenges for any national evaluation is to identify clearly what will be studied, as a way of shaping the tools and methods of inquiry that the evaluation will use. This chapter describes the general principles that the 21st CCLC evaluation will follow and a logic model of how programs can affect students. The logic model provides a starting point for structuring both the implementation and the impact analyses and for identifying linkages between the two. Subsequent chapters provide more discussion about the two analyses.[1]

---

[1]Much of the discussion in this chapter parallels the discussion in the companion paper that discusses activities for the enhancement study supported by the Mott Foundation.

A. PRINCIPLES AND CONCEPTUAL FRAMEWORK FOR THE EVALUATION

The national evaluation has two broad areas of inquiry: local implementation of after-school programs and the impact of programs on participants. Within the first, the evaluation seeks to understand how 21st CCLC programs are implemented and structured, as a means to identify ways to increase their effectiveness and enhance their sustainability beyond the federal 21st CCLC grant (which extends for three years). In the second area of investigation, the goal is to assess whether 21st CCLC programs improve students' in-school performance and out-of-school experiences and behaviors, and to learn more about for whom these programs work, how they work, and under what circumstances they work. The evaluation activities for accomplishing these purposes rest on three guiding principles for the successful conduct of national evaluations and on a conceptual framework that links after-school programs to positive changes in students' learning, behaviors, and personal growth.

1. Evaluation Principles

Credible Impact Measures. The national evaluation is designed to ensure credibility in measuring program impacts. It uses measurement techniques designed to uncover changes in students that are attributable to participation in 21st CCLC programs and not to normal growth or other school, after-school, or home experiences. Much of the research literature on effects of after-school programs has been hampered by a limited ability to distinguish changes due to these sources. In our choice of design, we strove to identify the most rigorous design possible. When random assignment was feasible, we used it. When random assignment was not possible, we identified the comparison design that we believed to be the most rigorous.

Broad Coverage. The national evaluation of 21st CCLC programs is intentionally broad in coverage. Comprehensiveness is critical for evaluations in which several types of interventions

make up an overall programmatic initiative. National evaluations, unlike evaluations of local programs, need to be broad enough to allow for exploring competing theories and explanations related to the potential outcomes that different interventions, or their differential implementation, may yield.

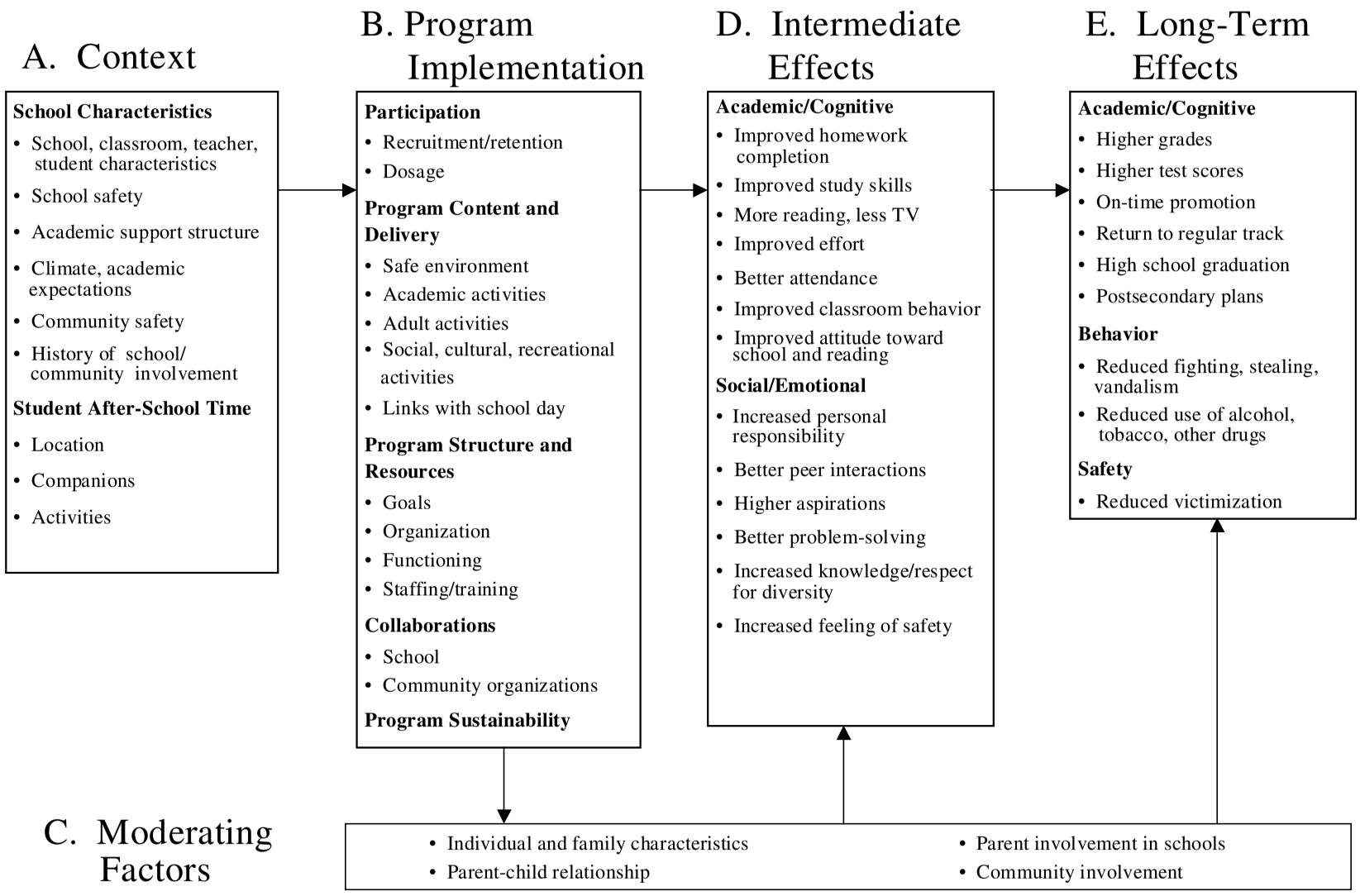
Longitudinal Approach. Finally, the national evaluation of 21st CCLC programs takes a longitudinal approach. As programs evolve, they change in terms of their organization, content, staffing, and approach. They can mature and overcome difficult startup challenges, and they can suffer setbacks (for example, changes in key staff positions, resources, or broader policies affecting after-school services). It is important that these changes be captured and results interpreted in their light. Similarly, the length of time required for a program to produce observable changes varies. How long it will take students to accumulate the benefits of a program is a subjective judgment at the outset of an evaluation.

2. Conceptual Framework

Although there is considerable variation in implemented after-school programs and a paucity of strong research findings on which to rely, it is possible to construct a conceptual framework and general logic model to guide the national evaluation. The framework helps identify appropriate areas of inquiry, data collection, and analysis.

Figure II.1 offers a schematic representation of relationships among five main topical areas that are central to the national evaluation. These areas include (A) the context in which an after-school program operates; (B) the implementation of the after-school program itself; (C) family, individual, and community moderating factors and (D and E) students' intermediate and long-term effects. The figure highlights how after-school programs are embedded in the larger constellation of school, community, and family influences that contribute to student outcomes in and out of school.

FIGURE II.1
LOGIC MODEL FOR UNDERSTANDING THE EFFECTS OF 21st CCLCs



The following sections elaborate on each topical area, indicating their relevance to the objectives of the national evaluation and how elements are related within and across areas.

a. Context

Identifying the circumstances under which after-school programs are implemented and may lead to impacts requires information about the educational and policy climate; perceptions about safety; prior community relationships; and demographic characteristics of the school, district, and community. The level of state, district, and school academic expectations and whether a school meets state performance standards establish the incumbent educational climate. Overall perceptions of safety in school and in the community are important for interpreting impact results related to students' sense of security and to differences in the success of after-school programs in recruiting and retaining participants. The historical relationship between schools and community organizations often affects how well they come together in establishing an after-school program aligned with school and community needs. Differences between urban and rural communities are important to interpreting after-school program implementation and effects on families and students, and information about student demographics is essential for assessing the accessibility of the after-school program.

Information about students' activities after the regular school day ends is critical to assessing whether 21st CCLC students have different learning experiences, form different relationships, and engage in more positive behaviors than similar students who do not participate in the program. In communities with several after-school programs, the 21st CCLC program's impacts may be lower because other students also receive structured after-school support.[2] If we know

[2]As noted in Chapter I, initial discussions with 21st CCLC program directors found that the inclusion of educational components distinguished these programs from other after-school

what students in the sample do in the hours after school, we can better interpret the impact results.

b. Program Implementation

Detailed knowledge of the operation of each program precedes an understanding of the practices and approaches that foster effectiveness in different settings and for different student groups. Variation in student participation and dosage, program content and structure, collaborations with host schools and community organizations, and efforts toward sustainability are all measures of program implementation and will play critical roles in our understanding of how programs are locally implemented and structured and what their differential effects on students are.

Participation and Dosage. Clearly, for 21st CCLC programs to change students' performance and behaviors, centers must first attract participants and sustain their involvement for some period of time. Although little empirical evidence informs the extent of after-school participation necessary to produce effects on different outcomes, some benchmarks are available. For example, ED's guidance for creating quality programs calls for after-school services to be offered a minimum of 15 hours each week (five days a week for 3 hours each day) (U.S. Department of Education 1998). If students attended at the ED-suggested weekly rate for the entire school year, the additional time spent in learning activities could be substantial.[3]

(continued)

programs in their areas. This suggests that the 21st CCLC programs may have greater impact on students' academic performance than other after-school programs in the neighborhood.

[3]For example, with a 180-day regular school calendar, a five-hour regular instructional day, and two-thirds of after-school time devoted to educationally enriched activities, a student participating in an after-school program could experience a 40 percent increase in learning time.

Program Content and Delivery. The content and delivery of a program can affect both student participation and student outcomes. Homework tutoring, focused instruction, hands-on learning, and enrichment activities are all ways in which after-school programs can address students' cognitive needs. Social, cultural, and recreational activities that present students with opportunities to explore interests, express talents, develop ownership, work in teams, and handle conflict are key to developing the social skills, confidence, and protective factors necessary for meeting challenges at home, with peers, and in the classroom. Important factors of content delivery, such as class size, the way participants select or are assigned to activities, the style and temperament of staff interactions with participants, and methods for linking parents and the community to the center, are all likely to affect participants and will be important to assess.

Program Structure and Staff Resources. Building programs with the features described above requires that centers have adequate support and sound organization. Formal role specifications and procedures also are critical to ensuring a smoothly functioning and predictable environment that results in high retention of staff and students. A stable, qualified staff undoubtedly is a major component of an effective program, and adequate compensation and supportive working conditions are likely to enhance staff retention. The educational level of the staff may prove important if only to support stronger relationships with regular school staff and further communication and content linkages. Professional training is likely to be a significant factor in centers' effectiveness, because it focuses on the skills staff need rather than their initial qualifications.

Collaborations. Regular interactions and communications that link an after-school program with school and community collaborators will be critical to a program's ongoing relationship with its host school and the planning and operational support it receives from the community. School collaborators can create learning and developmental activities that complement and

reinforce those of the school day, provide insight into the problems faced by individual children and the modes in which they might learn more effectively, and alleviate possible tensions stemming from shared space and resources. Community collaborators may expand the expertise available to the after-school program, provide opportunities for students to undertake community service, and offer opportunities for older students to intern in businesses or nonprofit organizations. It will also be important to identify the processes by which school and community collaboration operates, such as involvement in needs assessment, goal-setting, and program review processes; periodic meetings of designated lead staff from each organization; and development of clear understandings of the expectations of partnering organizations.

Program Sustainability. Developing self-sustaining after-school programs is a major goal of the 21st CCLC program. While sustainability constitutes an important outcome in and of itself, it may also play an important role in program's outcomes. After-school programs that show strong prospects for continuation are less likely to be considered marginal and may have stronger appeal to staff and parents. After-school programs that project continuity also are likely to command greater attention and time investments from regular school staff. The process of achieving sustainability is likely to depend on the extent to which the after-school program is a key component in the schools' system of support services and in the community's network of services for families.

c. Moderating Factors

External factors and relationships that intervene between the program and participating students influence the effects of after-school programs on students. Some factors and relationships will be linked with differential effects. For example, specific features of after-school programs may affect students of middle school age differently from the way they affect younger students. Students who have greater learning deficits, or who are not fluent in English

and thus more at risk academically, may benefit more from after-school programs than students less at risk. Uncovering why such differential effects occur is always challenging, but recognizing their existence can suggest steps for making program adjustments or providing strategic targeting of after-school activities.

After-school programs may alter certain factors and relationships in an effort to improve student performance and behaviors. Tracking whether changes occur in these arenas will provide important insights into how after-school programs can bring about changes in student outcomes. For example, parent-child interactions at home may improve through communicating more often and spending quality time together. Parents, for instance, may be encouraged to read more to their children and discuss their schoolwork. Increasing parents' knowledge of school programs and children's concerns in school can help parents become more proficient in discussing matters with their children and with regular school staff. Similarly, after-school programs are likely to provide a bridge for other organizations in the community to become more active in providing input and guidance to improve various elements of the regular school program.

d. Intermediate and Long-Term Effects

Consistent with the blend of needs fostering the 21st CCLC program (safe and structured childcare, learning opportunities, social/cultural/physical development), numerous student outcomes are expected of after-school programs. These include improved security, better academic performance, positive behavioral changes, and increased personal competence. In conceptualizing how after-school programs can affect these outcomes, the conceptual framework imposes a sequential order that separates effects into two stages: intermediate and long-term. Intermediate effects (for example, improved study skills, feeling of security, and increased interest in reading) are assumed to precede longer-term effects, and longer-term effects are

assumed to reflect changes resulting from the sustained influence of intermediate effects. For example, grades and test scores (long-term effects) are more likely to improve if students make greater effort in the classroom and miss less school (intermediate effects). Similarly, risky behaviors are more likely to be lessened if students first exhibit greater personal responsibility and associations with peers who share positive values.

The division of effects into intermediate and longer-term provides a temporal sense of when to expect certain changes and a reasoned basis for interpreting results over the multiyear course of the national evaluation. Arguably, some intermediate effects may be considered as stand-alone accomplishments for after-school programs. It is difficult to conclude that improved self-concept, for example, is not a benefit for youth, irrespective of whether it leads to higher grades or test scores. In the context of the 21st CCLC program, however, there is a clear expectation that the program improve measures of in-school performance, as well as measures of out-of-school behaviors. For these reasons, we consider increased aspirations and self-concept as intermediate steps to accomplishing these longer-term results.

The breadth or narrowness of goals espoused by 21st CCLC programs (and how well these goals align with centers' activities) is likely to exert a strong influence on the types of effects achieved by individual centers and by the program as a whole. While programs can produce a range of unintended results, it is reasonable to assume that programs will improve specific outcomes they seek to influence, rather than those they see as outside their mission. For example, programs that narrowly emphasize reading achievement may see reading improvements, but may not see other intermediate or long-term effects (for example, higher aspirations). After-school programs more closely approximating the broad mix of emphases that define the 21st CCLC "intended program" are likely to have impacts across a broader range of outcomes.

B. EVALUATION DESIGN AND ACTIVITIES

The conceptual framework suggests specific research questions for the implementation and impact analyses to address. Table II.1 shows specific research questions to be addressed in the implementation and impact areas.

TABLE II.1

RESEARCH QUESTIONS TO BE ADDRESSED BY THE NATIONAL EVALUATION

IMPLEMENTATION QUESTIONS
• What are the key contextual issues in the school, district, and community that affect the design and implementation of the 21st CCLC program?
• What are the program's goals and philosophies, and how are they translated into practice?
• What is the program's organizational structure, including staffing, management, and decision-making, and how does it affect implementation?
• What services are delivered to program participants, and what is the relative emphasis on academic and youth development activities?
• How do programs recruit participants, and how do participation patterns vary by subgroup and by type of activity? How are students, parents, and other adults involved in the after-school program? What factors enhance or limit their involvement?
• What collaborative structures are in place, and how do they affect implementation of the 21st CCLC program? What factors lead to strong community partnerships in after-school programs?
• What are the links between the program's activities and the regular school program and how are these links facilitated?
• What are the key challenges to implementing, operating, and sustaining 21st CCLC programs, and how are these challenges addressed?
IMPACT QUESTIONS
• Do centers improve academic and other in-school outcomes for students, such as academic achievement, on-time promotion, attendance, and classroom behavior?
• Do centers improve students' out-of-school behaviors, such as reducing incidents of violence, use of drugs, smoking, fears about being safe, and contact with gangs?
• How do centers' impacts differ for students of different ages, backgrounds, and levels of academic and English proficiency?
• What dimensions, models, and practices of centers are linked to impacts on student outcomes?
• Do students who frequently attend programs have more positive outcomes?

Figure II.2 broadly summarizes the configuration of the national evaluation, including the ED component and the enhancement study supported by the Mott Foundation. It emphasizes the two basic dimensions that frame both studies: (1) a study of program implementation that addresses the processes and experiences essential to understanding after-school programs, and (2) a study of program impacts on intermediate and long-term outcomes. A planned longer-term examination of impacts for middle school students (supported by the Mott Foundation enhancement) may extend the evaluation for four additional years. Common elements, in the form of questionnaires and site visits, will occur in the grantees sampled as ED and Mott sites. (The shaded areas in Figure II.2 represent these shared approaches.) These common elements will create the consistency needed for the data to be combined and results generated for the overall study. For example, results for program impacts will be estimated using data from all sites, regardless of their selection into the ED or Mott samples.

Figures II.3 and II.4 depict key research components and underlying questions for the impact and implementation dimensions of the national evaluation. These figures also provide an overview of the target sample sizes that make up each component. While impact and implementation data will be assessed independently to address several questions, we will combine data from both sources during the analysis phase of the study to explore how various features of programs link to greater levels of success and to help shed light on the presence or absence of results.

The implementation study as depicted in Figure II.3 encompasses a core set of case studies involving all sampled sites and an in-depth round of case study visits designed to capture important aspects of quality after-school programming in the Mott sites. Overall, the implementation study seeks to understand how 21st CCLC programs are locally implemented and

FIGURE II.2
ED AND MOTT STUDY COMPONENTS

2000 2003

IMPACT STUDY

ED Impact Study 30 middle school sites (comparison design)
Up to 20 elementary school sites (RA design)

Mott Enhancement 5 middle school sites (comparison design)
Up to 5 elementary school sites (RA design)

Mott enhancement
Student followup in selected MS sites

IMPLEMENTATION STUDY

ED Implementation Study
2 visits to 30 MS sites and up to 20 ES sites

Mott Enhancement
Principal and staff survey for all sites
3 visits to 5 MS sites and up to 5 ES sites

Mott Enhancement
Study of after-school time in 5 MS sites
Participant and nonparticipant surveys 10 sites
Lessons from non-21st CCLC programs

Gray shade indicates components sharing instruments or protocols

FIGURE II.3

COMPONENTS OF THE 21ST CENTURY IMPLEMENTATION STUDY

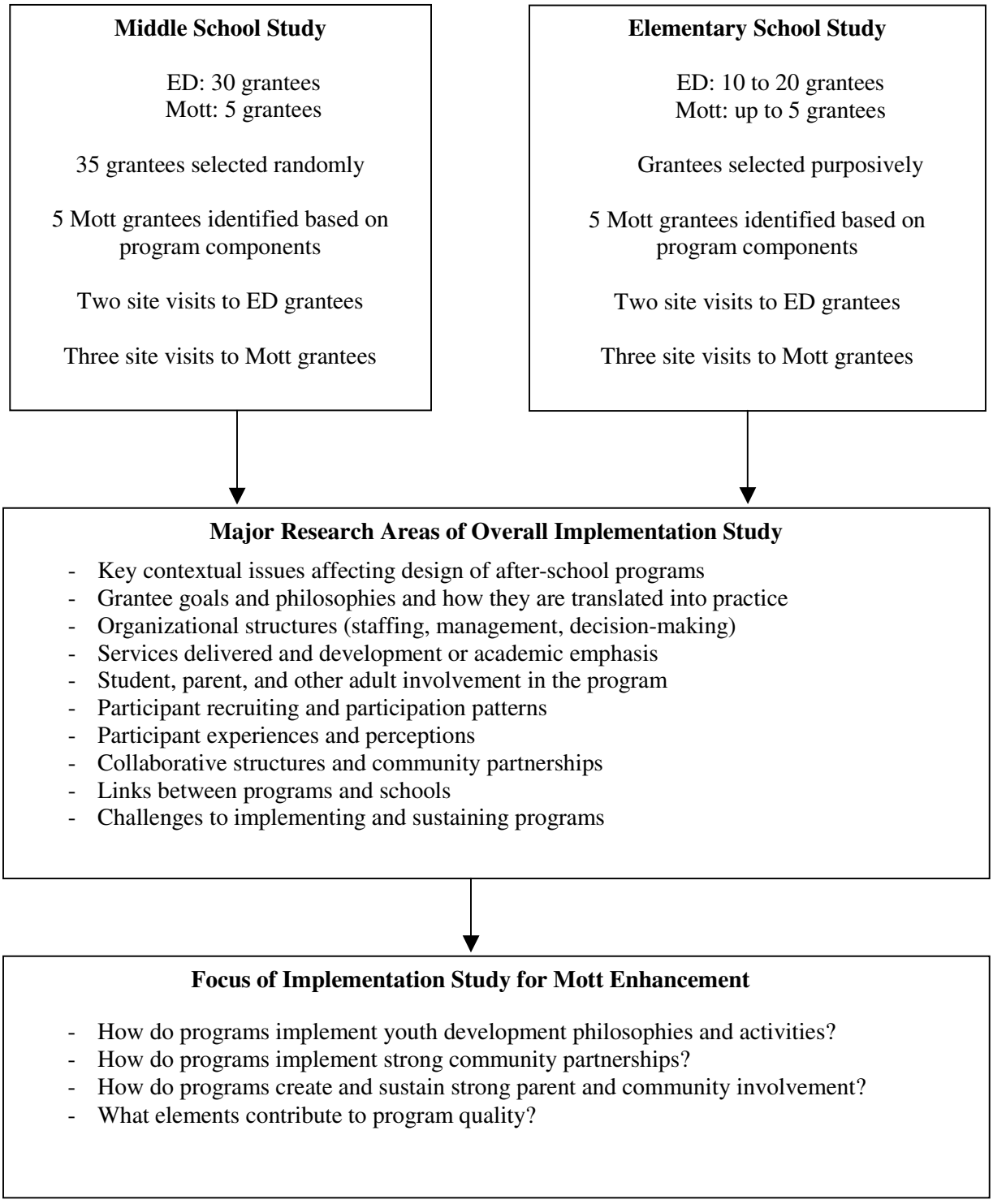
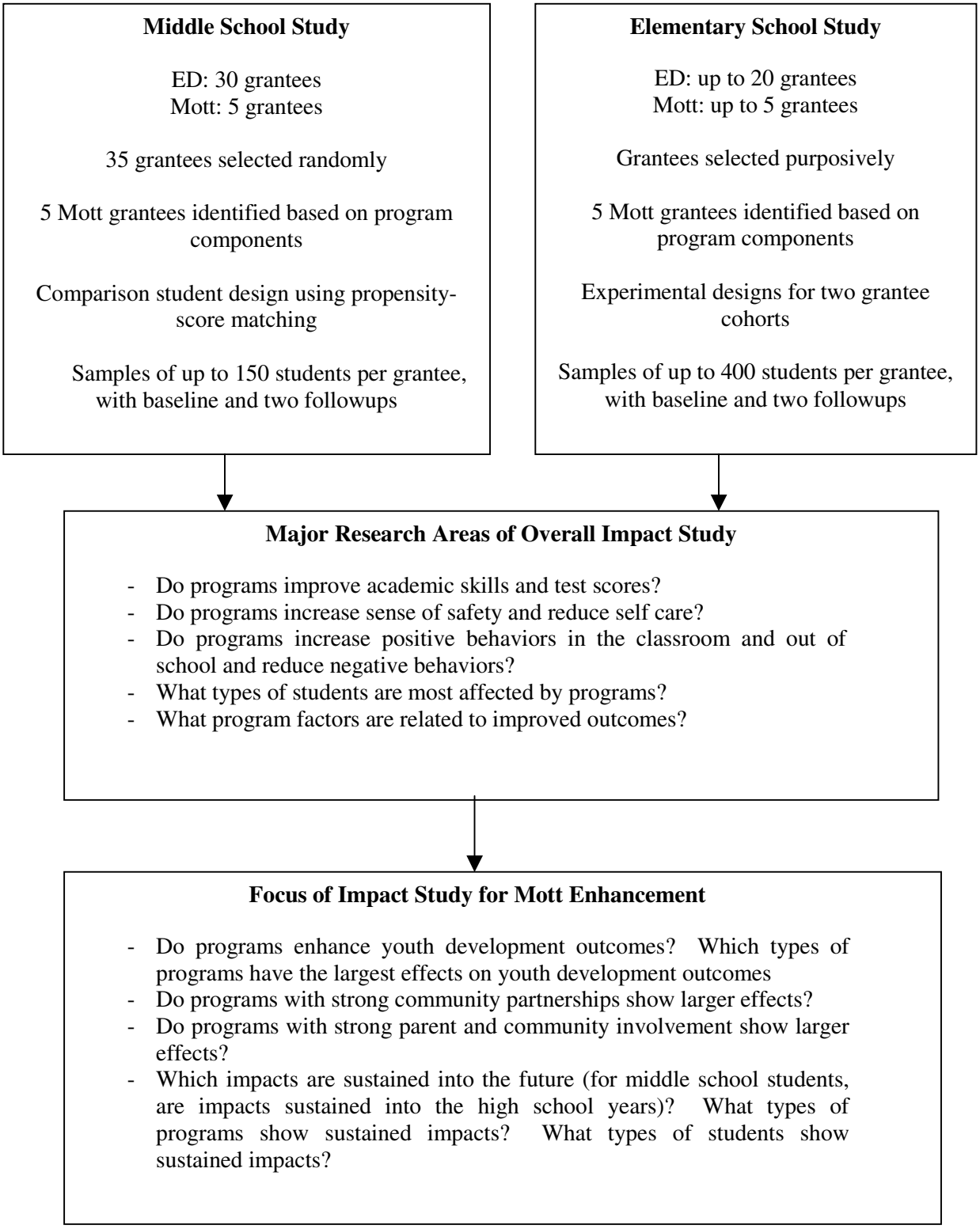


FIGURE II.4

COMPONENTS OF THE 21ST CCLC IMPACT STUDY



structured with the intent to identify ways to increase effectiveness and sustain local projects beyond the federal grant. The figure notes the major questions guiding the implementation study. They include contextual factors affecting program design, how grantee goals and philosophies translate into practice, how grantees structure their programs, what experiences are afforded students in after-school centers, how school and after-school staff align curricula and collaborate, and so on. The elements in the context and program implementation boxes of the logic model serve as the focal point for the implementation study.

The core implementation study, like the impact study, involves several common protocols, interview guides, and grantee/center assessment forms to ensure that both the ED- and the Mott-sampled sites generate consistent findings. The case studies that are part of the core implementation study will be based on two rounds of site visits during which relevant parties associated with each center, host school, district, and partner organizations will be interviewed or participate in focus groups. To gather a comprehensive picture of the design and functioning of the centers, we will tap a wide range of respondents.[4] A key topic explored in the site visit discussions will be the patterns of service delivery for different students (for example, the amount of time associated with different activities and the likely mix of activities for students during the course of a year). Further information about actual service delivery will be gathered through observations of a sampling of program activities at the center. Reviews of documents related to center operations (for example, needs assessments, handbooks, and evaluators' reports)

[4]The list includes project directors, center coordinators, center staff, community partners, district representatives, school principals, regular school teachers with students attending the after-school program, middle school students, parents of participants, and, in some cases, other adult participants in the center.

will also add insight about such topics as recruitment, plans for staff development, and sustainability challenges facing the center.

The Mott enhancement provides an additional opportunity to visit sites and obtain greater insights into their actual workings and their evolution. The third box in Figure II.3 indicates likely issues that this additional visit will explore in the Mott sites. These include how programs implement philosophies and activities directed toward youth development, how programs work to create and sustain strong community partnerships, and how programs work to create and sustain parental involvement. Some investigation of these issues will be undertaken in all sites, to provide for comparability across the two studies. However, the additional resources the Mott enhancement provides will facilitate deeper, more thorough investigations of these issues in selected sites. In addition, these focal issues may change after the first round of core site visits, if analyses drive other compelling issues to the forefront for the third site visit.

The focus of the impact study is an investigation of the elements in the logic model (Figure II.1) categorized as moderating factors, intermediate effects, and long-term effects. The core impact questions concern whether programs increase cognitive skills and lead to improved test score and grades; whether programs enhance safety; whether programs increase social and emotional skills of students; and whether they promote positive behaviors and reduce negative ones. The impact study also will explore the types of students who experience significant effects and the types of programs that have significant effects.

The impact measurement design for the overall evaluation combines two distinct evaluation designs. The first, for elementary schools, is a purposive selection of 21st CCLC programs that serve elementary school students, with random assignment of students into treatment and control groups. The second, for middle schools, is a random sample of 21st CCLC programs that serve middle school students, with comparison students selected to match students in the program

group. The data for the impact study will come from a variety of sources. Parent and student background data and information about outcomes (for example, reading and television watching, feelings of safety, aspirations, activities after school, parent/child interactions) will be gathered through baseline and follow-up questionnaires. Teacher questionnaires for each student in the impact sample will provide two rounds of information about students' academic/cognitive performance and behaviors in their regular classrooms. In addition, SAT 9 reading tests will be administered to all elementary students in the impact sample. School records will provide information about students' grades, progression, and school attendance. To acquire key information to examine dosage questions, the evaluation will gather data on daily attendance in after-school programs.

To carry out analyses of what kinds of programs work for whom and under what conditions, the national evaluation will pool subgroups of students and programs to correspond with key dimensions—for example, students who are more at risk, or programs that place a strong emphasis on parent involvement strategies. Chapter V provides more discussion about how programs will be coded into types.

III. CHOOSING SITES AND SAMPLING STUDENTS

Grantees that will participate in the national evaluation will play a crucial role in what the evaluation will learn. This chapter describes how grantees were identified and how students will be sampled. It also assesses statistical power for the evaluation.

The desire for rigor in the national evaluation argues for using random assignment as much as feasible. The desire for findings to be broadly representative of the 21st CCLC program argues for selecting a range of sites that span different kinds of programs and geographic regions. However, many programs funded by 21st CCLC grants did not have the high level of demand for their services that would support using random assignment to create control groups. The trade-off between rigor and representativeness was resolved during the evaluation's first year by setting up two distinct evaluation designs. Elements of the first, the "Elementary School Design," include a purposive selection of grantees that serve elementary school students and random assignment of eligible students into treatment and control groups. Elements of the second, termed the "Middle School Design," include random sampling of grantees that serve middle school students and the use of matching methods to create a comparison group of students similar to those who participate in the 21st CCLC programs. The two designs have distinct characters and the following discussion describes them separately.

A. IDENTIFYING SITES FOR THE ELEMENTARY SCHOOL DESIGN

The most powerful approach for estimating impacts is to identify a group of eligible persons that is larger than a program is able to serve, and then to randomly assign some persons to receive program services and others not to receive them. By the properties of random assignment, the persons not selected to receive services should, on average, resemble persons selected to receive services in terms of characteristics that can be observed, such as age, sex, and

race, and also characteristics that cannot be observed, such as motivation, resilience, and taste for risk. The experiences and outcomes of persons not selected to receive services provide the basis for understanding what would have happened to participants if they had not participated in the program—the "counterfactual."

For the national evaluation, the challenge for the research team was to identify grantees that serve elementary school students and that could carry out random assignment, in terms of having both excess demand for the after-school program and a willingness to support random assignment in the organizational sense. To identify suitable grantees, the evaluation used an extensive screening process with three key steps:

Step One — Calls focusing on program enrollment and recruitment processes were made to over 300 directors of 21st CCLC programs that served elementary school students. Directors were asked whether their centers were at capacity, whether centers at capacity had a waiting list, how large that list was, and how long students had been on it.

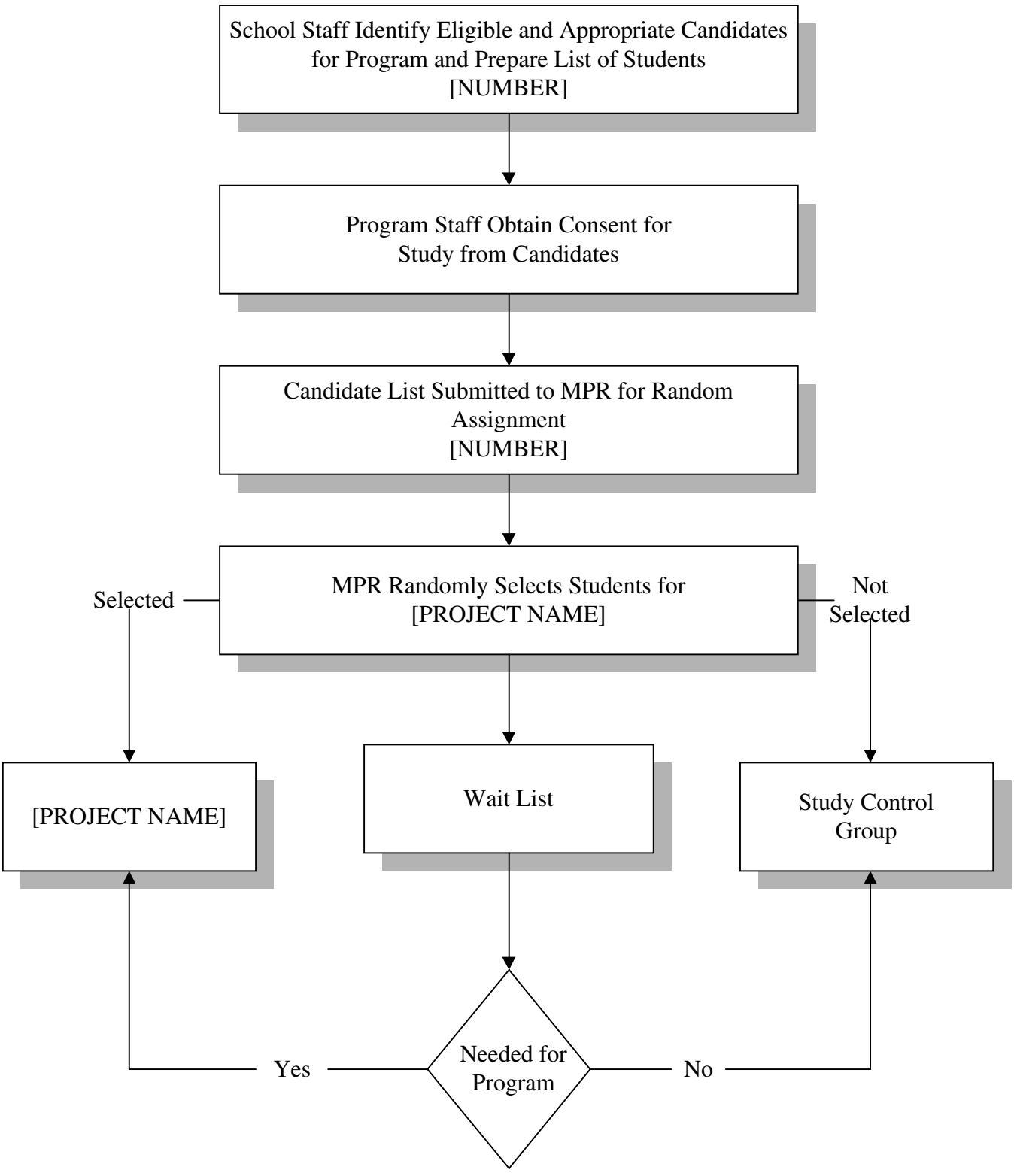
Step Two — For grantees that were at capacity or had waiting lists for one or more centers, a team member followed up and probed into the timing of the enrollment process, the type of participation patterns at each center, the flow of students in and out of the program over the course of the year, the size of the waiting list, and how frequently students moved from the waiting list into the program. After completing the second call, sites with high potential to implement a random-assignment design were identified. Sites that were dropped generally were those that had only transitory waiting lists or that planned to address their capacity problems by expanding.

Step Three — A senior team member visited sites with high potential and held discussions with district and program administrators on the feasibility of carrying out random assignment. A crucial objective of the meetings was to determine the nature of issues that administrators had regarding the use of random assignment. If discussions were fruitful, a memorandum of understanding was drafted that defined the responsibilities of the district, the program, and MPR and was signed by senior officials of the respective organizations.

The random-assignment process outlined in the agreements generally follows the hypothetical process illustrated in Figure III.1.

FIGURE III.1

HYPOTHETICAL RANDOM-ASSIGNMENT INTAKE PROCESS FOR [PROJECT NAME]



Discussions with grantee and district staff brought to light several issues that also were considered on a site-by-site basis in setting up the random-assignment process. In particular, procedures were set up to deal with situations in which (1) several students from the same household apply for services, (2) some students were judged to have a severe need for the program, and (3) more students needed to be selected after initial random assignment took place to fill in for students who left the program or decided not to participate.

- ***Students from the Same Household.*** In this situation, the evaluation randomly assigned families, rather than individual students, to the program and control groups. Siblings were grouped, and the same treatment or control assignment was given to the group.
- ***Exceptions.*** Teachers and principals judged a few students to be in severe need of the program, often because of risk factors in the home situation. Exceptions were made for these students by identifying their cases prior to the random-assignment process and admitting them to the program. Excepted students are not included in the evaluation sample.
- ***Filling New Openings.*** Centers had students who for whatever reason did not participate in the program after they had been selected, or students who left the program after they started. So that centers could operate as close to full capacity as possible, MPR randomly selected some students from the initial control group and admitted them to the program. This procedure essentially treats the randomly sorted control group as a waiting list. However, to preserve the evaluation's statistical power, we kept the size of the control group above one-third of the total sample of students.

1. Power and Sample Size

Ten grantees agreed to conduct random assignment at the start of the 2000-2001 school year. However, at that time, three programs had too few students applying to the programs to support random assignment, and the process was deferred for a year in those sites. The remaining seven sites resulted in a sample of about 1,000 students. To bolster the sample further, the evaluation will carry out the screening process described above for over 300 newly funded fourth- and fifth-cohort grantees during the 2000-2001 school year, with the goal of starting random assignment

with up to 13 grantees in fall 2001 for the ED-funded component of the national evaluation (and up to 5 additional grantees for the Mott-funded component).

The number of students that are randomly assigned directly affects the precision of estimated program impacts. Table III.1 shows the seven sites where random assignment is being done, as well as sample sizes in each site (as of December 1, 2000).

TABLE III.1

SITES AND SAMPLE SIZES FOR THE ELEMENTARY SCHOOL DESIGN
(First Set of Sites)

Site	Number of Participating Centers	Treatment Group Sample	Control Group Sample
Tampa, FL	1	82	62
Huntsville, AL	4	97	51
Markham, IL	5	154	44
East St. Louis, IL	4	66	60
Pasadena, CA	2	34	109
Redding, CA	2	32	31
DeKalb County, GA	4	88	88
Totals	**22**	**553**	**445**

Table III.2 shows levels of statistical precision for the sample sizes to date and for the planned sample sizes in the Elementary School Design. For the overall design, the minimum detectable effect size is 11.7 percent for the full sample on an outcome with a 50 percent mean (such as the number of students who are on grade level in reading).[1] Power levels are lower for subgroups, a result of their smaller sample sizes. Detectable effect sizes are about 23.4 percent for a 25 percent subgroup and 16.5 percent for a 50 percent subgroup. Table III.2 also shows the

[1]The minimum detectable effect size is the effect at which statistical tests of significance are likely to reject the hypothesis of no effect, expressed as a percentage of the outcome's standard deviation.

TABLE III.2

DETECTABLE EFFECT SIZES FOR THE ELEMENTARY SCHOOL DESIGN

	50 Percent Proportion			Test Score		
	Baseline Sample	Single-Site Effect Size	Pooled Effect Size	Baseline Sample	Single-Site Effect Size	Pooled Effect Size
First Set of 7 Sites	1,000	52.3%	16.5%	1,000	37.0%	11.7%
Subgroups						
25 percent subgroup	250	104.7%	33.1%	250	74.0%	23.4%
50 percent subgroup	500	74.0%	23.4%	500	52.3%	16.5%
Combined Set of 20 Sites	2,000	52.3%	11.7%	2,000	37.0%	8.3%
Subgroups						
25 percent subgroup	500	104.7%	23.4%	500	74.0%	16.5%
50 percent subgroup	1,000	74.0%	16.5%	1,000	52.3%	11.7%
Site Groups						
5 sites per group	500:500	(*)	42.7%	500:500	(*)	35.8%
10 sites per group	1,000:1,000	(*)	29.5%	1,000:1,000	(*)	24.5%

Notes to Table:

Parameters for calculating detectable effect sizes:

Outcome standard deviation	50 percent for proportion, 15 points for test score
Overall response rate	81.0 percent
Participation Rate	90.0 percent

80 percent power, 95 percent confidence level, one-tailed test

levels of statistical precision for comparing differences of effects of groups of sites. For example, there will be a detectable effect size of 42.7 percent when comparing the effect in one group of 5 sites and the effect in a different group of 5 sites, and a detectable effect size of 29.5 percent for comparing two groups with 10 sites in each group. The effect sizes are lower for test scores, because of the additional gain in precision from having a pre-test score, which reduces variability of the outcome.

The precision of the impact estimates shown in Table III.2 is lower for the first cohort of 7 programs than for the expected full sample of 20 programs. For the first cohort sample, an effect size of almost 17 percent is needed for statistical significance. Effect sizes at the level of a single site are large (more than 50 percent) and probably unreasonable to expect programs to generate.

B. SELECTING SITES AND STUDENTS FOR THE MIDDLE SCHOOL DESIGN

Random assignment is a preferred design for measuring program effects, but the method depends on excess demand for program services. Preliminary discussions with 21st CCLC program directors indicated that only a few centers serving middle school students had excess capacity. In the absence of excess capacity, random assignment would cause students to be turned away from a program that still had available space. The evaluation team deemed it unlikely that sites would agree to random assignment under these conditions.

Rather than not evaluating centers serving middle school students (for which ED had established a competitive priority in the first two grant competitions), the evaluation adopted an alternative design. A random sample of 21st CCLC grantees serving students in grades six through eight were selected, and comparison groups were to be formed from other students not participating in the centers. Where possible, comparison students would be identified from similar schools within the same district. In districts with only one middle school and in districts where all middle schools operated 21st CCLCs, comparison students would be drawn from within the same schools where centers are located.[2]

The middle school design had two sampling stages. Grantees were sampled first, and participants and comparison-group members were sampled second. To sample grantees, we

[2]Some consideration was given to a strategy of identifying comparison students from middle schools in neighboring districts. Ultimately, this was deemed infeasible.

stratified all first- through third-cohort 21st CCLC grantees that operated a center serving students in grades six through eight by region and type of community (urban or rural),[3] and sampled 35 grantees proportionately from the strata. We replaced eight of the initially sampled grantees by randomly selecting grantees from within the same sampling stratum. Three of the eight grantees that were replaced had been identified as having oversubscription in their elementary school centers, and the research team decided to consider them for the elementary school part of the study only. Five grantees that served only a few middle school students were judged too small to be included.

Table III.3 shows characteristics of the grantees in the sample and how they compare to the overall population from which they were drawn. The sample of grantees, like the population, is spread across the country, among urban and rural communities. The average grant award for the sample is larger than for all grantees, which may be because several small grantees were excluded and because the fourth-largest award winner, Kalamazoo, Michigan, was selected into the sample.

[3]Grantees outside the continental United States were excluded from the selection process.

TABLE III.3

CHARACTERISTICS OF MIDDLE SCHOOL GRANTEES IN THE SAMPLE

	All Grantees (N = 418)	Sample of Grantees (N = 35)
Average First-Year Award (in Dollars)	423,121	515,446
In Urban Area (Percentage)	46	51
Region (Percentage)		
Northeast	12.9	14.3
East	11.7	11.4
Southeast	13.6	11.4
North Central	15.3	14.3
Midwest	7.9	11.4
Midsouth	13.6	11.4
Northwest	10.5	14.3
Southwest	14.4	11.4

SOURCE: U.S. Department of Education database on 21st Century Community Learning Centers Grantees.

Table III.4 provides additional information on each grantee in the sample, including the grantee's region, type of community, city and state, and award amount received; whether the comparison design is "between" or "within" school (or, in two instances, both); and how many centers will be in the study.[4]

[4]For two grantees, both a "within" and a "between" design is needed. In Garden City, Kansas, one middle school received a 21st CCLC grant after being identified as a comparison school. In Lancaster, California, two school districts are part of the same 21st CCLC grant. One district has a "within" design and the other has a "within" design.

TABLE III.4

ADDITIONAL CHARACTERISTICS OF MIDDLE SCHOOL GRANTEES IN THE SAMPLE

Region	Urban or Rural	State	City	Award Amount	Type of Comparison Group Design	Number of Centers in Study
Northeast	Rural	NY	Sandy Creek	$84,800	Within	1
Northeast	Rural	VT	Barre	$200,000	Within	1
Northeast	Urban	MA	Boston	$305,977	Between	3
Northeast	Urban	MA	Worcester	$1,221,799	Between	1
Northeast	Urban	NY	New York	$957,413	Between	2
East	Rural	TN	Mohawk	$100,000	Between	1
East	Rural	VA	Machipongo	$722,595	Within	1
East	Urban	NJ	Union City	$200,000	Between	1
East	Urban	PA	Philadelphia	$540,192	Between	1
Southeast	Rural	NC	Marshall	$475,725	Within	1
Southeast	Rural	SC	Batesburg-Leesville	$430,729	Within	1
Southeast	Urban	AL	Montgomery	$599,704	Between	3
Southeast	Urban	SC	Greenville	$53,050	Between	1
North Central	Rural	MO	Mt. Vernon	$122,120	Within	1
North Central	Rural	WI	Franklin	$200,000	Within	1
North Central	Urban	MI	Grand Rapids	$620,613	Within	4
North Central	Urban	MI	Kalamazoo	$2,300,000	Within	3
North Central	Urban	MO	St. Louis	$592,958	Within	2
Midsouth	Rural	AZ	Coolidge	$159,241	Within	2
Midsouth	Rural	AZ	White River	$414,289	Within	1
Midsouth	Urban	AZ	Glendale	$163,875	Between	1
Midsouth	Urban	TX	Austin	$558,298	Between	1
Midwest	Rural	CO	Alamosa	$454,280	Within	4
Midwest	Rural	KS	Salina	$210,883	Between	1
Midwest	Urban	CO	Aurora	$171,749	Between	1
Midwest	Urban	KS	Garden City	$1,083,046	Both	3
Northwest	Rural	OR	Eugene	$1,377,613	Within	2
Northwest	Rural	OR	Sandy	$372,379	Within	3
Northwest	Rural	WA	Long Beach	$174,313	Within	2
Northwest	Urban	OR	Salem	$200,000	Between	3
Northwest	Urban	WA	Spokane	$716,618	Between	3
Southwest	Rural	CA	Alturus	$362,445	Within	3
Southwest	Rural	CA	Lancaster	$444,483	Both	2
Southwest	Urban	CA	Alameda	$300,000	Between	1
Southwest	Urban	NV	Reno	$1,149,433	Between	1

Source: U.S. Department of Education database on 21st Century Community Learning Centers Grantees.

Liaisons contacted directors of sampled programs to let them know they had been selected to be in the national evaluation and to explain that their participation was required under the terms of their grant. The liaison worked with each sampled grantee to identify issues and draft an agreement describing how the evaluation would be conducted. Common issues raised by grantee and district staff included the appropriate design for identifying a comparison group, the nature of the consent process that would be used for the evaluation, how data would be kept confidential, and how burden could be kept to a minimum.[5]

1. Power and Sample Size

The target sample size for the middle school design is 5,250 students, 150 students in each of 30 grantees (75 in the treatment group and 75 in the comparison group). For sampling purposes, participants are defined as students who attended a center operated by a sampled grantee for three days or more during a one-month period around October 2000 (the specific dates depended on when centers began operating and approached something near full enrollment). All students will be included in the sample if grantees serve fewer than 100 students.

In the second stage, comparison students whose characteristics most closely resemble those of program participants will be identified using the propensity score method developed by Rosenbaum and Rubin (1985 and 1983). This method uses statistical models of program

[5]Parents will provide active consent for the longitudinal study. Because some students will complete baseline questionnaires and not be selected for the longitudinal study, however, some districts opted for passive-consent procedures for the baseline questionnaire.

participation to identify comparison students who are most similar to program participants on a range of social and demographic variables.[6]

Table III.5 shows levels of statistical precision for the planned sample sizes and number of sites in the middle school design. The minimum detectable effect size for the 35-site sample is 9.6 percent if site variance is low and 11.5 percent if site variance is high.[7] Detectable effect sizes are about 15.5 percent for a 25 percent subgroup and 11.9 percent for a 50 percent subgroup (assuming low site variability). Detectable effect sizes are 37.8 percent for comparing effects of a group of 5 grantees to another group of 5 grantees, and 26.3 percent for comparing effects of 10 grantees to another group of 10 grantees (assuming low site variability). These large effect sizes indicate that the evaluation will need to be cautious on making statements about differences in impacts for various program models.

---

[6]Technical footnote: Under some technical conditions, the propensity scoring method generates comparison groups with properties equivalent to those of control groups created through random assignment. Operationalizing a propensity scoring model requires first estimating a statistical model of program participation. For two samples of students, n_P and n_{NP}, where P denotes the program group and NP denotes the nonprogram group, the model to be estimated is:

$$Pr(participation) = X\beta + e$$

where $Pr(participation)$ is set to one for participants (n_P) and zero for nonparticipants (n_{NP}), X is a set of social and demographic variables, β is a set of coefficients to be estimated, and ε is a random-error term. The estimated model is then used to generate predicted probabilities for program and nonprogram sample members, who are matched according to the closeness of the predicted probabilities.

[7]Technical footnote: Variances are affected by two factors that should be mentioned. (1) The "low" site variability assumption corresponds to 2.5 percent of the outcome variance being due to site-level factors. The "high" site variability assumption corresponds to 5 percent of the outcome variance being due to site-level factors. Generally, the degree of site-level variability is not known beforehand, and the assumptions here correspond to experiences from other evaluations. (2) The calculations shown in the table assume that variances of the impact estimates are calculated as if the sample were generated through random assignment. More sophisticated variance-estimation techniques that rely on bootstrapping will be used in practice. However, no rules of thumb exist for gauging the extent to which estimated variances will differ when bootstrapping is used.

TABLE III.5

DETECTABLE EFFECT SIZES FOR THE MIDDLE SCHOOL DESIGN

	Number of Sites	Total Sample	Single Site Effect Size	Pooled Effect Size (Low Var)	Pooled Effect Size (High Var)
Overall Sample	35	5,250	41.7%	9.6%	11.5%
Subgroups					
25 percent subgroup	35	1,313	83.4%	15.5%	16.7%
50 percent subgroup	35	2,625	59.0%	11.9%	13.5%
Strata					
5 sites in each stratum	5:5	750:750	(*)	37.8%	46.6%
10 sites in each stratum	10:10	1,500:1,500	(*)	26.3%	31.6%

Notes to Table:

Parameters for calculating detectable effect sizes:
Outcome standard deviation 50 percent
Overall response rate 81 percent
Participation Rate 90 percent
Regression R-Squared 20 percent
80 percent power, 95 percent confidence level, one-tailed test
Outcome variability due to site-specific factors is assumed to be 2.5 percent for "low var" and 5 percent for "high var."

IV. DESIGN OF THE IMPLEMENTATION STUDY

A major goal of the national evaluation is to learn how 21st CCLC programs are implemented and how they can be made more effective. Doing this requires analyzing (1) the context in which these programs operate, (2) participation by students and adults (including recruitment and dosage), (3) program content and delivery, (4) program structure and resources, and (5) collaborations and sustainability. This chapter lays out the implementation study design and describes its key data collection and analysis activities.

A. RESEARCH QUESTIONS FOR THE IMPLEMENTATION STUDY

The logic model in Chapter II (Figure II.1) provides the general framework for the implementation study. However, the model itself is too general to be used directly. Starting from the main research questions and subquestions derived from the logic model (noted in Chapter II), the design team identified 63 data elements, and appropriate respondents for collecting the elements, that will be useful to support the implementation study. Table IV.1 shows the data elements and respondents, listed by research question and the primary and secondary sources for each. This matrix will guide our data collection at all grantees and in all centers participating in the evaluation.

B. INFORMATION SOURCES

Each elementary and middle school site and center will be visited at least once across the 2000-2001 and 2001-2002 school years, with each visit typically lasting two to three days. While on site, visitors will be gathering information from three main sources: (1) interviews with individuals and small groups, (2) observations of program activities, and (3) reviews of relevant

TABLE IV.1

TOPICS, DATA ELEMENTS, AND SOURCES FOR IMPLEMENTATION STUDY DATA COLLECTION

	Project Director	Center Coordinator	Center Staff	Community Partners	District Rep	Principal	Teachers	Students	Parents	Other Adults	Observation	Doc Review
1. School/Community Characteristics												
a. How would you characterize the overall academic performance of this school? (For example, has the school been identified as exemplary or low performing by the district or state?)					X	X	X		O			O
b. How would you characterize the level of safety in the school?			O	O	X	X	X	O	O		O	
c. How would you characterize the level and nature of disciplinary problems at the school?					O	X	X	O	O		O	
d. How would you characterize the level of safety in the community?			O	O	X	O	X	O	O		O	
e. What other academic or youth development after-school programs or alternative activities are offered at this school or in the community to support students and adults?	X			O	O	X						
f. Are there other factors and issues about this school, district, or community that affect how and why the 21st CCLC program operates the way it does (for example, staff turnover, poor student attendance, poor parent/community support)?			O	O	O	X						
2. Program Goals												
a. What was the major impetus to apply for a 21st CCLC grant? (What major problems/issues would be resolved? What needs would be met? Why was this school selected to be part of the program?)		O			X	O						
b. What do you see as the primary goals of this program? Which goals do you see as most important?	X	X	O	O		O	O		O	O		
c. Did the program conduct a needs assessment when applying for a grant or when designing the program?	X			O	X	O						
d. How does the program measure progress toward achieving its goals?	X	X	O	O								
3. Organizational Structure												
a. How is this program staffed (full-time or part-time staff, use of volunteers, staff from other agencies), and what are the roles and responsibilities of the various staff and partners? What other (non-21st CCLC) responsibilities do staff have?	X	X		O								
b. Are centers able to identify, hire, and retain staff who meet program criteria? What are these criteria?	X	X	O									

NOTE: "X" identifies the primary respondent(s) for each question, and "O" identifies secondary respondent(s).

TABLE IV.1 *(continued)*

	Project Director	Center Coordinator	Center Staff	Community Partners	District Rep	Principal	Teachers	Students	Parents	Other Adults	Observation	Doc Review
c. What are the participant/staff ratios for various program activities? Were these ratios established intentionally or more the result of supply and demand? What are the challenges in trying to maintain target ratios?		X	X									
d. What training and professional development opportunities are provided or are available to staff? Do opportunities differ for different categories of staff?	X	X	X	O								
e. What is the level of staff attrition, and how does it affect the program? What factors lead to low/high attrition?	X	X	X									
f. How are decisions about the program made? Who has decision-making authority?	X	X	O	O	O							
g. What (if any) are the major sources of funding for this program other than the 21st CCLC grant? What proportion of the overall program budget do these sources contribute?	X	X			O							O
h. To what extent does the program use school resources (classroom space, computers, transportation, etc.) to implement the program? What additional resources are provided by other sources?		X	O	O		X	O				O	
i. Are these resources adequate? What other resources are needed? Have program activities ever been hindered by a lack of adequate equipment/supplies or other resources?		X	X									
4. Dimensions of Services												
a. What does the program schedule look like during a "typical" week (daily breakdown of time spent on different activities) and why was it developed this way? Has the program schedule changed over time and why?	O	X	O									O
b. To what extent are services/activities designed for a narrow age range versus broad or multiple age ranges (such as parents and children together)?		X	X								O	O
c. How do participants end up in particular services/activities? To what extent are services/activities voluntary versus mandatory? What are the rules/guidelines for participating in various services/activities?		X	X					O	O	O		
d. How much does program content vary within a semester/year?	O	X	O									O
e. To what extent are participants involved in solo versus small-group versus large-group activities?		X	X								X	O

NOTE: "X" identifies the primary respondent(s) for each question, and "O" identifies secondary respondent(s).

TABLE IV.1 *(continued)*

	Project Director	Center Coordinator	Center Staff	Community Partners	District Rep	Principal	Teachers	Students	Parents	Other Adults	Observation	Doc Review
f. To what extent are participants involved in self-directed versus adult-directed activities?		X	X								X	
g. How would you characterize the interaction among center staff, between staff and participants, between staff and parents, and among participants? What factors affect the quality and frequency of these interactions?		X	X					X	O	O	X	
h. What types of cognitive services are offered to students (such as homework assistance, general instruction, tutoring)?		X	O	O				X	O		O	
i. To what extent do these cognitive services focus on one or two content areas (such as reading/writing, math, science) versus all areas? What are those focused areas?		X	X								O	O
j. How frequently are these cognitive activities/services offered?		X	O	O				X				
k. What are the goals for these cognitive services/activities (that is, what would you like participants to get out of it)?		O	X				O					
l. What types of social, cultural, and physical activities are offered to students?		X	O	O				X	O		O	
m. How frequently are these social, cultural, and physical activities offered?		X	O	O				X				
n. What are the goals for these social, cultural, and physical activities (that is, what would you like participants to get out of them)?		O	X				O					
o. What types of services/activities are offered to adults (adult only or with children)? Why?		X	O						X	X	O	O
p. What are the goals of these various adult services/activities?		X	O						X	X		
q. Who is eligible for these adult services (adults in the community or students' parents), and how do people learn about the services?	X	X							X	X		
r. How frequently are these adult services and activities offered?		X	O						X	X		O
s. How popular are these adult activities? How many adults participate?		X							X	X		
5. Student Recruitment and Participation												
a. What are the eligibility requirements for participating in the program, and do they vary by site or by program activity?	O	X	O				O	O				
b. How does the program recruit students? Are certain students targeted for this program and, if so, which ones?		X										

NOTE: "X" identifies the primary respondent(s) for each question, and "O" identifies secondary respondent(s).

TABLE IV.1 *(continued)*

	Project Director	Center Coordinator	Center Staff	Community Partners	District Rep	Principal	Teachers	Students	Parents	Other Adults	Observation	Doc Review
c. What are the participation patterns (frequency of participation and extent of dropping out) for various program activities, and does participation vary by subgroup (such as gender, race, academic level)?		X	X					O				
d. What are the incentives/disincentives for participating? Who doesn't participate and why?	X	X	X					X	X	X		
e. How does the program ensure students' safety to, from, and within the program?	X	X	X									
f. What reasons do students give for participating in the program (for example, wanted to be with friends, thought it sounded fun, teacher/parents made them)?		O	O				O	X				
g. Why do parents want their children to participate in this program?									X			
h. What do students like/dislike about the program? What do they think they get out of the program?			O					X				
6. Collaboration												
a. Who are the program collaborators (includes community-based organizations, national organizations affiliates, schools), and how were they recruited? Had a relationship existed prior to this program?	X	X		X	O							O
b. Why do organizations and other partners participate in the program? What do they contribute/add to the program? What other organizations don't participate and why?	O	X		X								
c. Which collaborators were involved in initial planning of the program? How were they involved?	O	X		X	X							
d. How are collaborators involved in the management, operation, and activities of the after school program? Do they contribute staff, resources, run activities, etc.?	O	X	O	X								
e. How is the collaboration structured? How do partner organizations formalize agreements?	X	X		X		X						
f. How do members of the collaborative communicate? What topics/issues do collaborators communicate about? Do they discuss effective strategies, exchange information about students, share resources, conduct joint training, etc.?	X	X		X		X						
g. What are some of the challenges in recruiting and working with partners, and how have these challenges been addressed?	X	X	O	X	O							

NOTE: "X" identifies the primary respondent(s) for each question, and "O" identifies secondary respondent(s).

TABLE IV.1 *(continued)*

	Project Director	Center Coordinator	Center Staff	Community Partners	District Rep	Principal	Teachers	Students	Parents	Other Adults	Observation	Doc Review
h. What are the factors that make the collaboration strong/weak?	X	X		X								
7. Program/School Links												
a. How are academic services/activities linked to and affected by the regular school program or reform efforts?	O	X	X		O	O	X	O			O	
b. To what extent are academic services/activities seen as reinforcing the school curriculum (such as helping with homework) or supplementing it (presenting new/different material)?	O	X	O		O	O	X	O			O	
c. How and how often do program staff work with school staff (classroom teachers, special ed and remedial teachers, principal, district personnel) on issues related to the program curriculum and/or individual students? Who usually initiates that interaction?	O	X	X			O	X					
d. How do youth development services support, supplement, or fill gaps in the regular school program?	O	O	X			O	X	O				
e. What are the challenges to maintaining positive relationships with the school?	X	X		O								
8. Program Sustainability												
a. What have been the major challenges to implementing and operating the 21st CCLC program? How has the program responded? What has been successful? Unsuccessful?	X	X	O	O	O							
b. What changes would people like to see to the program content and delivery? In the organizational structure, resources, or collaborations?	X	X	O	O	O	O	O		O	O		
c. Have there been any barriers for students to participate in the after-school program (such as lack of transportation, distance from program) and how have you responded to them?		X	O	O				X	X			
d. Are there any plans in place to sustain the 21st CCLC program beyond the grant period? If so, what are they?	X	X		O	O	O						
e. What external supports has the program tapped for continuing the program beyond the grant period? What approaches have been successful? Unsuccessful?	X	X		O	O	O						

NOTE: "X" identifies the primary respondent(s) for each question, and "O" identifies secondary respondent(s).

While on site, visitors will be gathering information from three main sources: (1) interviews with individuals and small groups, (2) observations of program activities, and (3) reviews of relevant documents. At the conclusion of each visit, visitors will record descriptive information and complete a narrative summary and assessment of each center.

A major source of information will be interviews with key program personnel, participants, and others familiar with program implementation and operation. For each data element in Table IV.1, visitors will interview at least one primary respondent and several secondary respondents. Semistructured interview guides (see Appendix A) will structure the discussions, with some tailoring of interviews depending on local circumstances and data collection objectives. Respondents will include:

- ***Program Directors.*** Visitors will talk with project directors about program goals, organizational structure, collaborations, and sustainability.
- ***Center Coordinators.*** Visitors will talk with center coordinators about program goals, organizational structure, dimensions of service, recruitment and participation, collaboration, program/school links, and program sustainability. (Discussions will be adjusted accordingly in small programs where the program director also has this role.)
- ***Center Staff.*** During one-on-one or small-group interviews with center staff (whichever is more appropriate at a given center), visitors will discuss organizational structure, service dimensions, student participation patterns, and links between the program and the school.
- ***Community Partners.*** Visitors will talk with community partners one on one or in small groups about the nature of their collaboration: why they participate, which parts of the process they have been involved in, and the challenges associated with being partners.
- ***District Representatives.*** Visitors will talk with representatives of the school district to understand school and community characteristics, the goals behind the initial application, and the processes of identifying and working with program collaborators.
- ***School Principals.*** Visitors will talk with principals about school and community characteristics, program use of school resources, and links between the program and the school.
- ***Regular School Teachers.*** Visitors will talk with regular school teachers (as distinct from after-school program teachers) about the extent to which program staff and

classroom teachers communicate about students, curriculum, and after-school activities. Discussions will be one-on-one or small-group interviews.

- ***Middle School Program Participants.*** Visitors will talk in small groups with middle school students about their experiences with the program, especially about why they participate, what they actually do in the program, how they view their interactions with the staff, and what they like or dislike about the program.
- ***Parents of Participants.*** Visitors will talk with parents about why they want their children to take part in the program and about the barriers for students to participate fully.
- ***Adult Participants.*** For centers with adult programs, visitors will talk with participants about the services they wanted, the types of services offered, and their interactions with program staff.

Observations. Visitors also will observe program activities at each center. Observations will be informal and discrete, designed to allow visitors to develop an understanding of the program by watching it in operation. The goal will be to develop an understanding of service delivery, especially of such elements as the types of services offered, the prevailing pedagogy, the structure of activities (individual versus small-group versus large-group), the locus of direction (adult-directed versus participant-directed), the nature of staff/participant interactions, and program resources.

Document Review. Visitors will review documents on site for information that is historical, detailed, or readily available through a database. Since records kept by the program may vary in quality and thoroughness, visitors will need to make on-site judgments about their usefulness for the evaluation. Weekly activity schedules, recruitment and advertising materials, written needs assessments, school testing results, and summaries of program evaluations are some of the documents visitors will plan to review. Data elements guiding the document reviews are listed in the "Center Profile Form" in Appendix A.

C. INSTRUMENTS FOR GATHERING IMPLEMENTATION INFORMATION

At the conclusion of the site visit, visitors will chronicle detailed descriptive information and synthesize overall impressions about the project and the context in which it operates. To facilitate this and record the information for later analysis across sites, visitors will provide a narrative summary of the program using a (1) *write-up guide,* complete an assessment of the program using the (2) *grantee assessment form,* and record descriptive information about each center on a (3) *center profile form,* and a (4) *center assessment form.* These four instruments are shown in Appendix A. Data from these instruments will be used in both the impact study and the implementation study to categorize programs by particular program characteristics related to each research question.

Write-Up Guide. The write-up guide is a narrative instrument designed to organize the site observations and findings into an internal study document that will cover a single grantee and will facilitate cross-site analyses. The heart of the narrative will describe the site using the research questions and write-up prompts designed to illuminate each question. Visitors will use all relevant information, attribute key information, use clarifying examples and illustrative quotations when appropriate, and note significant discrepancies. The completed document will contain grantee and program information that is common within all centers and will clearly identify important and interesting differences among centers. The resulting narrative is intended to give a clear description of important aspects of the grantee relative to the research questions. The narratives as a group will become the basis for summarizing, analyzing, and reporting.

Grantee Assessment Form. Site visitors will use the grantee assessment form to describe the grantee on key dimensions, such as the degree to which it has difficulty finding, hiring, and retaining key staff, and the degree of similarity between its centers (if two or more are in our

study). The form features closed-ended, categorical items on which visitors report their informed judgments about grantee operations.

Center Assessment Form. The center assessment form is similar to the grantee assessment form in that it captures visitors' judgments, but the center assessment form is longer and more detailed. It collects information on more issues, including the degree of emphasis placed on various program objectives, the extent of involvement by collaborating partners, and the extent to which various issues pose a challenge for center operations.

Center Profile Form. Site visitors will use the center profile form to record information about each center participating in the study (filling out one form per center). The items on the form address issues such as staffing (for example, number of staff, professional qualifications), program policies (for example, recruitment and student eligibility), services offered (for example, schedule of activities, transportation to or from the program), and collaborations (for example, types of partners and the specific roles they play in program operations). The information will be obtained directly from the center coordinator or project director, or they will verify the information obtained from other sources.

D. ANALYSIS METHODS

The implementation analysis approach will vary depending on the stage of evaluation. In the first stage, during the 2000-01 school year, the focus will be on understanding and describing how the 21st CCLC program is operating in the sites that are participating in the evaluation. Data from the site visits will be used to address the eight major research questions outlined in Chapter II. During the first phase, we will analyze and report separately on implementation at the middle school and elementary school level. In the second stage of the study, during the 2001-02 school year, the focus will shift to exploring program characteristics and features that may be linked with program impacts. This shift is possible because impact data will be available from student

follow-up surveys, teacher surveys, parent surveys, students' school records, and, in the case of elementary school students, standardized test scores.

During the second phase of the evaluation, in the analysis of program characteristics and practices linked with program impacts, we will explore combining grantees across the elementary and middle school levels. For example, if some of the elementary school grantees and some of the middle school grantees show significant impacts on students' reading abilities, the analysis may want to consider grouping these subsets of grantees together, to see whether they collectively differ in some ways from the elementary and middle school grantees whose students show no impacts on reading ability.

1. Storing and Retrieving Data

Site visit notes will be entered into ATLAS/ti, a specialized qualitative data storage-and-retrieval software package, to increase the efficiency of the implementation analysis. The power of the software comes from its ability to store thematic codes for various sections of each write-up and sort, extract, and connect specific sections from all write-ups. The research team will assign codes to the research questions and subquestions and to site characteristics (for example, cohort, urban/rural, middle/elementary, site visitor, year visited), and use the codes in the analysis to retrieve text pertaining to particular research questions for particular types of sites. For example, if all site visits result in a total of 100 pages of narrative on collaboration, researchers can retrieve those pages and place them in a single file that would become the basis for an analysis of collaboration. In addition, searches could be restricted by particular site characteristics. A restricted search, for example, might include all narrative on collaboration in first-cohort elementary sites.

2. Analyzing the Data

Analyzing the mainly qualitative data from the implementation study will be an iterative process led by a core team of experienced researchers. Analysis will begin informally as soon as researchers begin site visits. After visits are complete and narratives and accompanying forms have been submitted, the analytical process will become more systematic. The core implementation research team, consisting of five to six senior researchers, will divide into overlapping subgroups, each assigned responsibility for addressing one or more of the eight major research questions. These subgroups will develop specialized knowledge of program characteristics and practices in certain areas. For example, one group of researchers will have more expertise in recruitment and participation, while another will have more in service delivery.

The core team will engage in a focused effort to explore similarities and differences across sites. They will categorize grantees on key dimensions and array them on various continua, using matrices and other analytic tools. They will consider not only what is typical or common among grantees, but also what is atypical or uncommon. They will look for relationships between key characteristics and practices. They will develop, test, and refine explanatory propositions, and revisit and reexamine the data until they are confident in their findings and conclusions.

Fundamentally, the process described here relies on subjective judgments. This is inherent in qualitative analysis. Nonetheless, it is important to minimize the degree to which a single person's impressions will influence the findings. One strategy for accomplishing this is to use small teams of experienced researchers, as mentioned above, to do the cross-site analysis. A second strategy to reduce individual subjectivity will be implemented before cross-site analysis begins, during the data collection phase. Two-person teams will conduct 12 of the early site visits (approximately 25 percent of all yearly visits) to strengthen common understandings and

perspectives on program features. For each grantee, visitor pairs will reach consensus on the write-up and its accompanying assessment and profile forms, and another researcher will review rating forms to ensure consistency within and across sites.

An example helps to explain the qualitative analysis process. In this case, consider analyzing the question, What are the key challenges to implementing and sustaining 21st CCLC programs, and how are these challenges addressed?

The analysis team first would review center and grantee assessment forms, which contain discrete and easy-to-analyze data, much like survey data. Item 20 on the center assessment form, for example, records visitors' judgments on the degree to which 10 distinct issues have been a challenge at each center. These issues include recruiting instructional staff, retaining students, gaining access to school facilities, and getting support and cooperation from regular school day teachers. The three response options are "little or no challenge," "modest challenge," and "major challenge." Item 3 on the grantee assessment form asks about the degree of difficulty that grantees have had in "finding, hiring, and retaining program staff such as center coordinators," with response options of "no difficulty," "some difficulty," and "a great deal of difficulty."

Data from these items would be tabulated to provide a distribution of the seriousness of various challenges across sites. Other data on grantees and centers could then be explored in conjunction with these data, for a better understanding of how the challenges are related to other program or contextual factors or characteristics. For example, simple cross-classification tables could be compiled to assess whether grantees that have had a great deal of difficulty finding, hiring, or retaining senior staff are more likely to be first-, second-, or third-cohort grantees, or more often located in urban than rural areas. Other possible relationships would be explored on the basis of hypotheses formed during and after field work.

After examining closed-ended data from the grantee and center assessment forms, the team would turn to the open-ended, narrative information contained in the write-up forms. This would have three major purposes. First, the write-ups may contain information on important challenges that were not listed in the assessment forms. Item 8.1 in the write-up form, for example, addresses the "top 3" challenges that the 21st CCLC programs faced, and item 8.2 addresses "other/minor" challenges. The team would analyze this information to classify challenges as relating to implementation, operation, or sustainability, and determine how often these challenges were noted across sites, how serious these challenges were perceived to be, and whether experiencing particular challenges was associated with program characteristics.

The second reason for looking to the write-ups is that they contain detailed information that may enable the team to describe the nature of program challenges better. The information might include why various challenges have arisen, whether various challenges are new or long-standing, and whether challenges have been becoming easier or more difficult.

Third, the write-ups are the sole source of information on how challenges are being addressed. Information in the write-ups will be useful for exploring whether different strategies had been used for particular challenges with differing levels of success, as well as why and under which circumstances various strategies have been effective in responding to certain challenges. Making these kinds of determinations requires sorting and categorizing programs based on subjective judgments. The team will form, test, and revise hypotheses to arrive at findings that are well supported by the cumulative evidence.

V. DESIGN OF THE IMPACT STUDY

The focus of the impact study will be on whether after-school programs improve student outcomes and whether specific program strategies and practices are more effective than others. This chapter discusses the research questions to be addressed in the impact study, the instruments that will be used to collect the data, and the methods that will be used to measure impacts. The chapter also discusses issues in measuring dosage effects.

A. RESEARCH QUESTIONS

The logic model laid out in Chapter II suggests that the impact analysis should address four key research questions:

1. Do centers improve academic and other in-school outcomes for students, such as academic achievement, on-time promotion, attendance, and classroom behavior?
2. Do centers improve student out-of-school behaviors, such as reducing incidents of violence, use of drugs, smoking, and contact with gangs?
3. Do students of different ages, backgrounds, and levels of academic and English proficiency experience different impacts?
4. Which program practices and models have the greatest effects? Do students receiving larger doses of program services show larger effects?

The first two questions can be explored by comparing outcomes of students in treatment groups with outcomes of students in control or comparison groups. The third question can be addressed by looking at impacts on students of a particular race, sex, age, or other characteristic. The fourth question can be addressed by using information from the center and grantee assessment forms (described in Chapter IV) and program participation data to assess impacts by characteristics of centers or grantees—such as the stability of staff, the involvement of the principal, and the frequency with which partners provide staff to the program—and by the size of a program "dose."

B. INSTRUMENTS TO SUPPORT THE IMPACT ANALYSIS

After-school programs may affect students in a variety of ways, and the effects may be seen through many perspectives. For example, students may put more effort into homework as a result of working with a mentor or tutor in an after-school program, and this effort may be observable to teachers, parents, and students themselves. Other outcomes may be observed only from one or the other perspective. For example, a teacher may be aware that a student is behaving better in class, but the student may be unaware of it, and parents may perceive that their child is safer in an after-school program even if their child does not.

Responding to the need to capture a variety of outcomes that may depend on perspectives, the evaluation will use a broad data collection approach that focuses attention on students but also gathers information from teachers and parents (see Table V.1). Questionnaires will be administered to students in grades three through eight at baseline (fall 2000) and two follow-up points (spring 2001 and 2002). Appendix B contains the full set of instruments for the study. The questionnaires differ for elementary school students and middle school students, to reflect their differences in maturity and developmental levels. A standardized reading test derived from the Stanford Achievement Test version 9 (SAT-9) also will be administered to students in kindergarten through grade five at the same times that questionnaires are administered (fall 2000, spring 2001, and spring 2002). In addition, information from school records will be collected, and after-school programs will provide student participation and attendance information.

Questionnaires will be administered to parents of elementary school students at the same baseline and two follow-up points, and to parents of middle school students at the two follow-up

TABLE V.1

DATA COLLECTION INSTRUMENTS AND SCHEDULE

Instrument	Time Frame
Elementary School Design*	
Student	
Baseline questionnaire	Fall 2000
Follow-up questionnaire	Spring 2001, Spring 2002
School records	Spring 2001, Spring 2002
Participation records	Spring 2001, Spring 2002
Reading standardized test	11/1/00-12/31/00, 4/1/00-6/30/00, 4/1/00-6/30/00, Fall 2000, Spring 2001, Spring 2002/00
Parent	
Baseline questionnaire	11/15/00-12/31/00, Fall 2000
Follow-up questionnaire	Spring 2001, Spring 2002
Teacher	
Follow-up questionnaire	Spring 2001, Spring 2002
Middle School Design	
Student	
Baseline questionnaire	Fall 2000
Follow-up questionnaire	Spring 2001, Spring 2002
School records	Spring 2001, Spring 2002
Participation records	Spring 2001, Spring 2002
Parent	
Follow-up questionnaire	Spring 2001, Spring 2002
Teacher	
Follow-up questionnaire	Spring 2001, Spring 2002

NOTE: For the second cohort of elementary school grantees expected to begin random assignment in fall 2001, the data schedule will be advanced by a year.

points.[1] Teacher surveys will be conducted at the two follow-up points with a core-subject teacher of each student in the sample. Data will be collected from the main classroom teacher for the elementary school sample and from the English/language arts teachers of the middle school sample.

The content of the various questionnaires is described in Table V.2. Sections of the table correspond to the main logic model topic, with the first column indicating the subtopic within the main topic. Circles indicate whether the subtopic is addressed in the various questionnaires. The same tables are reproduced in Appendix B, with numbers of the questions indicated in place of circles, to guide readers to specific items. The last column indicates whether complementary information is being collected from site visits.

C. METHODS FOR MEASURING IMPACTS

The two designs used for estimating impacts for elementary and middle school students share some elements but also differ in important ways. For both designs, differences between average group outcomes are estimators of program impacts. For example, the difference between the average grades of the treatment group and the average grades of the control or comparison group measures the program's impact on grades. A simple *t*-test of the difference in average grades enables the evaluation to assess the likelihood that the difference was due to chance.

For random-assignment designs, the treatment and control groups are similar at baseline by construction (though tests will be conducted to verify the similarity). However, for comparison group designs, the analysis needs to explore carefully whether comparison groups are

[1]Administering a parent baseline questionnaire is not feasible for middle school parents, because parents of comparison group students will not be known until several months after the treatment group parents are known.

TABLE V.2

INSTRUMENT TOPICS AND SURVEY ITEM NUMBERS

Topic	Subtopic	Questionnaires: Student Baseline[a]	Student Followup[a]	Parent[b]	Teacher[c]	Records	Site Visits
CONTEXT							
School Characteristics	a) School, classroom, teacher, student (peer) characteristics	●	●	●	●		
	b) School safety	●	●	●	●		
	c) Academic support structure	●	●	●			
	d) Climate, academic expectations	●	●	●	●		
	e) Community safety	●	●	●			
	f) School/community involvement						●
Out-of-School Time	a) Location	●	●	●	●		
	b) Supervision	●	●	●			
	c) Activities	●	●	●	●		
PROGRAM IMPLEMENTATION							
Participation	a) Recruitment/ retention					●	●
	b) Dosage					●	●
Program Content	a) Safe environment		●		●		
	b) Academic activities	●	●	●	●		
	c) Youth development (social, cultural, and recreational activities)	●	●	●	●		
	d) Links with school day						●

TABLE V.2 (*continued*)

Topic	Subtopic	Questionnaires: Student Baseline[a]	Questionnaires: Student Followup[a]	Questionnaires: Parent[b]	Questionnaires: Teacher[c]	Records	Site Visits
Program Structure and Resources	a) Goals						
	b) Organization			●	●		
	c) Functioning	●	●	●	●		
	d) Staffing/training	●					●
Collaborations	a) School				●		●
	b) Community organizations						●
Program Sustainability							●
INTERMEDIATE EFFECTS							
Academic and Cognitive	a) Improved homework completion	●	●	●	●		
	b) Improved study skills	●	●		●		
	c) More reading, less TV	●	●	●			
	d) Improved effort				●		
	e) Better attendance	●	●		●	●	
	f) Improved classroom behavior	●	●		●	●	
	g) Improved attitude toward school and reading	●	●	●	●		
Social and Emotional	a) Increased personal responsibility	●	●	●	●	●	
	b) Improved peer interactions	●	●	●	●		
	c) Higher aspirations	●	●	●			
	d) Better problem-solving	●					
	e) Increased knowledge/ respect for diversity			●			
	f) Increased feeling of safety	●	●	●			

TABLE V.2 (*continued*)

Topic	Subtopic	Questionnaires: Student Baseline[a]	Questionnaires: Student Followup[a]	Questionnaires: Parent[b]	Questionnaires: Teacher[c]	Records	Site Visits
LONG-TERM EFFECTS							
Academic and Cognitive	a) Higher grades	●		●	●	●	
	b) Higher test scores				●	●	
	c) On-time promotion					●	
	d) Return to regular track					●	
Behavior	a) Reduced fighting, stealing, vandalism [#]	●	●	●	●		
	b) Reduced use of alcohol, tobacco, other drugs [#]	●	●	●			
MEDIATING FACTORS							
Mediating Factors	a) Individual and family characteristics	●		●	●	●	
	b) Parent-child relationship	●	●	●			
	c) Community involvement	●	●	●			
	d) Parent involvement in schools			●			

Notes: Some items are included only in elementary school questionnaires or middle school questionnaires. The table presented in Appendix B indicates the numbers of the items in the various questionnaires.

systematically different from the treatment groups and develop analytic adjustments if the evidence suggests that the groups differ.

1. General Impacts

A straightforward approach for estimating program impacts is to use regression models in which an outcome is related to (1) an indicator of whether a sample member was a treatment group member, and (2) other explanatory variables (constructed using baseline data items). Regression models are easy to estimate and yield more precise impact estimates than simple differences-in-means estimators. An example of a regression model for estimating impacts of a 21st CCLC program is given by:

$$[1] \qquad Y = \alpha + P\beta + X\delta + \varepsilon$$

where α, β, and δ are coefficients to be estimated; P is whether a student was assigned to the treatment group, and X represents explanatory variables assumed to affect the outcome Y. The estimated value of the coefficient β is an estimate of the program's impact (that is, the difference in means between the treatment group and the control group after adjusting for other characteristics). Ordinary-least-squares estimation is used for estimating impacts when the dependent variable is continuous (such as a test score). Other estimation methods, such as discrete-choice modeling, are used when the outcome of interest is an indicator variable (such as whether students engaged in risky behaviors).

2. Subgroup Impacts

The second type of impact analysis looks at impacts for various groups of students. Subgroup analysis is particularly useful for identifying the types of students who benefit most from program participation. Some types of students might be affected more by program

participation, and programs may want to consider targeting these children to achieve the greatest impacts with fixed resources.[2] Another use of subgroup analysis is to classify students into groups depending on the program they attend. Looking at impacts by type of program involves calculating site-level impact estimates (adjusted for student characteristics) and examining whether there are any patterns between the site-level impacts and site characteristics. For example, the analysis might find that programs with a strong academic focus had larger impacts on academic outcomes.

Estimating subgroup impacts entails adding interaction terms between the treatment indicator and indicators of whether sample members are in the relevant subgroup under consideration. An example of a model for subgroup analysis is:

$$[2] \qquad Y = \alpha + P\beta + P \cdot W\lambda + X\delta + \varepsilon$$

where the terms are defined as in equation [1] and W denotes membership in a particular group, such as urban students. For example, if we were interested in the differential impact of programs on students with low achievement, we might construct a binary variable equaling one if a student has a test score below a cutoff value and zero otherwise. The impact of the program on students with low achievement is then estimated by $\beta + \lambda$. An estimate of λ that is statistically significant and positive is evidence that the impact of the program is larger for low-achieving students. The particular subgroups that can be created using the data items in Table V.2 are discussed in the last section of the chapter.

---

[2]Myers and Moore (2000) explore this issue using findings from the National Evaluation of the Upward Bound program.

Another type of subgroup analysis explores the impact of particular program characteristics on student outcomes. Because site-level sample sizes are modest, a useful approach for determining the impact of particular program characteristics is to combine sites with similar characteristics and assess whether groups of sites ("pools") have larger impacts. For example, pools could be created based on the strength of the academic focus at each site, and the impact estimates of the various academic strength levels compared. The analysis may find that sites with the strongest academic focus have a larger impact on academic outcomes than sites with a weaker one. Other analyses could combine several factors—such as pooling sites with high academic focus and high parent involvement—assuming there are enough programs in the pools.

Before conducting this analysis, we will need to decide whether the center or grantee will be used as the unit of analysis. This will depend upon the variability that exists among centers within a grantee. If there is little variation, then we can conduct the analysis at the grantee level. If the variability between centers within a grantee is high, then the analysis is appropriately conducted at the center level, since grantee-level variables will not accurately characterize each center.

The grantee assessment form is a useful tool for assessing whether the centers within a grantee are similar. For example, grantees can be coded as having low intra-grantee variability if centers are rated "very similar" or "somewhat similar" on at least 6 of 10 items designed to assess the similarity of centers within a grantee.

If centers are the unit of analysis, center characteristics can be obtained from the center assessment form. For example, center characteristics could be based on

- The major objective of the center
- The extent to which other local programs provide similar services
- The principal's degree of involvement with the program
- Regular school teachers' degree of involvement with the program

- Students' degree of engagement in center activities
- The program's number of collaborating partners

Hierarchical linear models (HLM) also can be estimated to assess relationships between impacts and program characteristics. An example of a HLM model is:

[3] $$Y = \theta_1 + P\theta_2 + X\theta_3 + Z\theta_4 + PZ\theta_5 + (\mu_0 + \mu_1 P + \varepsilon)$$

where *Y, P,* and *X* are as defined for equation [1], *Z* represents a program characteristic, θ_j (j = 1, 2, 3, 4, 5) are the parameters to be estimated, and μ_0 and μ_1 are random error terms. For example, suppose *Z* represents a variable for whether programs have trouble retaining staff. The impact for students in programs that are *not* having trouble retaining staff is estimated by θ_2. The impact for students in programs that *do* have trouble retaining staff is estimated by $\theta_2 + \theta_5$.

An important center or grantee-level characteristic is the extent to which control or comparison group students are engaged in other after-school programs. Student questionnaires contain items designed to understand this context better. The questionnaire asks students where they go after school and who is supervising them Monday through Friday of a typical week. The evaluation can use these items to create categories for the setting and for the type of supervision. For example, categories for after-school settings can include:

- Own home
- Home of someone else
- School or other program
- Somewhere to "hang out"
- Mixed location (not in any one location for at least three days)

As an initial construct, students will be included in a category if they are in the supervision setting at least three days a week. Similarly, supervision categories can include:

- ***Self-Care.*** This category includes students who are either alone after school or with friends who are about the same age.
- ***Parental Care.*** This category includes care by the student's mother or father.
- ***Non-parental Adult Care.*** Included in this category is care by any adult who is not the child's parent.
- ***Sibling Care.*** Included in this category is care by either an older or younger sibling.
- ***Mixed Care.*** This category will include students who do not receive care in any one category for at least three days of the week.

To investigate whether impacts affected by the degree of exposure to other after-school activities, the evaluation can categorize grantees or centers based on the average percentage of control or comparison group students who participate in other after-school programs. Categories for high, medium, and low after-school program activity can be created and impact estimates calculated for each category. If participation in after-school activity improves student outcomes, sites with high levels of other after-school programs are likely to have lower impacts.

3. Adjusting for Crossover

Two additional considerations affect measured impacts. The first is that some treatment group students may not participate in the program, implying that the estimated impact underestimates the impact of participating in the program. Adjustments for nonparticipation are described in this section. The second consideration is that some students will receive a larger "dose" of program services and thus might experience larger impacts from the program. Estimating impacts that vary by dosage is appealing, and the discussion in the next section describes several methods for doing so.

Impact estimates described above will underestimate the impact of participation in the program when some treatment group members drop out of the program. Impact estimates also are underestimated if some control group members "cross over" and participate in the 21st CCLC program.

Two simple adjustments can be used to measure the impact of program participation. The first involves dividing the overall impact estimate by the proportion of the treatment group that participated in the program. This adjustment is based on the assumption that the program has no impact on students who do not participate, which is innocuous in most evaluation settings.

The second adjustment can be applied if some "crossover" of control group members into program services occurs. In this case, the overall impact estimate is divided by the difference between the proportion of the treatment group who participated and the proportion of controls who "crossed over" and received the treatment. This adjustment is based on the assumption that the program has no impact on treatment group members who do not participate and that the program has an average impact on the control group members who do participate.

A more complex method for estimating the impact of program participation involves computing complier average causal effects (CACE). The CACE estimator provides an estimate of the impact of the program on those who decide to participate in the program after they are offered the opportunity (Angrist et al. 1996). To calculate the CACE estimator, a statistical model is estimated with two equations: (1) the relationship between being randomly assigned to participate in the program and participating in the program, and (2) the relationship between after-school program participation and outcomes. These relationships can be expressed as:

[4] $$P = \alpha_0 + \alpha_1 T + \alpha_2 X + \varepsilon$$

[5] $$Y = \beta_0 + \beta_1 P + \beta_2 X + \mu$$

where T equals one if the student had been randomly assigned to the treatment group and zero otherwise, X is a vector of student and family characteristics, P represents whether the student participated in the 21st CCLC program, Y represents the outcome of interest, and ε and μ are error terms. The coefficient β_1 can be interpreted as the impact of the program for compliers—students who participated in the program after being assigned to the treatment group.[3]

4. Adjusting for Different Doses

Another variant of dosage analysis involves calculating impacts for subgroups defined by levels of program participation. Technical issues need to be considered carefully in this approach. Previous research that explored dosage estimates did not control for possible biases that may arise because students who receive larger doses of program services may differ systematically from those receiving smaller ones. For example, Baker and Witt (1996) found that students who participated in more after-school program activities had higher grades than students who participated in fewer activities. However, it is difficult to know whether the higher numbers of activities were responsible for the higher grades or if students who participated in more activities were naturally higher achievers, which seems at least plausible.

One approach for reducing bias is to create a comparison group of students who are similar to students who receive larger doses, in terms of their basic demographic characteristics. Using a propensity-score method, for example, the analysis would (1) estimate a model of whether participants attend frequently, (2) use the estimated model to compute propensity scores and

[3]Technical footnote: the CACE is estimated using a two-stage least-squares approach. The first stage involves regressing program participation on the treatment indicator and the explanatory variables. The estimated coefficients are then used to calculate predicted program participation. The second stage involves regressing outcomes on predicted program participation (from the first stage) and explanatory variables.

match comparison group members with frequent-attender treatment group members, and (3) estimate program impacts as the difference between frequent-attender treatment group members and matched comparison group members.

Another approach for estimating the effect of varying after-school program doses is to group grantees or centers by the average dosage among its students, forming high-, medium-, and low-dosage pools. Program attendance records and Annual Performance Reports (APRs) will allow the evaluation to calculate after-school program dosage as a means of classifying centers. Student attendance records provide information on the number of days the student attends the program, and the APRs provide the number of hours each program is in operation each day. This will allow the evaluation to calculate the number of hours students spend in the after-school program. Classifying centers into high-, medium-, and low-dosage pools will provide suggestive evidence about the effects of dosage at the center level.

5. Analysis Plan

The last step in planning the analysis is to combine the instrument items (Table V.2) with the measurement methods described in the previous section. From an impact-measurement perspective, the data items have three uses: (1) as outcome variables, (2) as explanatory variables in regression models, or (3) as variables to define subgroups. Items in the latter two categories also support hierarchical linear modeling.

Tables V.3 through V.5 lay out a scheme in which the items are categorized into one of the three groups. Table V.3 indicates student intermediate and long-term outcomes that will be analyzed. For example, the impact on homework completion, the first intermediate outcome shown in Table V.3, would be estimated using a regression model that includes student age, race, ethnicity, sex, mother's and father's education, and previous-year absences, among other

TABLE V.3

PLAN FOR ESTIMATING IMPACTS ON STUDENT OUTCOMES

Outcomes	Explanatory Variables	Subgroups
ELEMENTARY AND MIDDLE SCHOOL STUDENTS: INTERMEDIATE OUTCOMES		
Homework completion Attendance Participation in band, drama, or other school clubs and activities Disciplinary problems Academic engagement Participation in academic activities Participation in music, art, or dance Participation in clubs Perception of safety after school Television watching Time spent reading for fun	Age Race Ethnicity Sex Mother's and father's education Previous-year absences Previous-year suspensions Self-reported description of grades Mother's and father's age Mother's and father's employment status Family income Family receipt of TANF or food stamps Language spoken in home	Age Race Ethnicity Sex Mother's education Previous-year absences Previous-year suspensions Grades Mother's employment status High or low family income Receipt of TANF or food stamps Language spoken in home Urban/rural location Single-parent family
ELEMENTARY AND MIDDLE SCHOOL STUDENTS: LONG-TERM OUTCOMES		
Grades Standardized test scores Advancement and promotion Delinquent behaviors Positive behaviors	See above	See above
MIDDLE SCHOOL STUDENTS: ADDITIONAL INTERMEDIATE OUTCOMES		
Frequency of skipping school, skipping class, and getting detention	See above	See above
MIDDLE SCHOOL STUDENTS: ADDITIONAL LONG-TERM OUTCOMES		
Delinquent behaviors, frequency of getting arrested, using or selling drugs, and shoplifting	See above	See above

TABLE V.4

PLAN FOR ESTIMATING IMPACTS ON TEACHER-REPORTED OUTCOMES

Teacher-Reported Outcomes	Explanatory Variables for Regression Model	Subgroups
TEACHER OUTCOMES: INTERMEDIATE OUTCOMES		
Teachers' perceptions of school safety Teachers' perception of school quality (students placing high priority on learning, school's expectation for homework completion, willingness of teachers to help students before or after school)	Teacher's age Teacher's race Teacher's ethnicity Teacher's sex Teacher's education Total years of teaching experience	Teacher's race Teacher's education Teacher's experience Urban or rural location Percent free lunch
ELEMENTARY AND MIDDLE SCHOOL STUDENT OUTCOMES: INTERMEDIATE OUTCOMES		
Teacher's rating of student's level of effort Absences from class Disruptive behavior Homework completion	Age Race Ethnicity Sex Mother's and father's education Previous-year absences Previous-year suspensions Child's description of grades Mother's and father's age Mother's and father's employment status Family income Family receipt of TANF or food stamps Language spoken in home Teacher's race and ethnicity Teacher's sex	Age Race Ethnicity Sex Mother's education Previous-year absences Previous-year suspensions Grades Mother's employment status High or low family income Welfare receipt Language spoken in home Urban/rural location Single-parent family
ELEMENTARY AND MIDDLE SCHOOL STUDENT OUTCOMES: LONG-TERM OUTCOMES		
Teacher's rating of student achievement level Course taking (regular, remedial, gifted, etc.)	See above	See above

TABLE V.5

PLAN FOR ESTIMATING IMPACTS ON PARENT-REPORTED OUTCOMES

Parent-Reported Outcomes	Explanatory Variables	Subgroups
ELEMENTARY AND MIDDLE SCHOOL STUDENTS: INTERMEDIATE OUTCOMES		
Whether child did homework after school each day Child's homework completion Concern about getting into trouble after school Concern about safety after school Child's participation in clubs and activities Television watching Parent's perception of child well-being	Age Race Ethnicity Sex Mother's and father's education Previous-year suspensions Child's description of grades Mother's and father's age Mother's and father's employment status Family income Family receipt of welfare Language spoken in home	Age Race Mother's education High or low family income Receipt of TANF or food stamps Urban/rural location Single-parent family
ELEMENTARY AND MIDDLE SCHOOL STUDENTS: LONG-TERM OUTCOMES		
Aspirations for child's education Parent's perception of school	See above	See above

variables. In addition, impacts will be measured within subgroups such as younger and older students, boys and girls, racial and ethnic groups, and high- and low-absence students.

Tables V.4 and V.5 show the analysis plans based on data collected from teachers and from parents. Teachers will provide information on student performance and behavior, as well as their own perceptions of the school environment. Teacher-reported student outcomes that will be analyzed include the teacher's rating of each student's level of effort, the teacher's rating of each student's level of achievement, and the frequency with which the student is absent or late for class. Teacher outcomes that will be analyzed include the teacher's perception of school safety and school quality. Parent-reported outcomes include aspirations for the child's education, concern about the child's safety after school, the child's homework completion, and grading of the school.

REFERENCES

Baker, Dwayne and Peter Witt. "Evaluation of the Impact of Two After-School Recreation Programs." *Journal of Park and Recreation Administration*, vol. 14, no. 3, 1996, p. 23-44.

Brooks, Pauline, Cynthia Mojica, and Robert Land. "Final Evaluation Report: Longitudinal Study of LA's BEST After-School Education and Enrichment Program, 1992-94." Los Angeles, CA: Center for the Study of Evaluation, Graduate School of Education and Information Studies, University of California at Los Angeles, Spring 1995.

Card, David and Alan Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, vol. 100, no. 1, February 1992, pp. 1-40.

Cohen, P.A., J. A. Kulik, and C. C. Kulik, "Educational Outcomes of Tutoring: A Meta-Analysis of Findings," *American Educational Research Journal*, vol. 19, no. 1, 1982, pp. 237-248.

Cooper, Harris, Kelly Charlton, Jeff Valentine, and Laura Muhlenbruck. "Making the Most of Summer School: A Meta-Analytic and Narrative Review." Forthcoming in the *Monographs of the Society for Research in Child Development*. May 23, 1999.

Council of Economic Advisors. "Families and the Labor Market, 1969-1999: Analyzing the Time Crunch." [http://www.whitehouse.gov/WH/EOP/CEA/html/famfinal.pdf]. May 1999.

de Kanter, Adriana, Leila Fiester, Andrew Lauland, and Valerie Romney. "Keeping Schools Open As Community Learning Centers: Extending Learning in a Safe, Drug-free Environment Before and After School." Washington, DC: US Department of Education, July 1997.

Evans, Ellis and Dan Marken. "Longitudinal Follow-up Comparison of Conventional and Extended-Day Public School Kindergarten Programs." Paper presented at the annual meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).

Fashola, Olatokumbo. "Review of Extended-Day and After-School Programs and their Effectiveness." http://www.csos.jhu.edu/crespar/CRESPAR%20Reports/report24entire.htm]. Accessed May 17, 1999.

Fiester, Leila, Elizabeth Reisner, Carolyn Marzke, and Anne Castle. "Beating the Odds from 3 to 6 p.m.: Evaluation Results from the TASC After-School Program's First Year." Washington, DC: Policy Studies Associates, Inc., June 1999.

Fusaro, Joseph A. "The Effect of Full-Day Kindergarten on Student Achievement: A Meta-Analysis." *Child Study Journal*, vol. 27, no. 4, 1997, pp. 269-277.

Hattie, John, H.W. Marsh, James T. Neill, and Garry E. Richards. "Adventure Education and Outward Bound: Out-of-Class Experiences That Make a Lasting Difference." *Review of Educational Research,* vol. 67, no. 1, Spring 1997, pp. 43-87.

Marshall, Nancy, Cynthia Garcia Coll, Fern Marx, Kathleen McCartney, Nancy Keefe, and Jennifer Ruh. "After-School Time and Children's Behavioral Adjustment." *Merrill-Palmer Quarterly*, vol. 43, no. 3, July 1997, pp. 497-514.

Pierce, Kim, Jill Hamm, and Deborah Lowe Vandell. "Experiences in After-School Programs and Children's Adjustment in First-Grade Classrooms." *Child Development*, vol. 70, no. 3, April 1999.

Posner, Jill and Deborah Lowe Vandell. "Low-Income Children's After-School Care: Are There Beneficial Effects of After-School Programs?" *Child Development Special Issue: Children and Poverty*, vol. 65, no. 2, April 1994, pp. 440-456.

Quinn, Jane. "Where Need Meets Opportunity: Youth Development Programs for Early Teens." *The Future of Children*, vol. 9, no. 2, Fall 1999, pp. 96-116.

Roderick, Melissa, Anthony Bryk, Brian Jacob, John Easton, and Elaine Allensworth. "Ending Social Promotion: Results from the First Two Years." Chicago, IL: Consortium on Chicago School Research, December 1999.

Rodriguez, Eunice, Thomas Hirschl, June Mead, and Stephen Goggin. "Understanding the Difference 4-H Clubs Make in the Lives of New York Youth: How 4-H Contributes to Positive Youth Development." [http://www.cce.cornell.edu/youth/4h/news/final_report.html]. August 1999.

Rosenthal, Robert and Deborah Lowe Vandell. "Quality of Care at School-Age Child-Care Programs: Regulatable Features, Observed Experiences, Child Perspectives, and Parent Perspectives." *Child Development*, vol. 67, 1996, pp. 2434-2445.

Ross, S. M., L. J. Smith, J. Casey, and R. E. Slavin. *Evaluation of the Extended Day Tutoring Program in Memphis City Schools.* Baltimore, MD: Johns Hopkins University for Research on the Education of Students Placed at Risk, 1996.

Schinke, Steven, Kristin Cole, and Mario Orlandi. "The Effects of Boys and Girls Clubs on Alcohol and Other Drug Use and Related Problems in Public Housing." New York, NY: Boys and Girls Club of America, March 1991.

Schinke, Steven, Kristin Cole, and Stephen Poulin. "Thirty Month Data and Process Findings: Evaluation of Educational Enhancement Program of Boys and Girls Clubs of America." New York, NY: Carnegie Corporation, December 1998.

Seligson, Michelle. "School-Age Child Care Comes of Age." *ChildCare Action News*, vol. 14, no. 1, January-February 1997.

Seppanen, Patricia, John Love, Dianne deVries, Lawrence Bernstein, Michelle Seligson, Fern Marx, and Ellen Kisker. "National Study of Before- and After-School Programs: Final Report." Report submitted to the Department of Education. Portsmouth, NH: RMC Research Corporation, 1993.

Taylor, Mathew J., Karl R. White, and Alice Kusmierek. "The Cost-Effectiveness of Increasing Hours per Week of Early Intervention Services for Young Children with Disabilities." *Early Education and Development*, vol. 4, no. 4, October 1993, pp. 238-255.

Teacher Center of New York State. "Goals 2000: National Education Goals." [http://www.tier.net/schools/tcenters/goals.htm]. Accessed January 20, 2000.

Tierney, Joseph, Jean Baldwin Grossman, and Nancy Resch. "Making a Difference: An Impact Study of Big Brothers/Big Sisters." Philadelphia, PA: Public/Private Ventures, November 1995.

Trousdale, Donna. "First-Year Evaluation of an After-School Program for Middle School Youth." *ERS Spectrum*, Summer 2000, [http://www.ers.org/spectrum/sum00a.htm]. Accessed September 18, 2000.

US Department of Education. "Safe and Smart: Making the After-School Hours Work for Kids." [http://www.ed.gov/pubs/SafeandSmart/index.html]. June 1998.

US General Accounting Office. *Welfare Reform: Implications of Increased Work Participation for Child Care.* Report GAO/HEHS-97-75. Washington, DC: GAO, May 1997.

Vandell, Deborah Lowe and Mary Anne Corasaniti. "The Relation between Third-Graders' After-School Care and Social, Academic, and Emotional Functioning." *Child Development*, vol. 59, 1988, pp. 868-875.

Vandell, Deborah Lowe and Kim Pierce. "Can After-School Programs Benefit Children who Live in High-Crime Neighborhoods?" Poster Symposium conducted at the meeting of the Society for Research in Child Development, Albuquerque, NM, April 14-18, 1999.

Walberg, Herbert J. and Shiow-ling Tsai. "Reading Achievement and Diminishing Returns to Time." *Journal of Educational Psychology*, vol. 76, no. 3, 1984, pp. 442-451.

Warren, Constancia, Prudence Brown, and Nicholas Freudenberg. "Evaluation of the New York City Beacons: Summary of Phase I Findings." Washington, DC: Academy for Educational Development, 1999.