Appendix A

Response Rates and Data Quality

The results presented in this report were based on two waves of data collected from a variety of respondents and from school and program records. In the fall of 1999-2000, we conducted baseline surveys with middle school students, elementary school students, and elementary school parents, and administered standardized reading tests to elementary school students. In the spring of 1999-2000, we surveyed students, parents, teachers, school principals, and center directors, coordinators, and staff. We also collected students' school records and program attendance, and we again administered standardized reading tests to elementary school students.

We collected data on schools and centers (from principals, after-school program directors, center coordinators, and staff) in 41 sites (34 middle and 7 elementary school sites). At two middle school sites, the baseline administration of student surveys was delayed because of the time needed for obtaining parental consent. When reporting data collected on individual students (from students, parents, teachers, and school records), we excluded those two sites. As part of the enhanced study supported by the grant from the C. S. Mott Foundation, another group of students not participating in the 21st-Century program at six middle school sites completed a questionnaire on their after-school activities.

A. Data Collection Procedures for Middle School Sites

1. Baseline

We surveyed 21,156 students in 32 middle school sites at the baseline. Questionnaires were generally self-administered during the school day. The questionnaire covered family background, after-school activities, school experiences, in-school and out-of-school behavior, and experiences in and knowledge of after-school programs. In most schools, students completed questionnaires two to six weeks after the 21st-Century program began operating for the year. A few weeks before administering the questionnaires at a site, we either mailed (or

gave to students to take home, depending on the preference of the school) letters explaining the study and requesting parental consent. Parents who declined to participate returned the consent form in the postage-paid, addressed envelope provided. Three sites required parents to return the forms to approve their child's participation. We also asked middle school students to assent to participate in the study. The questionnaire cover served as a detachable assent form and explained the study components, its voluntary nature, and the confidentiality of data. Only students who signed the cover completed the questionnaire.

Using after-school program attendance records collected for a four-week period at the start of each program, we classified surveyed students as participants (those who attended the 21st-Century center three or more times), under-attenders (students that had attended one to two times), or potential comparisons (all students at comparison schools, or students at the host schools that had not attended centers). We excluded under-attenders from the rest of the study. We then used propensity score methods to match participants with students in the potential comparison group (see Appendix B, Technical Methods, for a description of the matching process). Table A-1 shows the number and types of students that were surveyed at baseline.

Table A.1

Classification of Students Completing the Baseline Survey:
Middle School Sites

Student Status	Number	Percentage
Participant Group	2,472	11.7
Potential Comparison Group	17,596	83.1
Matched Comparisons	3,921	18.5
Nonmatched Comparisons	13,675	64.6
Under-attenders	1,088	5.1
Total Surveyed	**21,156**	**100.0**

NOTE: Because of rounding, percentages do not sum to 100.0.

After matching was completed, we requested parental consent for participants and matched comparison students to participate in the study. Two-thirds of the 6,393 parents who were asked to participate (67 percent) gave their consent (Table A-2), with the proportion ranging by site from 45 to 100 percent (Table A-3).

Table A.2

Percentage of Parents Consenting to Participate in the Study: Middle School Sites

	Sample Size					
	Total		Treatment		Comparison	
Parental Consent	N	%	N	%	N	%
Asked to Consent	6,393		2,472		3,921	
Consented	4,264	67	1,782	72	2,482	63

Table A.3

Consent Rates by Site: Middle School Sites

Percentage of Parents Consenting	Number of Sites
90 to 100	2
80 to 89	2
70 to 79	11
60 to 69	11
50 to 59	5
40 to 49	1
Total	**32**

2. Follow-Up

Approximately six weeks before the end of each school's 2000-2001 academic year, field staff returned to middle school sites to administer the follow-up questionnaire. These were nearly identical to the baseline questionnaires except that items on demographics and after-

school program participation in the previous (1999-2000) school year were dropped. Ninety-five percent of the 4,264 students in the study completed the questionnaire (Table A-4), and response rates were more than 90 percent at all but two sites (Table A-5). Nearly all students completed the survey in school (84 percent). The others (16 percent), who were primarily transfer students, completed the questionnaire with computer-assisted telephone interviewers.

B. Data Collection Procedures for Elementary School Sites

1. Baseline

We surveyed 90 percent of the 522 third- to sixth-grade elementary school students at baseline (Table A.4). Response rates ranged from 84 to 96 percent across the six sites (all students at the seventh elementary school site were in kindergarten through second grade and were not surveyed) (Table A.5). Questionnaires were generally self-administered during the school day (in a few instances teachers read the questions aloud to their class).

Like middle school students, elementary students were asked to assent to participate in the study by signing the cover of the questionnaire, and only students who gave their assent completed the questionnaire. MPR interviewers conducted telephone questionnaires with a small number of students who were not surveyed at school (primarily transfer students).

MPR field staff also administered the reading component of the Stanford Achievement Test 9 (SAT-9) in school to 70 percent of students in kindergarten through sixth grade who had not completed a district-administered version of the SAT-9 that fall or the previous spring (Table A.4). Response rates across sites ranged from 44 to 93 percent (Table A.5), excluding one site that provided SAT-9 test scores for students in grades 2 to 5 but did not allow kindergarteners and first-graders to be tested. MPR field staff administered tests in student homes to a small number of students who were not tested in school.

Table A.4

Sample Sizes and Response Rates for Student Data

	Sample Size					Response Rate					
	Total	Treatment		Comparison		Total		Treatment		Comparison	
Instrument	N	N	%	N	%	N	%	N	%	N	%
Middle School Follow-Up											
Student Survey	4,264	1,782	42	2,482	58	4,059	95	1,700	95	2,359	95
Elementary School Baseline											
Student Survey[a]	522	333	64	189	36	467	90	304	91	163	86
Student Test[b]	798	497	62	301	38	561	70	358	72	203	67
Parent Survey	973	589	61	384	39	861	88	528	90	333	87
Elementary School Follow-Up											
Student Survey[a]	522	333	64	189	36	441	85	285	86	156	83
Student Test[b]	621	394	63	227	37	522	85	342	87	180	79
Combined Elementary and Middle School Follow-Up											
Parent Survey	5,237	2,371	45	2,866	55	4,224	81	1,898	80	2,326	81
Middle school	4,264	1,782	42	2,482	58	3,595	84	1,495	84	2,100	85
Elementary school	973	589	61	384	39	629	65	403	68	226	59
Teacher Survey[c]	5,237	2,371	45	2,866	55	3,969	76	1,834	77	2,135	74
Middle school	4,264	1,782	42	2,482	58	3,307	78	1,425	80	1,882	76
Elementary school	973	589	61	384	39	662	68	409	69	253	66
School Records	5,237	2,371	45	2,866	55	4,923	94	2,253	95	2,670	93
Middle school	4,264	1,782	42	2,482	58	4,069	95	1,716	96	2,353	95
Elementary school	973	589	61	384	39	854	88	537	91	317	83

[a]Sample includes only grades 3 to 6.

[b]SAT-9 tests were administered only to students for whom districts did not have test scores.

[c]Sample size and response rates are based on number of students, not teachers; 82.5 percent of the 939 teachers in the sample completed surveys.

Table A.5

Response Rates by Site for Student Data

Instrument	Number of Sites: Total	Percentage: 90 to 100	80 to 89	70 to 79	60 to 69	50 to 59	Less than 50
Middle School Follow-Up							
Student Survey	32	30	2	0	0	0	0
Elementary School Baseline							
Student Survey[a]	6	2	4	0	0	0	0
Student Test[b]	7	1	1	2	0	1	2
Parent Survey	7	3	4	0	0	0	0
Elementary School Follow-Up							
Student Survey[a]	6	1	2	3	0	0	0
Student Test[b]	7	2	3	1	0	1	0
Combined Elementary and Middle School Follow-Up							
Parent Survey	39	8	18	6	6	1	0
Middle school	32	8	18	5	1	0	0
Elementary school	7	0	0	1	5	1	0
Teacher Survey[c]	39	11	9	8	4	2	5
Middle school	32	11	9	5	2	1	4
Elementary school	7	0	0	3	2	1	1
School Records	39	31	5	1	1	1	0
Middle school	32	27	4	0	0	1	0
Elementary school	7	4	1	1	1	0	0

[a]Surveys were administered only to third- to sixth-grade students; one elementary school site had no sample in those grades.

[b]SAT-9 tests were administered only to students for whom districts did not have test scores.

[c]Sample size and response rates are based on number of students, not teachers; 82.5 percent of the 939 teachers in the sample completed surveys.

We also asked elementary school parents to complete a baseline questionnaire about their academic expectations for their child, safety concerns, interactions with their child, and their child's social and behavioral outcomes, school experiences, and after-school activities the previous spring. Almost 9 of 10 parents (88 percent) completed a questionnaire (Table A.4).

Response rates across sites ranged from 82 to 95 percent (Table A.5). About one-fourth (26 percent) returned questionnaires by mail, and three-fourths (74 percent) completed them by telephone.

2. Follow-Up

Approximately six weeks before the end of each school's 2000-2001 academic year, field staff returned to elementary school sites to administer follow-up questionnaire, which were identical to the baseline questionnaires except for one item on language that was not asked again. Eighty-five percent of students in grades 3 to 6 completed the follow-up questionnaire (Table A.4), and response rates across sites ranged from 72 percent to 92 percent (Table A.5). Nearly all the students who completed the questionnaire did so in school. MPR interviewers administered telephone questionnaires to the rest (primarily transfer students).

We again administered the reading component of the SAT-9 in school to students in kindergarten through sixth grade who would not be given a district-administered version of the SAT-9 that spring. Eighty-five percent of students completed the test (Table A.4), with response rates ranging by site from 57 to 95 percent (Table A.5). MPR field staff administered make-up tests at students' homes to a small number of children who were not tested in school.

C. Elementary and Middle School Sites Combined: Follow-Up

Beginning in the late spring of 2000-2001, we collected data on individual students from parents, teachers, and school records. We also collected data on schools and centers from principals and program staff.

1. Data Collected on Individual Students from Parents, Teachers and Records

The parent follow-up questionnaire included many items from the baseline questionnaire administered to elementary school parents, as well as items on family and child characteristics, academic expectations for their child, safety concerns, interactions with their child, and their child's social and behavioral outcomes, after-school activities, school experiences, and after-school program experiences. Eighty-one percent of parents completed the follow-up questionnaire—84 percent of middle school and 65 percent of elementary school parents (Table A.4). Slightly more than half (54 percent) responded to a mail survey; we interviewed the rest by telephone (46 percent). Response rates ranged by site from 51 to 96 percent (Table A.5).

We asked the English teacher of students at middle school schools and the homeroom teacher of students at elementary schools to complete a questionnaire on the student's classroom behavior and academic performance, teacher views of the after-school program and the school environment, and teacher demographics. About 83 percent of the teachers completed questionnaires, which provided data on 76 percent of the students—78 percent of middle school students and 71 percent of elementary school students (Table A.4). Most teachers responded by mail (70 percent) or telephone (28 percent), though a few completed the survey via the Web (2 percent). Response rates across sites ranged from 0 to 100 percent (Table A.5).

At the end of the 2000-2001 school year, we collected student records, which contained information on the students' demographics, attendance, suspensions, retention, academic services received, disabilities, standardized test scores, and grades. We obtained school records for 94 percent of students—95 percent of middle school students and 88 percent of elementary school students (Table A.4). We collected more than 80 percent of records at all but three sites, with response rates ranging from 52 percent to 100 percent (Table A.5). Generally, students for whom we were unable to collect school records had transferred to other schools.

Table A.6 summarizes the data collected on individual students in the follow-up, showing the percentage of students for whom data were obtained from one instrument (school records), two instruments (school records and student survey), three instruments (school records, student survey, and parent survey), and four instruments (school records, student survey, parent survey, and teacher survey). At the middle school sites, for example, we collected data for 95 percent of students for one instrument, 91 percent for two instruments, 78 percent for three, and 62 percent for four. Response rates for students at elementary school sites are divided between those to whom student surveys were and were not administered (grades 3 to 6 and kindergarten to grade 2, respectively).

Table A.6

Follow-Up Response Rates for Individual Student Data

Students	Total	School Records		School Records and Student Survey		School Records and Student and Parent Surveys		School Records and Student, Parent, and Teacher Surveys	
		N	%	N	%	N	%	N	%
Middle School	4,264	4,069	95	3,895	91	3,339	78	2,653	62
Elementary School Grades 3-6	522	459	88	403	77	279	53	205	39
Elementary School Grades K-2	451	395	88	(a)	(a)	258	57	203	45

[a]A survey was not administered to students in kindergarten through second grade.

2. Data Collected from Center and School Staff

Principals completed questionnaires on the relationship between the school and the 21st-Century program and their views of the program's objectives, facilities and resources, sustainability, benefits, and challenges. Ninety-five percent of principals completed a questionnaire (Table A.7)—79 percent by telephone and 21 percent by mail.

Table A.7

Sample Sizes and Response Rates: Data Collected from School and Center Staff

Instrument	Sample Size	Response Rate	
		N	%
Principal Survey[a]	80	76	95
Project Director Survey[a]	41	39	95
Center Coordinator Survey[b]	89	77	87
Staff Survey[a]	894	609	68
Program Attendance Records	75	69	92

[a]Includes 41 sites

[b]Nine after-school programs had two center coordinators; both coordinators returned surveys at five after-school programs.

We asked all 21st-Century program staff to complete a questionnaire that included items on staff roles and responsibilities, program objectives, experiences, interactions with school-day teachers and administrators, interactions with parents, professional development, professional background, and demographics. Center coordinators responded to those questions, as well as to another module that asked about interactions with parents, size of program, staff recruitment and retention, program challenges, facilities and resources, sustainability, and additional items on their role and responsibilities in the program. Like the principal questionnaire, the project director questionnaire covered program objectives, sustainability, benefits, and challenges. Project directors also answered questions on their role and responsibilities in the program and on their experience.

Questionnaires were mailed to project directors to distribute to center coordinators and to (paid) staff that were age 19 and older. We conducted follow-up telephone interviews with nonrespondents. Ninety-five percent of project directors, 87 percent of center coordinators, and

68 percent of staff completed a questionnaire (Table A.7). Most responded by mail (70 percent of center coordinators and 65 percent of staff).

We collected program attendance records from 92 percent of 21st-Century program centers (Table A.7). The centers provided copies of their records in whatever form they typically maintained attendance, such as by day or by activities offered each day. In a few cases, centers provided the total number of days students attended, rather than the daily attendance records.

Although the elementary school study design precluded attendance by students in the control group, records showed that 8 percent attended the 21st-Century program at least once. There were a variety of reasons for controls being able to attend the program. For example, because of changes in program staff, some staff were not aware of the students who should have been excluded from the program. Of those controls that attended the program, about three-fourths (76 percent) attended from 1 to 25 days, and the average attendance was 17 days.

The middle school study design did not bar any students from attending the 21st-Century program. About 14 percent of students in the comparison group attended the program at least once. Most (89 percent) attended from 1 to 25 days, and the average attendance was 10 days.

3. Data Collected from Nonparticipants

As part of the enhanced study, we surveyed students not participating in the 21st-Century program in six sites. In these sites, we drew a random sample of nonparticipating students in the schools that had centers. Comparison students participating in the larger study had lower probabilities of selection than did other nonparticipating students. The nonparticipant questionnaire asked about students' after-school activities, self-concepts, homework, and demographics. We included a module of questions on awareness of and familiarity with the after-school program, reasons for not attending, and ways nonparticipants would be encouraged

to attend. Eighty-two percent of sampled students completed the survey (868 of 1,062 students). We surveyed most students by telephone, and obtained parental permission before beginning the interview. A small number of students completed the survey by mail.

D. Tests for Response Bias

Not all consenting middle school students completed the follow-up questionnaire, which introduces the possibility of response bias. Table A.8 shows means for a range of characteristics for the sample of students that consented to be in the study and for the sample of students that completed a follow-up questionnaire.

Comparing characteristics that differ significantly for the comparison group and for participants indicates that participants generally were at higher risk of academic difficulty. For example, participants had lower average grades and test scores, more disciplinary incidents, less parental education, and less parental income. Parental characteristics and test scores were not part of the matching process, and the differences evident in the table indicate that matching did not yield groups that were equivalent on these characteristics. However, essentially the same differences are evident in the sample for which follow-up questionnaires were obtained, which indicates that the process of responding to the questionnaire did not introduce further differences. This is an expected result considering the high follow-up response rate of 95 percent (see Table A.4).

In addition, not all consenting elementary school students completed the follow-up questionnaire. Table A.9 shows baseline characteristics of the treatment and control group students both at baseline and at follow-up. The few differences in baseline characteristics compared to the middle school sample is attributable to the random assignment design. As in middle school sites, the follow-up process evidently did not introduce bias, as four characteristics

Table A.8

Participant and Comparison Group Characteristics,
Middle School Centers

Variable	Baseline Characterics of Students Consenting to Be in the Study			Baseline Characteristics of Students Responding to Follow-Up Questionnaire		
	Participants	Comparison Group	p-value[a]	Participants	Comparison Group	p-value [a]
Race						
Black	27.6	24.0	0.01***	27.1	24.4	0.05*
White	32.7	35.9	0.03**	32.9	36.1	0.03**
Hispanic	27.9	26.2	0.22	28.1	25.9	0.12
Other	7.4	8.3	0.29	7.6	8.2	0.51
Grade						
6	17.6	17.8	0.85	17.7	18.2	0.69
7	41.5	42.0	0.74	41.8	42.0	0.94
8	34.9	35.6	0.64	34.6	35.3	0.65
Other	6.0	4.6	0.04**	5.9	4.6	0.07*
Average Grades	83.0	84.1	0***	83.19	84.14	0.01***
Homework						
The student does the homework teachers assign	3.48	3.54	0.07*	3.48	3.53	0.15
Mother or father helps student with homework	63.5	63.2	0.86	63.3	63	0.86
Mean of homework habits index (Low=Does Not Do Homework)	2.83	2.88	0.03**	2.83	2.88	0.02**
Number of Hours Read for Fun Yesterday	0.30	0.32	0.05*	0.30	0.32	0.06*
Number of Hours Watched TV Yesterday	2.14	2.01	0.01***	2.16	1.99	0***
Mean of Index of Confidence in Reading Skills (Low=Little Confidence)	3.12	3.21	0***	3.13	3.22	0***
Student Expects to Drop out of High School	2.5	1.9	0.14	2.5	1.6	0.04**
At Least One Parent Is a College Graduate	29.0	32.5	0.02**	29.5	33.0	0.02**
At Least One Parent Is a High School Dropout	17.3	15.5	0.13	17.0	15.6	0.25

Table A.8 (c*ontinued*)

Variable	Baseline Characterics of Students Consenting to Be in the Study			Baseline Characteristics of Students Responding to Follow-Up Questionnaire		
	Participants	Comparison Group	p-value[a]	Participants	Comparison Group	p-value [a]
Mean of Index of Discipline Problems (Low=Few Problems)	1.40	1.33	0***	1.39	1.32	0***
Mean of Index of Bad Behavior (Low=Never)	1.54	1.51	0.04**	1.54	1.51	0.03**
Mean of Index of Good Behavior (Low=Never)	3.03	3.02	0.61	3.03	3.03	0.9
Mean of Index of Using Drugs/Alcohol	1.12	1.1	0.06*	1.11	1.09	0.01**
Mean of Index of Empathy (Low=Poor)	3.10	3.09	0.92	3.10	3.10	0.81
Mean of Index of Controlling Destiny (Low=Poor)	3.02	3.02	0.96	3.02	3.03	0.71
Mean of Parental Discipline Index (Low=Least Strict)	2.94	2.95	0.52	2.94	2.95	0.5
Mean of Social Position Index (Low=Least Engaged/High Isolation)	3.43	3.46	0.04**	3.43	3.46	0.03**
Mean of Safety Index (Low = Not Safe)	3.33	3.36	0.05**	3.33	3.36	0.06*
Mean of Index of Been Harmed or Threatened (Low=Little Harm)	1.52	1.49	0.11	1.51	1.48	0.06*
Mother's Education						
Eighth grade or less	8.3	7.5	0.41	8.5	7.6	0.3
Some high school (did not graduate)	11.3	14.0	0.02**	11.3	14.0	0.02**
High school equivalence (GED)	5.8	5.3	0.47	5.9	5.2	0.38
High school graduate	25.3	22.2	0.03**	25.0	22.1	0.05*
Vocational, trade, or business school after leaving high school	8.8	7.7	0.21	8.8	7.7	0.25
Some college	18.3	18.6	0.82	18.2	18.5	0.79
Graduated from a two-year college	10.5	9.5	0.36	10.4	9.5	0.35
Four-year college degree or other advanced degree	10.5	14.4	0***	10.6	14.7	0***

Table A.8 (c*ontinued*)

Variable	Baseline Characterics of Students Consenting to Be in the Study			Baseline Characteristics of Students Responding to Follow-Up Questionnaire		
	Participants	Comparison Group	p-value[a]	Participants	Comparison Group	p-value [a]
Household Income						
Less than $10,999	14.2	13.9	0.79	13.5	13.9	0.72
$11,000 to $24,999	24.8	22.5	0.11	24.9	22.3	0.07*
$25,000 to $39,999	22.4	19.5	0.03**	22.6	19.6	0.04**
$40,000 to $59,999	16.7	18.3	0.22	16.9	18.3	0.27
More than $60,000	14.0	19.6	0***	14.2	19.7	0***
Sample Size	**1,483**	**2,090**		**1,431**	**2,024**	

SOURCE: Student Survey and Parent Survey, School Records.

[a]The p-value is the smallest level of significance at which the null hypothesis that the impact equals zero can be rejected. If the p-value is less than .01, an impact is significant at the 1 percent level; if the p-value is less than .05, the impact is significant at the 5 percent level, and so on.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

Table A.9

Treatment Group and Control Group Characteristics,
Elementary Schools

Variable	Baseline Characteristics			Baseline Characteristics of Students Responding to Follow-Up		
	Treatment Group	Control Group	p-value	Treatment Group	Control Group	p-value [a]
Gender						
Male	46.4	50.6	0.24	43.2	55.3	0.02**
Female	53.6	49.4	0.24	56.9	44.7	0.02**
Race						
Black	66.7	72.5	0.13	70.3	74.9	0.40
White	9.7	8.3	0.55	9.6	6.1	0.30
Hispanic	18.9	14.1	0.13	17.5	14.3	0.49
Other	1.6	2.3	0.57	1.6	1.7	0.97
Grade						
Kindergarten	10.5	12.9	0.25	0	0	
1	18.1	17.1	0.68	0	0	
2	17.0	18.0	0.67	0	0	
3	13.4	10.5	0.17	24.2	17.7	0.11
4	17.4	18.5	0.67	32.2	34.1	0.68
5	19.5	16.5	0.24	35.5	34.3	0.80
6	4.0	6.5	0.08*	8.1	14.0	0.05**
Mother's Average Age (Years)	36.4	35.5	0.21	38.9	38.4	0.66
Father's Average Age (Years)	37.5	39.0	0.09*	39.2	41.4	0.07*
Number of Tardy Arrivals During 1999-2000 School Year	3.4	3.6	0.78	4.0	3.0	0.30
Number of Absences During 1999-2000 School Year	7.2	7.5	0.62	6.6	6.5	0.83
Parent Feels It Is Safe for Child to Walk in Neighborhood	59.8	54.9	0.16	64.8	62.0	0.57
Parent Feels It Is Safe to Walk in Neighborhood	72.3	78.5	0.14	71.4	77.7	0.16

Table A.9 (c*ontinued*)

Variable	Baseline Characteristics			Baseline Characteristics of Students Responding to Follow-Up		
	Treatment Group	Control Group	p-value	Treatment Group	Control Group	p-value [a]
Sample Size for Items Above	**587**	**381**		**285**	**153**	
Baseline Reading Test Score (Percentile)	36.3	36.6	0.94	31.1	27.0	0.24
Sample Size for Reading Test Score	**378**	**206**		**227**	**102**	
Sample Size[b]	**278**	**148**		**158**	**71**	

SOURCE: Student Survey and Parent Survey, School Records.

[a]The p-value is the smallest level of significance at which the null hypothesis that the impact equals zero can be rejected. If the p-value is less than .01, an impact is significant at the 1 percent level; if the p-value is less than .05, the impact is significant at the 5 percent level, and so on.

[b]Sample sizes for the control group range from 148 to 381, sample sizes for the consenting treatment group range from 278 to 587. Sample sizes for control group members and treatment group members who completed followup surveys (third through fifth graders) range from 71 to 153 and from 158 to 285, respectively.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

are significantly different at baseline and the same four and one other were significantly different at follow-up. This also is an expected result, considering the high follow-up response rate of 85 percent for the student questionnaire (see Table A.4).

Appendix B

Technical Methods

This appendix describes the technical approach for estimating impacts of middle and elementary school centers. Section A provides details for the methods used to estimate impacts of middle school centers, presented in Chapter IV. Section B provides details for the methods used to estimate the impacts of elementary school centers, presented in Chapter V.

A. Methods for Estimating Impacts: Middle School Centers

As described in Chapter III, we estimated impacts for middle school centers using a comparison group design. We used propensity score matching techniques to select a comparison group in the 34 middle school grantees, of which 32 provided follow-up data used in the analysis. We used regression models to estimate impacts, and then weighted the site impacts to represent all first- through third-cohort grantees that served middle school students. We conducted a separate analysis to explore the relationship between attendance and outcomes.

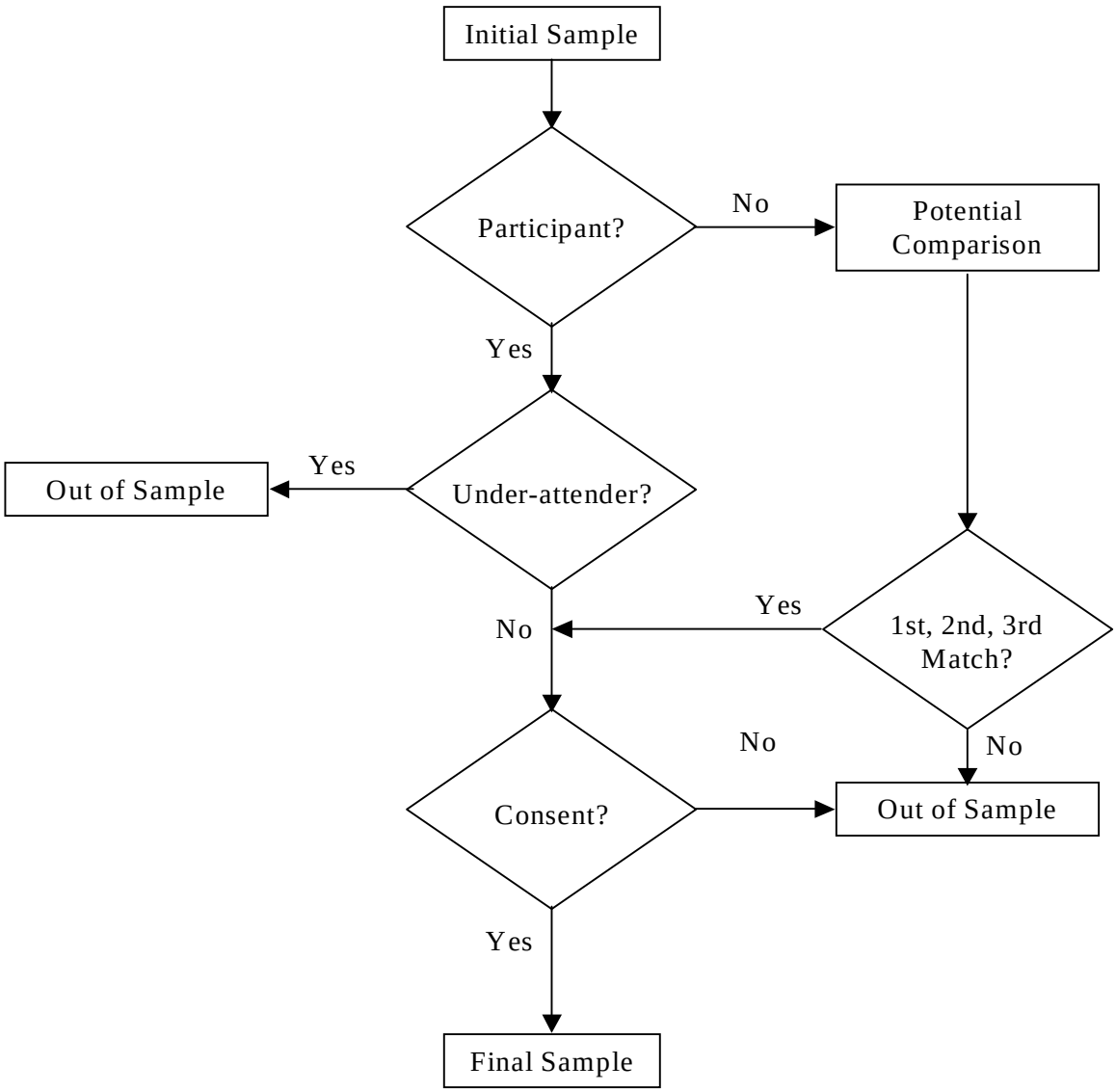
1. Identifying a Comparison Group Using Propensity Score Matching

To implement the comparison group design, we had a large group of students from schools operating centers and from similar schools complete questionnaires at the beginning of the 2000-2001 school year. Appendix A describes this part of the data collection in more detail and how the potential comparison group was identified from within the sample of students completing questionnaires. Ultimately, the potential comparison group was about seven times as large as the participant group. Having such a large potential comparison group provided a basis for the matching method to identify good matches. Figure B.1 depicts the various steps in the process of creating the comparison groups.

Using propensity score matching (PSM), we selected from the potential comparison group the students whose characteristics most closely resembled those of participants. PSM techniques were most prominently developed by Rosenbaum and Rubin (1983; 1985), who showed that the

Figure B.1

Overview of Process for Creating Middle School
Participant and Comparison Groups



technique yields a comparison group that is equivalent to a control group created using random assignment under the assumption that unobserved student characteristics are not correlated with outcomes. We applied the PSM method separately for each of the 34 middle school grantees, following five steps:

1. ***Logistic Regression Model Estimated.*** We estimated a logistic regression model in which the dependent variable was participation status and the independent variables were student demographic characteristics, indicators of student social development, measures of academic performance, and measures of student behavior (see Table B.1 for a listing of matching variables). Data on student characteristics used as a basis for matching were drawn from the baseline student questionnaire, the only data available at the time matching was conducted.[52] In most sites, 38 student characteristics were used in the matching.

2. ***Propensity Scores Calculated.*** For participants and potential comparison group students, we calculated propensity scores using data on each student's characteristics and parameter estimates from the logistic model estimated in the first stage. Presumably, students who actually were participants would have on average higher propensity scores than potential comparison group students; however, not every participant had a higher propensity score than every potential comparison group student.

3. ***Matching to Identify the "Closest" Comparison Students.*** For each participant, we selected, as that participant's "first-best" match, the potential comparison group student whose propensity score was numerically closest to the participant's score. To allow for possible attrition if parental consent was not received for the first-best match, we also identified potential comparison group students whose propensity scores would rank them as the second- and third-best match. We repeated the process for each participant, allowing individual students in the potential comparison group to be selected as matches for more than one participant.[53]

[52]Because school records data and parent questionnaire data were not available at the time matching was done, we could not include family income, parents' education and employment status, and students' baseline scores on standardized tests as matching variables.

[53]The number of participants to which a single potential comparison group student could be matched as a first-best match was limited to 10. The restriction was needed after it became evident that in some sites, a large proportion of participants were being matched to a single potential comparison group student, which resulted in a small comparison group in that site. Allowing such a comparison group to be selected would have dramatically reduced the statistical power of our analysis.

Table B.1

Percentage of Grantees for Which Participants and First- and Third-Best Matches Had Statistically Equivalent Mean Characteristics

	Original Match	
Variable	First Best	Third Best
Race		
Black	93.9	90.9
White	100.0	100.0
Hispanic	97.0	93.9
Other	97.0	93.9
Student Is a Female	100.0	87.9
Grade		
6	90.9	100.0
7	93.9	90.9
8	97.0	93.9
Other	97.0	93.9
Average Grades	100.0	97.0
Homework		
The student does the homework teachers assign	100.0	97.0
Mother or father helps student with homework	97.0	93.9
Mean of homework habits index (Low=Does Not Do Homework)	100.0	90.9
Number of Hours Read for Fun Yesterday	97.0	90.9
Number of Hours Watched TV Yesterday	97.0	93.9
Mean of Index of Confidence in Reading Skills (Low=Little Confidence)	93.9	90.9
Mean of Index of Helping Students Learn (Low=Little Help)	100.0	93.9
Overall Grade Student Gives School, 5=A through 1=F	97.0	100.0
Student Expects to Drop out of High School	97.0	93.9
Student Expects to Graduate from High School	100.0	93.9
At Least One Parent Is a College Graduate	100.0	93.9
At Least One Parent Is a High School Dropout	90.9	97.0
Child Doesn't Know Parents' Education Level	97.0	93.9
Mean of Index of Friends Encouraging Bad Behavior (Low=Never)	97.0	97.0
Mean of Index of Discipline Problems (Low=Few Problems)	100.0	97.0
Mean of Index of Bad Behavior (Low=Never)	97.0	93.9
Mean of Index of Good Behavior (Low=Never)	100.0	93.9

Table B.1 *(continued)*

Variable	Original Match	
	First Best	Third Best
Mean of Index of Using Drugs/Alcohol (Low=Does Not Use Drugs/Alcohol)	97.0	100.0
Mean of Index of Empathy (Low=Poor)	100.0	97.0
Mean of Index of Controlling Destiny (Low=Poor)	100.0	100.0
Mean of Parental Discipline Index (Low=Least Strict)	97.0	93.9
Mean of Social Position Index (Low=Least Engaged/High Isolation)	100.0	93.9
Overall Safety After School (1=Not Safe)	100.0	97.0
Mean of Safety Index (Low=Not Safe)	100.0	97.0
Mean of Index of School Climate (Low=Low Engagement)	97.0	97.0
Mean of Index of Serious School Problems (Low=Few Problems)	100.0	90.9
Mean of Index of Altering Behavior Because of Fear (Low=Never Alter Behavior)	97.0	97.0
Mean of Index of Been Harmed or Threatened (Low=Little Harm)	100.0	97.0
Average Percentage of Sites Matching Across All Characteristics	**97.8**	**95.0**

SOURCE: Student Survey.

NOTE: For each of 32 sites, t-tests are performed for the difference in means between participants and nonparticipants. This table shows the percentage of sites for which there is no significant difference between participants and nonparticipants. For example, in 100 percent of sites there was no difference in the percent of participants who were female and the percent of nonparticipants who were female for all samples except for the "third-best match" sample (in which there were no significant differences in 87.9 percent of sites).

4. ***Quality of Match Tested.*** Once we identified matching students for each participant, we could test the equivalence of the overall participant and matched comparison groups. We conducted an F-test of the equality of the set of characteristics for participants and their first-best matches a and used the *p*-value from the F-test as the indicator of match quality.[54] Higher *p*-values indicated that the groups were more similar.

5. ***Alternative Matching Specifications Assessed.*** To assess whether other comparison groups could be more equivalent, we used an algorithm to generate 2,000 logistic regression specifications by drawing randomly from the much larger set of all combinations of characteristics and second-order interactions of characteristics (squared terms and interacted terms). For each specification, we carried out steps 1 through 4 and used the comparison groups from the models with the highest, second-highest, and third-highest *p*-values from the F-test.

The matching process resulted in 3,921 comparison group students matched to 2,472 participants. Statistical tests verified the similarity of the matching students. The *p*-values for the joint test of equality of the 38 matching characteristics averaged 0.96, whereas the *p*-value of the F-test that these characteristics were the same for the participant group and the entire potential comparison group was 0.08. In other words, whereas participants and the potential comparison group had significantly different characteristics, participants and the matched comparison group were statistically indistinguishable at reasonable levels of significance.

We also examined the mean values of individual characteristics for participants and matched comparison students. Table B.1 consists of 38 rows corresponding to the 38 matching characteristics, showing the percentage of the 32 grantees for which there were significant differences between the mean value among participants and mean value among the first- (or third-) best matches. For each characteristic, the mean value in nearly all sites among participants and among matched comparison students was not significantly different at the 10

[54]The *p*-value indicates the probability that differences in the values of the participant and comparison group means that were observed could have resulted by chance under the null hypothesis that the means were jointly equal. Higher *p*-values imply that observed differences were more likely to have resulted from chance.

percent level. For example, the average characteristic had a significantly different mean value among participants and first-best matches in only 1 of 32 sites.

2. Impact Estimation

The basic approach for estimating the impact of middle school centers consisted of comparing the follow-up outcomes of participants and matched comparison group members, using regression models to adjust for baseline characteristics that may have influenced the outcomes. Outcomes of interest were regressed on an indicator of whether sample members were in the participant or comparison group, as well as a set of other explanatory variables. The basic regression model was:

$$(1) \quad Y = \alpha + P\beta + X\delta + \varepsilon$$

where $\alpha, \beta, \& \delta$ are coefficients that were estimated; P is the indicator of whether a student was in the participant group, and X represents the set of explanatory variables assumed to affect the outcome Y. (See Table B.2 for a list of the explanatory variables included in the model.) The estimated value of the coefficient β is an estimate of impact (that is, the difference in means between the participant group and the comparison group after adjusting for other characteristics). Because we used a complex sample design in selecting the sites and weights in the analysis, we used the SUDAAN® statistical package to estimate the standard errors of the coefficients of the model.

After estimating the regression models, we estimated "regression-adjusted" mean values of outcomes to facilitate interpreting the estimated impacts. Conceptually, the regression-adjusted mean value of an outcome for participants is the value of the outcome that the estimated model predicts for the average characteristics of the full sample. Similarly, the regression-adjusted

Table B.2

Explanatory Variables Included in the Basic Regression Model

Variable
Race/Ethnicity (White Excluded)
Black
Hispanic
Other race
Mixed race
English is not student's native language
Student is a female
Grade (grade 5 excluded)
Student is overage for grade level
Average grades
Average grades squared
Student-reported confidence in reading skills composite variable
Student-reported peer interaction/empathy composite variable
Overall grade student gives school, 5=A through 1=F
Student expects to drop out of high school or graduate from high school but not attend college
Student-reported discipline problems composite variable
Student-reported index of controlling destiny
Student-reported parental discipline composite variable
Student-reported social position composite variable
Student-reported safety index
Parent-reported variables
Family receives Food Stamps/Temporary Aid to Needy Families/Medicaid/housing assistance
Household income
Whether student's mother has a two- or four-year college degree
Whether student moved during previous year
Household structure (two-parent households excluded)
Student lives with single parent and no other adults
Student lives in other household arrangement
Student suspended during 1999-2000 school year
Number of times suspended during 1999-2000 school year
Number of absences during 1999-2000 school year
Number of times late to class during 1999-2000 school year
Student retained in grade prior to current year

SOURCE: Baseline Student Survey, Followup Parent Survey, Baseline School Records.

NOTE: All student-reported variables were drawn from the baseline student survey. Parent-reported variables were drawn from the follow-up parent survey but were limited to those variables that were unlikely to have been influenced by program participation. Variables based on school records data were limited to those that measured baseline outcomes. In addition to the variables listed, the explanatory variables also included missing value flags—binary indicators of observations in which a particular characteristic was missing and its value was imputed.

mean for comparison group students is the value of the outcome that the estimated model predicts for a comparison student who had the average characteristics of the full sample. The difference between the regression-adjusted mean outcome of participants and comparison group students is the estimated impact of participating in centers. The regression-adjusted mean values are calculated as follows:

1. ***Regression Model Estimated.*** Using data from the full sample, the regression model is estimated and coefficient estimates generated.
2. ***Predicted Outcome Values Calculated for Each Student.*** For every student in the sample, the coefficient estimates and the student's actual characteristics are used to calculate a predicted value of the outcome for that student under two different scenarios. A predicted value is calculated using all the student's characteristics except participation status, which is set to one under the assumption that the student is a participant. Another predicted value is calculated under the assumption that the student is a comparison group member (the participation status variable is set to zero).
3. ***Mean Values of the Two Predicted Values Calculated.*** Among all students in the sample, the mean values of the two predicted values are calculated using sample weights to ensure that the resulting mean value is representative of the population. The mean of the predicted values calculated under the assumption that each student was a participant is the regression-adjusted mean among participants. The mean of the predicted values calculated under the assumption that each student was a comparison group member is the regression-adjusted mean among comparison group students. The difference between these two regression-adjusted mean values is the estimated impact of participation and should be equal to the estimated coefficient β.

We used a variant of the basic regression model to estimate impacts for subgroups of students, supplementing the model with an interaction term between the treatment indicator and an indicator of whether sample members were in the subgroups being considered. An example of a subgroup model is:

(2) $Y = \alpha + P * S1\beta_1 + P * S2\beta_2 + X\delta + \varepsilon$

where the terms are defined as in equation (1) except that S1 is a binary variable denoting membership in a particular subgroup and S2 is a binary variable denoting membership in its

complementary subgroup. For example, to estimate the differential impact of program participation on outcomes for males and females, S1 might equal 1 for males and S2 would equal 1 for females. In this model, the estimated coefficient β_1 would be the model's estimate of the effect of center participation for males and β_2 would be the model's estimate of the effect of participation for females.

3. Analysis of the Relationship between Center Attendance and Outcomes

The approach to estimate the relationship between center attendance and outcomes was to adjust for observable differences between frequent and infrequent participants using regression models. The regression model that was estimated was:

$$Y_i = X_i b + a_1 P_i + a_2 D_i + u_i$$

where Y_i is the outcome for student i, X_i is a set of student characteristics, P_i is an indicator variable for whether a student is a center participant, and D_i is a variable indicating the number of days the student attended the center during the year. The observable characteristics (X_i) were the same as those used to estimate impacts in the basic model. The estimate of the coefficient a_2 represents the "effect" of attending the program for additional days.

To calculate regression-adjusted mean outcomes, we estimated the coefficient estimates from the above model to calculate predicted outcome values for an assumed level of attendance. We then averaged the predicted values to generate the predicted value of the outcome, which was termed the "moderate participation" value. We used similar techniques to estimate mean regression-adjusted outcomes for infrequent and frequent participants, which are presented in the text.

Because students could differ in unobserved characteristics that were not accounted for in the regression model, the estimated differences in outcomes do not represent the causal effect of the difference in attendance. An alternative approach to estimate causal effects would be to identify characteristics that are related to attendance but unrelated to outcomes, termed "instrumental variables," and apply well-known methods to estimate the effects of attendance. We considered several potential instrumental variables from the student and parent surveys but ultimately rejected them. One was mother's employment status, because mothers who work may have a greater need to place their child in an after-school program. However, mother's employment status proved to be nearly uncorrelated with attendance. We rejected other potential instruments for similar reasons.

4. Sample Weights

Because grantees included in this evaluation were sampled from among all grantees in cohorts one, two, and three that served middle school students, weights needed to be applied so that impact estimates could be applied to the full population of middle-school students served by centers. The construction of sample weights had two parts. First, a basic weight was constructed as the inverse of a grantee's probability of selection in its stratum. Second, the basic weight was modified so that the number of students in the sampled grantees represented the number of students in the stratum from which the grantees were sampled. The formula for the sample weight of a student was:

$$(3)\quad W_{ij} = \frac{1}{p_{j1} * p_{j2}}$$

where

$$p_{j1} = \frac{\#\, grantees \quad sampled \quad from \quad stratum}{total\,\# \quad grantees \quad in \quad stratum}$$

$$p_{j2} = \frac{n_p + n_c}{N_p}$$

where n_p and n_c are the number of treatment and comparison students for each grantee.

Table B.3 shows the 16 strata from which grantees were selected, along with the associated selection probabilities of grantees sampled from those strata. The effect of the first part of the sample weight is to make each student's data representative of students in grantees in the sample stratum that were and were not selected. The second part of the weight ($1/p_{j2}$) ensures that the weight given to all sampled students in a particular grantee depended on the number of eligible students served by the grantee (N_p,) rather than the number of treatment and comparison students included in the sample ($n_p + n_c$). When the weights from sample members within a site are summed, the site's cumulative weight is:

$$\sum_j W_{ij} = \sum_j \frac{1}{p_{j1} * p_{j2}} = \frac{n_p + n_c}{p_{j1} * p_{j2}} = \frac{n_p + n_c}{p_{j1}} * \frac{N_p}{n_p + n_c} = \frac{N_p}{p_{j1}}$$

B. Methods for Estimating Impacts: Elementary School Centers

We used an experimental design to estimate the impact of centers that served elementary school students. Students who were eligible to participate in elementary school centers were randomly assigned into a treatment group that was allowed to participate in the centers or a control group that was not allowed to participate. The experimental design ensured that the treatment group and control group were statistically similar in their baseline characteristics (both observed and unobserved). Thus, any outcome differences between the two groups at follow-up could be attributed to participation in the center.

The elementary school design was distinct from the middle school design in another respect. Whereas middle school grantees were selected for the evaluation at random from among all

Table B.3

Middle School Grantee Selection Probabilities

Stratum Number	Stratum Name	Number of Grantees in Population	Number of Grantees in Sample	Selection Probability
1	Northeast, Rural	16	2	0.13
2	Northeast, Urban	38	2	0.05
3	East, Rural	27	2	0.07
4	East, Urban	22	2	0.09
5	Southeast, Rural	32	2	0.06
6	Southeast, Urban	25	2	0.08
7	North Central, Rural	28	2	0.07
8	North Central, Urban	35	3	0.09
9	Midwest, Rural	22	2	0.09
10	Midwest, Urban	11	2	0.18
11	Mid-South, Rural	34	2	0.06
12	Mid-South, Urban	23	2	0.09
13	Northwest, Rural	37	3	0.08
14	Northwest, Urban	7	2	0.29
15	Southwest, Rural	27	2	0.07
16	Southwest, Urban	33	2	0.06

grantees serving middle school students, elementary school grantees were selected for the evaluation purposefully in order to ensure that they would be able to successfully implement random assignment. The impact findings for the elementary school centers in the evaluation have high internal validity but do not generalize to all grantees serving elementary school students.

If random assignment is correctly implemented, a comparison of average outcomes for the treatment and control groups is an estimate of the impact of participation in elementary school centers. However, the variance of the estimates can be reduced by estimating impacts using regression models to adjust for chance differences in baseline characteristics. The model used to estimate elementary school impacts was slightly different from that used to estimate middle school impacts. The model is

$$Y = \alpha + X\delta + \beta_1 P*G1 + \beta_2 P*G2 + ... + \beta_7 P*G7 + \varepsilon$$

In this model, we estimated separate impacts for each elementary school grantee (G1 through G7). To generate the estimate of the overall impact of elementary school centers, we calculated a simple mean of the seven site-specific impacts. We calculated the standard error of the overall impact estimated using the information from the variance-covariance matrix of the estimates of the seven coefficients representing the site-specific impacts. The regression model included whether students were overage for grade, race, parental education, parental income, household structure, whether the family received public assistance, the number of times the family moved in the past year and, when possible, the baseline value of the outcome variable.

For elementary schools, we used the same procedures to estimate regression-adjusted treatment and control means, subgroup impacts, and the attendance-outcome relationship described for middle schools. We also weighted students to offset differential probabilities of

selection that arose because centers had different numbers of applicants for slots and therefore required different ratios of treatment to control group assignments. In this case the weight simply was the inverse of the selection probability, suitably normalized to sum to the number of students that were randomly assigned.

References

Bissell, J., and J. Malloy. "Evaluation of California's After-School Learning and Safe Neighborhoods Partnerships Program: 1999-2001." Irvine, Calif., February 2002.

Brooks, Pauline, *et al*., "Longitudinal Study of LA's BEST After-School Education and Enrichment Program, 1992-1994." Los Angeles, Calif.: UCLA Center for the Study of Evaluation, 1995.

Cooper, Harris and Jeffrey Valentine. "Using Research to Answer Practical Questions about Homework." *Educational Psychologist,* 36(3), 2001, pp. 143-153.

Dynarski, Mark, Susanne James-Burdumy, Daniel Mayer, Mary Moore, John Mullens, Tim Silva, Carol Pistorino, and Douglas Hermond. *A Broader View: The National Evaluation of the 21st-Century Community Learning Centers Program.* Design Report submitted to the U.S. Department of Education. Princeton, N.J.: Mathematica Policy Research, March 2001.

Grossman, J., M. Price, V. Fellerath, L. Jucovy, L. Kotloff, R. Raley, and K. Walker. "Multiple Choices After School: Findings from the Extended-Service School Initiative." Philadelphia, Pa.: Public/Private Ventures, June 2002.

Halpern, Robert, Julie Spielberger, and Sylvan Robb, "Evaluation of the MOST (Making the Most of Out-of-School Time) Initiative: Final Report. Summary of Findings." Discussion Paper PS-32. Chicago: Chapin Hall Center for Children, 2001.

Langford, Barbara Hanson. "State Legislative Investments in School-Age Children and Youth." Washington, D.C.: The Finance Project, June 2001.

Moore, Mary, Mark Dynarski, John Mullens, Susanne James-Burdumy, and Linda Rosenberg. "Enhancing the 21st-Century Community Learning Centers Program Evaluation: A Concept Paper." Submitted to the Charles Stuart Mott Foundation. Princeton, N.J.: Mathematica Policy Research, November 2000.

Reisner, E., R. White, J. Birmingham, and M. Welsh. "Building Quality and Supporting Expansion of After-School Projects: Evaluation Results from the TASC After-School Program's Second Year." Washington, D.C.: Policy Studies Associates, February 2001.

Rosenbaum, Paul and Donald Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, 1983, pp. 41-55.

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

Trousdale, Donna. "First Year Evaluation of the After-School Program for Middle School Youth." Educational Research Service, Arlington, Va., Summer 2000. [http://www.ers.org/spectrum/ sum00a.htm].

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

U.S. Department of Education. "Working for Children and Families: Safe and Smart After-School Programs." [http://www.ed.gov/pubs/parents/SafeandSmart/index.html]. Washington, D.C., May 2000.

Warren, Constancia, P. Brown, and N. Freudenberg. "Evaluation of the New York City Beacons." Phase 1 Findings. New York: Academy for Educational Development, 1999.