When Elementary Schools Stay Open Late: Results from The National Evaluation
of the 21st Century Community Learning Centers Program

October 7, 2005

Susanne James-Burdumy
Mark Dynarski
John Deke

Mathematica Policy Research, Inc.

Abstract: *The U.S. Department of Education's 21st Century Community Learning Centers program has supported after-school programs since 1998. Research on the effects of after-school programs has been inconclusive, leading to an ongoing debate about their effects on children's outcomes. This paper presents evidence from a national evaluation which examined the impacts of 26 21st Century after-school programs in 12 school districts. A total of 2,308 elementary students who were eligible for and interested in attending a 21st Century Community Learning Center were randomly assigned either to the 21st Century center group (1,258 students) or to a control group (1,050 students). Impact findings are based on school records data, student surveys, parent surveys, teacher surveys, and the Stanford Achievement Test in reading. Key findings include reduced parent care after school, increased care by other adults after school, no impact on self-care, improved feelings of safety after school, no impact on academic achievement as measured by reading test scores or grades, and higher levels of negative behavior among the treatment group relative to the control group on multiple outcomes, including suspensions, teachers calling students' parents about behavior, and students being disciplined by teachers.*

INTRODUCTION

The number of after-school programs has grown rapidly in recent years, spurred by growing employment rates of mothers, pressure to increase academic achievement, and concerns about risks to children who are unsupervised during after-school hours. The percentage of public schools offering “extended-day” programs (which include before- and after-school programs) more than tripled between 1987 and 1999, from about 16 to 47 percent (DeAngelis and Rossi 1997; National Center for Education Statistics 2002).

In 1994, Congress authorized the 21st Century Community Learning Centers (21st Century) program to open up schools for broader use by their communities. In 1998, the program was refocused on providing academic, enrichment, and recreational activities in public schools during the after-school hours (centers also could be open before school, on weekends, and during the summer). The program grew from an appropriation of \$40 million in 1998 to \$1 billion in 2002, where it has remained.

Research on the impacts of after-school programs has accumulated in the past decade but is inconclusive. Some studies have reported that after-school programs increase academic achievement, increase safety, and reduce negative behaviors such as drug and alcohol use (Brooks et al. 1995; Hamilton and Klein 1998; Tierney et al. 1995; Welsh et al. 2002; Massachusetts 2020 and Boston Public Schools 2004; Reisner et al. 2004). Most of these studies also report negative or neutral findings, a pattern that has been noted by observers (Fashola 1998; Hollister 2003; National Research Council and Institute of Medicine 2002; Roth et al. 1998). In addition, most studies of after-school programs have used non-experimental designs with varying degrees of validity, adding to the difficulty of synthesizing the literature’s findings.

In 1999, we began a national evaluation of the U.S. Department of Education's (ED) 21st Century program. The challenge facing the national evaluation was to address three key questions about a rapidly growing and popular program at a level of rigor that would support policymakers in their efforts to develop the program.

1. Did the program improve student outcomes, such as supervision and safety after school, academic achievement, behavior, and social and emotional development?
2. What types of students benefited the most? For example, did particular groups of students, such as boys or students with low levels of academic achievement, experience more positive impacts than girls or students with high levels of academic achievement?
3. What were the features and characteristics of programs? How often did students attend the programs? What types of services did programs provide, and how often? What type of staff did programs hire? How did programs structure program activities?

The wide range of outcomes examined in the study was guided by program content and ED's priorities in the 21st Century program grant competitions, which called for programs to include extended learning opportunities, but also allowed them to include enrichment activities, such as recreation, music, and art.

To address these key research questions, the evaluation conducted an implementation study and an impact study. The implementation study was conducted to provide a better understanding of program features and operations and relied primarily on site visits and staff surveys for its data. The impact study was based on random assignment, in which outcomes of students assigned to the centers were compared to outcomes of students not assigned to the centers. The elementary grantees and centers in our study were purposively selected because they could implement random assignment, and the results apply to these grantees and centers. The results should not be interpreted as findings from the universe of 21st Century centers serving elementary school students.

The 21st Century Learning Centers Program

During spring 1998, ED made its initial 21st Century grants, totaling \$40 million, to the first cohort of school districts. Within a few months, Congress increased the program's funding to \$200 million. When the evaluation began in 1999, funding had increased tenfold in three fiscal years. Program funding continued to increase, rising to \$1 billion in 2001, where it has remained. Ultimately, ED made grants to seven cohorts of school districts, with funds going to almost 1,600 districts and 6,800 schools. In 2002, the No Child Left Behind Act (NCLB) substantially changed the program's structure, and money now is provided to states in proportion to the state allocation of Title I funds. States conduct their own grant competitions to fund programs that local school districts operate.

School districts receiving 21st Century funding were not necessarily operating after-school programs for the first time. For example, most districts in the study had operated some type of after-school program, though the programs may have been smaller or included fewer services and activities than those the districts could offer with the 21st Century grants. In fact, 57 percent of the elementary school grantees in the study had operated after-school programs in one or more schools that were part of the 21st Century grant. Data collection for the evaluation began after programs had been operating under their grants for at least one year, and sometimes for two years. Grants lasted for three years.

The legislation authorizing the 21st Century program did not require programs to focus on academic activities, but ED's implementation did. ED issued a "notice of priorities" in September 1997 that it would fund only programs that proposed academic activities in addition to other kinds of activities. In its response to public comments about the academic focus it was creating, ED provided a glimpse of its vision for the program:

For younger children who are not reading as well as they should, Community Learning Centers can provide extended time in which to overcome the obstacles that have in the past prevented them from becoming good readers. The competitive priorities will also encourage schools to develop strategies to address the needs of students who can benefit from additional enrichment or challenge in mathematics or science, or who are not performing as well as they should. Community learning centers can provide extended hours for students to learn and review basic concepts they may have missed during class, to delve deeper into a more challenging curriculum, or to participate in enjoyable hands-on activities and experiments. (*Federal Register*, December 2, 1997, page 63774)

Consistent with this vision, the initial grant announcement in the *Federal Register* on December 2, 1997, indicated that ED would fund only applicants that proposed “an array of inclusive and supervised services that include extended learning opportunities (such as instructional enrichment programs, tutoring, or homework assistance) but may also include recreational, musical and artistic activities.” ED also awarded additional points to applications that proposed activities that would help students meet state and federal standards in core academic subjects.

The academic focus was clarified further in subsequent grant competitions. In its competition for grants in 2001 (*Federal Register*, January 3, 2001), ED stated, “Applicants must describe in their application the elements of their projects that are designed to assist students to meet or exceed state and local standards in core academic subjects, as appropriate to the needs of the participating children.” The academic focus was put into law when the program was reauthorized by NCLB. NCLB defined the program’s first purpose as being to “provide opportunities for academic enrichment, including providing tutorial services to help students, particularly students who attend low-performing schools, to meet state and local student academic achievement standards in core academic subjects, such as reading and mathematics” (P.L. 107-110, Title IV, Part B, sec. 4201).

The Evaluation Design

The evaluation identified grantees with oversubscribed 21st Century centers serving elementary school students and implemented experimental designs. In fall 2000, roughly 1,000 students from 18 centers in seven school districts applied to 21st Century centers and were randomly assigned. Random assignment was conducted at the center level; for example, students who applied to center A were randomly assigned at that center and students who applied to center B were assigned at that center. In fall 2001, more than 1,300 students from eight centers in five additional school districts applied to 21st Century centers and were randomly assigned (the second cohort). In total, 2,308 elementary students who were eligible for and interested in attending a 21st Century Community Learning Center were randomly assigned either to the 21st Century center group (1,258 students) or to a control group (1,050 students). Results from the full sample of students in 26 elementary centers after two school years are presented in this paper.

National data from program performance reports indicate that schools operating 21st Century centers have higher poverty and a higher percentage of minority students than the average school nationwide. While this suggests that schools with 21st Century centers are different from the average school, this is consistent with the NCLB legislation, which noted an intent to serve students attending low-performing schools.

Data

The evaluation collected data on a wide array of outcomes, including grades, test scores, classroom behavior and effort, absences, suspensions, location and supervision after school, social development, parental involvement, negative behavior, and feelings of safety after school. The wide range of outcomes examined by the study reflects the many objectives embraced by after-school programs, which are suggested by the research literature, by other descriptions of

programs (see, for example, U.S. Department of Education 1998), and by ED's priorities in the 21st Century program grant competitions.

Baseline, first follow-up, and second follow-up data collection were conducted for all students in the treatment and control groups. In the study's first year of data collection, baseline and first follow-up data were collected for roughly 1,000 elementary students in 18 centers in 7 school districts (baseline data were collected in fall 2000, and first follow-up data were collected in spring 2001). In its second year of data collection, the evaluation added a second cohort of eight elementary school centers and administered the baseline and first follow-up surveys to students in these centers (for these new students, baseline data were collected in fall 2001, and first follow-up data were collected in spring 2002). First follow-up data from the first and second cohorts of elementary students were then combined. The second follow-up data collection was conducted for the first cohort of elementary students during spring 2002.

In spring 2003, during the evaluation's third and final year of data collection, the study conducted the second followup for the second cohort of elementary students. This paper combines these data with the second followup from the first cohort, to explore whether outcomes are affected by a second year of being able to attend the after-school programs in the evaluation.

The evaluation's data sources include questionnaires completed by students, parents, teachers, principals, and program staff, as well as reading tests, school records, center attendance records, and site visits. Second follow-up response rates were high, ranging from 76 percent for reading test scores to 88 percent for the student survey.

The CONSORT figure (see page 60, James-Burdumy et al. 2005) documents the flow of elementary students through the evaluation, starting from students' application to programs through the analysis of the second follow-up data. Of the 2,396 students who applied to the programs, 88 were excluded from the evaluation for various reasons, such as refusing to

participate or being ineligible. The remaining 2,308 students were randomly assigned to the treatment or control groups. Data collected from all students were analyzed. Twelve treatment-group students and 9 control-group students dropped out of the evaluation after the baseline survey.

Analysis Methods

To estimate program impacts, we used regression models that included outcomes at the second followup as dependent variables. Simple differences in mean outcomes of the treatment and control groups can provide unbiased estimates of the 21st Century program's impact, however, estimating impacts using regression models can increase the precision of the impact estimates and adjust for any residual differences between the treatment and control groups.

We included two types of independent variables in the regression models: (1) student characteristics (based on baseline data collected on students), and (2) variables created by interacting treatment status with the 12 site indicators. The student characteristics include students' grade, whether the student is overage for grade, race/ethnicity, number of absences, tardies, and suspensions in the year prior to the start of the study, whether the student has been retained in any year prior to the start of the study, math scores from the year prior to the start of the study, and reading scores from the year prior to the start of the study. We also included baseline measures of the outcome variables whenever possible.

The models yielded 12 impact estimates, one for each site, and the overall impacts were then calculated as the simple mean of the 12 site-specific impacts. The variance of the estimator was derived from the variance-covariance matrix of the 12 site impact estimates. Because the elementary sites and centers in our study were purposively selected, the results do not generalize statistically to the universe of 21st Century centers serving elementary students. If the evaluation had instead randomly selected sites and randomly selected centers within sites to participate in

the study, a more complex multilevel model would have been needed to account for sampling variability between centers and between sites.

It is common in program evaluation for some treatment-group members not to participate in the program after random assignment occurs. Therefore, we also estimated impacts on elementary school participants using a two-stage procedure. In the first stage, an indicator for whether students participated in the program was regressed on treatment status and baseline characteristics; in the second stage, outcomes at the second followup were regressed on predicted participation from the first-stage and the baseline characteristics.

Nonresponse weights were used in the regression models. They were constructed using a propensity-score approach, where the probability of responding to the follow-up survey was modeled as a logistic function of student baseline characteristics similar to those used as control variables in estimating impacts.

We also conducted sensitivity tests of the impact results and found that none of the checks alter our conclusions about the impacts of the program. We examined the effect of weights and regressors on the impact estimates by comparing four sets of impacts: (1) impacts presented below, which use nonresponse weights and regression adjustment; (2) impacts that use the weights but not regression adjustment; (3) impacts that do not use the weights but use regression adjustment; and (4) impacts that do not use the weights or regression adjustment. We also examined alternative definitions of self-care, and found that this did not change the self-care impact findings. The four definitions of self-care we investigated include: a student was defined to be in self-care if (1) the student did not spend at least three days with a parent, nonparent adult, or older sibling in a typical week; (2) the student did not spend at least one day with a parent, nonparent adult, or older sibling in a typical week; (3) the student was alone at least three

days in a typical week; or (4) the student was alone at least one day in a typical week. Finally, we examined site-level results, and found that impacts were not driven by individual sites.

IMPLEMENTATION STUDY RESULTS

As noted above, we conducted site visits to each grantee participating in the evaluation. We collected data from program directors, staff, and principals, and observed centers to analyze program objectives, activities, and staffing. Key findings include: (1) the most important program objectives were to provide a safe environment for students and to offer activities to help students improve academically, (2) many center staff were teachers, and (3) program leadership over the course of the study was stable, but line staff turnover was high. More details on the implementation study findings can be found in James-Burdumy et al. (2005).

Program Attendance

Students in the full treatment group attended an average of 81 days during the two-year follow-up period—49 days in the first year and 32 in the second. An important reason for the decline in average attendance in the second year compared with the first year is that one-fourth of the treatment group did not have access to the program in the second year, because they changed schools and their new school did not operate a 21st Century center. Figure 1 shows how the full sample of students breaks down in terms of whether they had access to centers and whether they attended if they did. Among the three-fourths who had access, 53 percent attended in the second year and 47 percent did not.

About 40 percent of the treatment-group students attended the program in the second year for at least one day. Attendance for these students (those who continued to attend in the second year) averaged about 81 days, which translates into roughly 2.7 days a week (centers were open

for 30 weeks on average) or 63 percent of days centers were open (on average, centers were open for 129 days).

Attendance varied throughout the school year. Figure 2 shows the average weekly attendance during the first and second year for students in the treatment group. Attendance patterns are similar in both years, but at a lower overall level in the second year because some students did not attend at all. Figure 3 shows the average attendance pattern only for students who attended in the second year, which removes students with zero attendance. Again, the patterns are similar in both years. The pattern is relatively constant, with sharp dips around major holidays.

Closer examination revealed large differences in average attendance across grantees. Average attendance for the three grantees with the lowest attendance was 29 days, compared to 104 days for the three grantees with the highest attendance. Districts with high and low average attendance were compared in terms of their attendance policies, urban and rural locations, and academic focus, but no patterns were evident, perhaps because of low statistical power—the sample was only 12 districts.

Student characteristics also did not explain most of the variation. We used regression models to examine relationships between the number of days attended and 15 student and family characteristics including grade, race/ethnicity, receipt of public assistance, mother's education, family income, household structure, whether the family had moved frequently in the past, whether the student was over age for their grade, whether the student was suspended in the year before the study, number of suspensions, absences, and tardies in the year before the study, whether students had ever been retained in the years prior to the study, and baseline test scores. Only one characteristic—high levels of maternal education—was significant (students whose

mothers had a high education level attended less). For the most part, variation in attendance seems to be related to factors not observed by the evaluation.

The study also examined whether different types of students participated in the program in the second year. Students from two-parent families and students with high baseline reading test scores were more likely to participate in the program in the second year, but other characteristics were not statistically significant.

IMPACT STUDY RESULTS

In this paper, “treatment group” refers to students who were randomly assigned and were able to attend centers, and “control group” refers to students who were randomly assigned and not able to attend centers. The term “participants” is used for treatment group members who attended a program in one or both years, as indicated by program attendance records.

At baseline, treatment and control groups were similar on a range of characteristics, such as gender, grade level, mother’s age, absences, suspensions, and reading test scores (Table 1). These similarities are expected because the groups were constructed using random assignment. One of the baseline variables differed significantly (students in the treatment group were more likely to report doing their homework). Considering the large number of variables reported in the table, some differences naturally would arise by chance, which also applies to the large number of impact estimates reported in this paper.

The evaluation also looked at attendance lists to determine whether control-group students attended centers. In principle, none of the control-group students would have attended centers. Over the course of two years, however, about 16 percent of control-group students were found in the attendance records, and overall, the control group averaged about 9 days of center attendance (compared to 81 days for the treatment group).

The tables show two types of impact estimates. The first are known as "intent-to-treat" estimates, and are based on the full treatment and control groups. The second set of estimates, which are shown in the column labeled "Estimated Impact on Participants" and sometimes are termed the impacts of "treatment on the treated," adjust for the 8 percent of treatments who never attended centers ("no show adjustment") and for the 16 percent of controls who attended centers for one day or more ("cross-over adjustment"). The properties of the estimation methods that the evaluation used make it possible for "intent-to-treat" impacts to be statistically significant while impacts for participants are not, and vice versa. In general, the participant impacts are qualitatively similar to the intent-to-treat impacts. The text below notes some outcomes for which differences were observed.

The Program Affected Whom Students Were With and Where Students Were After School

Treatment-group students were less likely than control-group students to be with their parents after school and more likely to be with other adults (Table 2). For example, 68 percent of treatment-group students were with parents after school at least three days in a typical week, compared to 75 percent of control-group students (effect size of 0.15), and 40 percent of treatment-group students were with other adults after school at least three days in a typical week, compared to 33 percent of control-group students (effect size of 0.14).

Supervision by older siblings did not differ significantly between the treatment and control groups. First-year findings indicated that supervision by siblings was lower among the treatment group relative to the control group.

The program had no impact on the frequency with which parents reported students to be in self-care. For the self-care estimate in Table 2, students were defined as being in self-care if, for at least three days in a typical week, they were not with a parent, a nonparent adult, or an older sibling. Just over one percent of treatment-group (and control-group) students were in self-care

in a typical week using this definition. Estimates based on other definitions of self-care, such as whether students were home alone at all during the week, or were home alone three or more days during the week, resulted in the same finding of no impact.¹

Students in the treatment group were more likely than students in the control group to be at school or in another place outside the home during the after-school hours (Table 2). Forty-six percent of treatment-group students were at school or another place outside the home at least three days in a typical week, compared to 36 percent of control-group students (effect size of 0.20). Students in the treatment group were less likely than students in the control group to be at home during the after-school hours, with 67 percent of treatment-group students at home after school at least three days in a typical week, compared to 76 percent of control-group students (effect size of 0.20). Treatment-group students stayed after school more frequently than control-group students, averaging 1.5 days per week, compared to 1.0 day for control-group students (effect size of 0.31).

Impacts of centers are likely related to whether students otherwise would have attended after-school programs. Using data from parent surveys on children's participation in after-school programs at the time of either of the two follow-up surveys, the study found that about 40 percent of parents of control students indicated that their child was attending an after-school program at the time of one or both of the surveys. In contrast, 69 percent of treatment parents indicated that their child was attending an after-school program at the time of one or both of the parent surveys.² The data do not provide more information about whether control and treatment students were attending similar types of programs, however, the data support the notion that there was a difference in rates of after-school program participation between the treatment and control groups.

The study also looked at activities of students after school. Treatment-group students were more likely than control-group students to report that they surfed the Internet or “did other things on the computer” (42 percent of treatment-group students and 36 percent of control-group students reported doing this activity at least one day in the previous week).³ Program participants reported watching TV or videos less often than nonparticipants, but this finding was statistically insignificant for the full sample.

The Program Had No Impact on Homework Completion

The evaluation gathered data from students and parents about participation in homework and from teachers about completion of homework. According to students, there was no impact of the program on homework or tutoring participation. According to parents, there was no impact of the program on homework participation, but treatment-group students were more likely than control-group students to participate in tutoring (Table 2). Student surveys were administered to students in grades 3 to 6, while parent and teachers surveys were administered to parents and teachers of all students regardless of grade, and differences in the samples may explain the differences in impacts on tutoring. In the second year, according to teachers, there was no impact of the program on homework completion (see Table 4). In the first year, the evaluation found that students in the treatment group were less likely than students in the control group to complete their homework.

It is not clear why the impacts on homework completion differ between the first and second years, but it could be related to the increase in the percentage of centers offering homework help sessions from the first year to the second year (85 and 100 percent, respectively). Consistent with this, the percentage of parents of treatment-group students reporting that their child did their homework after school increased from 84 percent in year 1 to 90 percent in year 2 ($p < .01$; compared to 89 and 91 percent, respectively, for parents of control-group students).

Table 3 provides a potential explanation for the lack of an effect on homework completion. The table shows that programs did not affect homework assistance received by students. Treatment-group students were no more likely than control-group students to have their parent or another adult ask if their homework is complete, look at their homework to see if it is complete, look at their homework to see if it is correct, or to explain homework in a way that is easy to understand. Almost 90 percent of control-group students reported that a parent or other adult asked them if their homework was complete, and about 80 percent reported that a parent or other adult looked at the homework to check it if was complete or correct.⁴

In addition, the structure of the homework help sessions that centers provided may not have been conducive to completing homework. Site visitors reported that most centers did not track the homework students should have been doing, or ensure that students completed their homework during the sessions. In addition, students mostly worked on their own, asking session monitors for assistance if needed.

The Program Had No Impacts on Academic Outcomes

There were no impacts of the program on reading test scores or course grades in math, English, science, or social studies (Table 4). Treatment-group students had an average reading score of 32.7 (in percentiles) on the Stanford Achievement Test—Version 9 (SAT-9) reading test, and control-group students had an average reading score of 32.4.⁵ Course grades were similar for the two groups. For example, treatment students had an average math grade of 79.9, while control students had an average math grade of 80.6 (Table 4).

The evaluation looked at other academic outcomes, including classroom effort and teacher reports of achievement (Table 4). According to teachers, effort and achievement were lower for treatment-group students relative to control-group students. Teachers reported that 22 percent of treatment-group students achieved at an “above-average” or “very high” level, compared to 28

percent of control-group students (effect size of 0.14). Similarly, 47 percent of treatment-group students “usually try hard” in reading and English according to their teachers, compared to 52 percent of control-group students (effect size of 0.11).⁶

There were no differences between the groups for related academic outcomes, such as completing assignments to the teacher’s satisfaction, coming to school prepared, and performing at or above the student’s ability level. However, the subgroup findings noted below indicate that there were some differences in students’ readiness for learning according to teachers (the impact of the program for boys on being “ready to learn” was significantly different than the impact for girls).

The Program Had Negative Impacts on Behavior

There was evidence of higher levels of behavior problems among treatment-group students relative to control-group students in the study’s second year (Table 5). For example, teachers reported having to call parents about behavior problems for 28 percent of treatment-group students and 23 percent of control-group students (effect size of 0.12). Twenty-two percent of treatment-group students reported that they were disciplined for behavior by having to miss recess or sit in the hall, compared to 17 percent of control-group students (effect size of 0.16). Twelve percent of treatment-group students were suspended from school, compared to 8 percent of control-group students (effect size of 0.16). The year one impact results were similar to these results, but were not statistically significant. The impact on suspensions was more significant for participants (at the .01 level) relative to the intent-to-treat model (significant at the .05 level).⁷

No impacts were observed on the extent to which teachers reported disciplining the child for misbehaving, sending the child to the office for misbehaving, or giving the child detention. However, a composite variable based on teacher reports of discipline problems was significantly

higher for treatment-group students relative to control-group students, as was a composite variable based on student reports of discipline problems.

Site visitors observed some evidence of behavior problems in centers, such as center staff struggling to maintain control of students or students talking back to staff, but it is not clear how negative behaviors in centers relate to negative behaviors during the school day. Previous research about whether after-school programs are related to negative behavior is mixed. A Massachusetts study reported significant increases in negative behaviors while students were in after-school programs (Massachusetts 2020 and Boston Public Schools 2004), and researchers studying a school district in suburban Dallas reported that children who attended day care centers after school (including after-school programs) were more likely to be viewed negatively by their peers (Vandell and Corasaniti 1988). Another study found that nonmaternal care (including after-school programs) was not related to behavior problems (National Institute of Child Health and Human Development Early Child Care Research Network 2004). Findings for subgroups discussed below indicate that the negative-behavior impacts are concentrated among boys and students who had high levels of disciplinary problems at baseline, which may provide some insight about the pathways.

The Program Had No Impacts on Parent Involvement

Parents of treatment-group and control-group students were equally likely to help their child with homework, check homework completion, and ask about things their child was doing in class (Table 6).⁸ Parents of treatment-group and control-group students were also equally likely to attend open houses at school, parent-teacher organization meetings, and after-school events and to volunteer to help out at school. Parents of participants were more likely than parents of nonparticipants to attend parent/teacher organization meetings.

The Program Increased Feelings of Safety

Treatment-group students reported feeling safer after school than control-group students (Table 7). Three percent of treatment-group students reported feeling “not at all safe” after school, compared to seven percent of control-group students (effect size of 0.21).⁹ The increase in feelings of safety for treatment students relative to control students suggests that centers were meeting one of their key objectives. The evaluation did not gather other data about whether student safety after school improved from other perspectives, which might be suggested, for example, by a lower incidence of victimization in the neighborhood area around schools that operated centers.

The Program Had Negative Impacts on Some Developmental Outcomes

Treatment-group students were less likely than control-group students to rate themselves highly on working with others on a team or group (Table 7): 78 percent of treatment-group students rated themselves highly on this measure, compared to 85 percent of control-group students (effect size of 0.19). In addition, according to teachers, treatment-group students were less likely than control-group students to get along well with others (effect size of 0.15). Treatment-group students were as likely as control-group students to (1) report getting along with others their age, (2) rate themselves highly on sticking to their beliefs even if their friends do not agree, and (3) report helping other students.¹⁰ Differences between students and teachers in the finding about getting along well with others could be attributable to sample differences (as noted above, only students in grades 3 to 6 completed questionnaires, whereas teachers completed questionnaires about students in grades K to 6).

Some Subgroups Had Different Patterns of Impacts

The evaluation estimated intent-to-treat impacts for six subgroups: (1) grade level, (2) whether students had low or high reading test scores at baseline, (3) whether students had low

or high behavior problems at baseline, (4) student race and ethnicity, (5) student gender, and (6) whether students lived in two-parent or one-parent households.^{11,12}

Two subgroup patterns are worth noting (see Tables 8-10). One is that students who had low test scores at baseline had impacts on grades which differed significantly from the impacts on grades for students with high test scores at baseline (Table 8). The differences in grade impacts between the high and low score groups were statistically significant for English (reading) and science, though not statistically significant for math and social studies (significant differences in impacts between two subgroups are indicated by bold text). While impacts on English and science grades for these two subgroups differed significantly from each other, only the impact on English grades differed significantly from zero for students in the low baseline score group (treatment students with low baseline scores had higher English grades than control students with low baseline scores).

Impacts for low-score students also differed significantly from impacts for high-score students on teacher reports of trying hard in reading and on a composite variable for classroom effort (not shown). Again, while impacts for these subgroups differed significantly from each other, only teacher reports of students trying hard in class in the high baseline group differed significantly from zero (treatment students with high scores at baseline were significantly less likely than control students with high scores at baseline to try hard in class, according to teachers). Low-score students also had impacts which differed significantly from impacts for high-score students on confidence in their reading abilities (Table 8). Treatment students with low baseline scores reported feeling less confident in their reading abilities than control students with low baseline scores.

The second subgroup pattern is that the negative behavior findings noted above (treatment-group students were more likely than control-group students to be suspended, have their parents

called about behavior problems, to be disciplined for behavior problems by having to miss recess or sit in the hall, and to have their parents come to school about a problem) are concentrated in two subgroups: (1) boys, and (2) students with high levels of disciplinary problems at baseline. The behavior results in Table 9 show differences for boys and girls. For four of the seven behavioral outcomes, there was a negative impact for boys. On none of the behavioral outcomes was there a negative impact for girls. The difference in impacts for boys and girls was statistically significant for one of the behavioral outcomes. For example, for girls, the estimated impact on teachers calling parents about behavior was zero, and, for boys, the estimated impact was an increase of 10 percentage points. Other behavior outcomes for boys and girls show a similar pattern, though the differences are not statistically significant.

The difference in impacts for boys and girls was statistically significant for teacher reports of the extent to which students come to class prepared and ready to learn (not shown). Treatment-group boys were significantly less likely to come to class prepared than control-group boys, and this impact differed significantly from the impact for girls (which was statistically insignificant).

Similar patterns also are evident in Table 10 for students with high- and low-discipline problems at baseline.¹³ For example, there was no impact of the program on being disciplined by missing recess or sitting in the hall for students with low-discipline problems, but high-discipline problem treatment students were more likely to be disciplined than high-discipline problem control students.

DISCUSSION

Directors of 21st Century programs in the national evaluation stated that their two main goals were to provide a safe environment for students after school and to help students improve

academically. The analysis and findings above suggest that 21st Century programs are meeting the first, but not the second, goal. In particular, elementary students in the treatment group were more likely to feel safe after school than elementary students in the control group, but treatment-group students scored no higher than control-group students on reading tests or grades in math, English, history, or science. For elementary school students who had low test scores at baseline, however, the program had a positive impact on English grades, possibly reflecting an ability to help low-achieving students. The difference was about 2 points on a 100-point scale.

Explanations for the first finding –improved feelings of safety after school – are not clear. This might be expected if fewer students were in self-care. However, the findings show that programs were shifting students from parent care to the care of other adults with no change in self-care. The answer may lie in how the study measured care outcomes. The study did not measure whether programs reduced the total amount of time students spent in self-care, which is distinct from the frequency of days in which they experienced self-care. A reduction in time spent in self-care could generate improved feelings of safety even though the incidence of self-care is unchanged.

There are a number of potential explanations for the lack of improved academic achievement. One potential explanation is a lack of coordination with the school day curriculum and teachers. While most programs reported that they understood the need to coordinate their activities with curriculum in the regular school day, and the need to have information flow between teachers in classrooms and staff in programs, they had varying degrees of success in facilitating the flow. Coordination was smoother when regular schoolteachers were also program staff. Coordination was weak or nonexistent in centers that either relied on outside staff, focused on noncognitive activities, or used burdensome processes, such as having teachers send homework assignments to programs or share lesson plans with them.

Another potential explanation is low levels of program attendance. Elementary students in the treatment group attended only about two days a week. The study did not find relationships between more frequent attendance and positive outcomes. However, more frequent and steadier attendance would likely help programs manage service delivery and integrate school-day and after-school instruction.

A third potential explanation for the lack of improvement in academic outcomes is that programs had no impact on the extent to which students completed homework or received help with it. Help with homework might have been one pathway by which programs could improve achievement. However, the study found that nearly all elementary school students already received homework help.

Programs also appeared to increase negative behaviors. Discussions with program directors indicated that students were not likely to be suspended because of their behavior during the after-school program, suggesting that, like the teacher discipline outcome, suspensions are related to negative behavior during the regular school day. A potential explanation is that attending centers created a long day for elementary children, which may have led to fatigue and negative behavior.

Subgroup analyses showed that nearly all the negative behaviors could be attributed to boys (behavior impacts for girls were close to zero and statistically insignificant) and to students who had a higher level of disciplinary problems at baseline, providing some insights about the pathways of behavior problems.

We conclude that while 21st Century after-school programs may help elementary students feel safer after school, they are not improving outcomes in the full range of areas suggested by the literature.

Some have suggested that this study's findings diverge from the literature in this area. We view our results as consistent with literature reviews in this area (Fashola 1998, Hollister 2003, Roth et al. 1998), which show a dearth of rigorous evidence of positive effects of after-school programs. In addition, a recent study found similar results. Vandell et al. (2005) studied 38 "high-quality" after-school programs serving elementary and middle school students that were believed to be efficacious, and found that a wide range of outcomes (such as work habits, aggressiveness with peers, and academic performance) did not differ between the treatment and comparison groups.

A limitation of the 21st Century study is that it was based on 27 after-school centers in 12 school districts who could support random assignment due to having more applicants than available slots. It is not clear why these programs were oversubscribed; perhaps the programs were popular, perhaps they were designed to be small to facilitate close relationships with students, or perhaps they were located in areas of high demand for after-school programs. However, that the programs in the study were able to support random assignment while most others could not does not mean the impact estimates necessarily would be substantially different if other programs were studied. This is a topic that future studies need to address. We can say that the findings are based on programs serving the type of students called for by the NCLB legislation.

References

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

DeAngelis, Karen, and Robert Rossi. (1997). *Schools Serving Family Needs: Extended-Day Programs in Public and Private Schools*. National Center for Education Statistics Issue Brief (NCES 97-590). Washington, DC: National Center for Education Statistics.

Dynarski, Mark, Susanne James-Burdumy, Mary Moore, Linda Rosenberg, John Deke, and Wendy Mansfield. (2004). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: New Findings*. Report submitted to the U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. Government Printing Office.

Dynarski, Mark, Mary Moore, John Mullens, Phil Gleason, Susanne James-Burdumy, Linda Rosenberg, Carol Pistorino, Tim Silva, John Deke, Wendy Mansfield, Sheila Heaviside, and Daniel Levy. (2003). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, First-Year Findings*. Report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc.

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

Fashola, Olatokunbo. (1998). *Review of Extended Day and After School Programs and Their Effectiveness*. CRESPAR Report No. 24. Center for Research on the Education of Students Placed at Risk. Available at [<http://www.csos.jhu.edu/crespar/techReports/Report24.pdf>].

Hamilton, Laura, and Stephen Klein. (1998). *Achievement Test Score Gains Among Participants in the Foundations School Age Enrichment Program*. Report prepared for Foundations, Inc. Santa Monica, CA: RAND Corporation.

Hollister, Rob. (2003). *The Growth in After-School Programs and Their Impact*. Washington, DC: Brookings Institution.

James-Burdumy, Susanne, Mark Dynarski, Mary Moore, John Deke, Wendy Mansfield and Carol Pistorino. (2005). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: Final Report*. Report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc.

Massachusetts 2020 and Boston Public Schools. (2004). *The Transition to Success Pilot Project*. Boston, MA: Massachusetts 2020.

National Center for Education Statistics. (2002). *Schools and Staffing Survey, 1999-2000. Overview of the Data for Public, Private, Public Charter, and Bureau of Indian Affairs Elementary and Secondary Schools*. NCES Report 2002-313 (Table 1.05). Washington, DC: NCES.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). *Are Child Developmental Outcomes Related to Before- and After-School Care Arrangements?* *Child Development*, 75, 280–295.

National Research Council and Institute of Medicine. (2002). *Community Programs to Promote Youth Development*. Committee on Community-Level Programs for Youth. Edited by Jacquelynne Eccles and Jennifer Appleton Gootman. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Office of the Federal Register, National Archives and Records Administration. (January 3, 2001). *21st Century Community Learning Centers Program Notice Inviting Applications*. Federal Register, 66(2), 353–355. Washington, DC: OFR.

Office of the Federal Register, National Archives and Records Administration. (December 20, 1999). *21st Century Community Learning Centers Program Notice Inviting Applications*. Federal Register, 64(243), 71263–71266. Washington, DC: OFR.

Office of the Federal Register, National Archives and Records Administration. (December 7, 1998). *21st Century Community Learning Centers Program Notice Inviting Applications*. Federal Register, 63(234), 67463–67465. Washington, DC: OFR.

Office of the Federal Register, National Archives and Records Administration. (December 2, 1997). *21st Century Community Learning Centers Program Notice Inviting Applications*. Federal Register, 62(231), 63776–63778. Washington, DC: OFR.

Reisner, Elizabeth, Richard White, Christina Russell, and Jennifer Birmingham. (2004). *Building Quality, Scale, and Effectiveness in After-School Programs*. Report submitted to The After-School Corporation. Washington, DC: Policy Studies Associates, Inc.

Roth, Jodie, Jeanne Brooks-Gunn, Lawrence Murray, and William Foster. (1998). Promoting Healthy Adolescents: Synthesis of Youth Development Program Evaluations. *Journal of Research on Adolescence*, 8(4), 423–459.

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

U.S. Department of Education. (1998). *Safe and Smart: Making After-School Hours Work for Kids*. Washington, DC: ED.

Vandell, Deborah Lowe, and Mary Anne Corasaniti. (1988). The Relation Between Third-Graders' After-School Care and Social, Academic, and Emotional Functioning. *Child Development*, 59, 868–875.

Vandell, Deborah, Elizabeth Reisner, B. Bradford Brown, Kimberly Dadisman, Kim Pierce, Dale Lee, and Ellen Pechman. (2005). *The Study of Promising After-School Programs: Examination of Intermediate Outcomes in Year 2*. Draft report. Washington, DC: Policy Studies Associates, Inc.

Welsh, Megan, Christina Russell, Imeh Williams, Elizabeth Reisner, and Richard White. (2002). *Promoting Learning and School Attendance Through After-School Programming: Student-Level Changes in Educational Performance Across TASC's First Three Years*. Draft report. Washington, DC: Policy Studies Associates, Inc.

Figure 1
Attendance in Second Year, Treatment Students

Source: Program Attendance Records and Student Tracking Data.

Figure 2

Figure 3

Source: Center Attendance Records.

Note: Figures exclude students who transferred during the year.

Table 1
Characteristics of Treatment- and Control-Group Students at Baseline,
Elementary School Centers

Characteristic	Treatment Group	Control Group	p-value ^a
Demographic Characteristics			
Gender			
Male	48.0	49.6	0.48
Female	52.0	50.4	0.48
Race/Ethnicity			
White (Non-Hispanic)	6.6	4.8	0.06
Black (Non-Hispanic)	54.2	55.0	0.06
Hispanic	35.3	36.2	0.06
Other	1.0	2.2	0.06
Mixed	2.9	1.7	0.06
Grade Level (Percentages)			
Kindergarten	10.5	10.3	0.95
1st grade	17.9	18.2	0.95
2nd grade	17.7	19.1	0.95
3rd grade	14.9	13.7	0.95
4th grade	16.6	17.1	0.95
5th grade	16.3	16.0	0.95
6th grade	6.2	5.7	0.95
Mother's Age (Years)	34.7	34.3	0.28
Academic and Other Characteristics at Baseline			
SAT-9 Reading Score (Percentile)	32.6	30.4	0.18
Number of Absences from School	6.4	6.5	0.87
Percentage of Students Who Were Suspended at Least Once in Previous School Year	2.9	3.0	0.93
Percentage of Students Who Report Feeling the Following Level of Safety After School up Until Dinnertime:			
Very safe	73.4	74.9	0.06
Somewhat safe	25.1	21.3	0.06
Not at all safe	1.9	3.8	0.06
Percentage of Students Who Report Doing the Homework Teachers Assign	42.9	38.8	0.04**
Sample Size ^b	1,247	1,041	

Source: Student Survey, Parent Survey, School Records.

^aThe p-value is the smallest level of significance at which the null hypothesis that the difference in means between program participants and control group members equals zero can be rejected. If the p-value is less than .01, the difference is significant at the 1 percent level. If the p-value is less than .05, the difference is significant at the 5 percent level, and so on. Chi-squared tests were conducted for categorical variables; for other variables, t-tests were conducted.

^bSample sizes differ depending on the data source. Sample sizes for demographic variables range from 746 to 1,041 for treatments and 936 to 1,247 for controls. Sample sizes on academic and other outcomes at baseline range from 501 to 721 for treatments and 567 to 847 for controls.

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

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

Table 2

Impacts on Students' Location, Supervision, and Activities After School,
Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Percentage of Students with the Following Individuals at Least Three Days After School in a Typical Week, According to Parent Reports:				
Self-care ^a	1.1	1.1	-0.1	1.3
Parent	67.8	74.5	-6.7**	-8.4***
Nonparent adult	39.5	32.8	6.7**	9.4***
Sibling	22.6	27.5	-4.9	-4.9
Mixed (Not in any one category for at least three days)	2.7	2.2	0.5	-0.8
Percentage of Students in the Following Locations After School at Least Three Days in a Typical Week, According to Parent Reports:				
Own home	67.3	76.0	-8.7***	-12.5***
Someone else's home	13.7	14.5	-0.8	-4.2
School or other place for activities	45.9	36.4	9.5***	17.2***
Somewhere to "hang out"	2.1	2.3	-0.2	-1.2
Mixed location (not in one location for at least three days)	1.4	0.9	0.4	0.7
Mean Number of Days Stayed After School for Activities in Typical Week, According to Parent Reports				
	1.5	1.0	0.5***	1.0***
Percentage of Students in the Following Activities After School at Least One Day in the Prior Week, According to Parent Reports:				
Homework	90.3	91.6	-1.3	-1.0
Tutoring	27.5	16.1	11.4***	15.1***
School activities (band, drama, etc.)	19.4	18.8	0.6	1.7
Organized sports	27.1	26.3	0.8	2.8
Clubs (Boy and Girl Scouts, Boys and Girls Club, etc.)	13.5	16.3	-2.9	1.3
Watched TV or videos	76.8	78.7	-1.9	-7.2**
Surfed the Internet or did other things on the computer	42.4	35.7	6.6**	8.2**
"Hung out" with friends	43.0	44.6	-1.6	1.0
Sample Size ^b	991	812		

Source: Parent Survey.

Note: The tables show two types of impact estimates: (1) "intent-to-treat" estimates (in the "Estimated Impact" column) use the full treatment and control groups, and (2) impacts on participants (in the "Estimated Impact on Participants" column) are the impacts after adjusting for the percentage of treatments who did not attend centers ("no-shows") and the percentage of controls who attended centers ("cross-overs"). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included students' demographic characteristics, students' baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates. Percentages may not sum to 100 because of rounding.

^aStudents are defined as being in self-care if they were not with a parent, nonparent adult, or older sibling at least three days in a typical week.

^bSample sizes differ for some outcomes due to nonresponse.

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

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

Table 3

Impacts on Homework Assistance, Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Percentage of Students Who Report That Their Parent “Often” or “Always” Does the Following:				
Asks if homework is complete	83.3	80.2	3.1	2.4
Looks at homework to see if it is complete	72.3	74.8	-2.5	-1.3
Looks at homework to see if it is correct	72.3	71.8	0.5	-1.5
Explains homework in a way that is easy to understand	73.5	73.8	-0.3	3.5
Percentage of Students Who Report That an Adult Who Is Not Their Parent “Often” or “Always” Does the Following:				
Asks if homework is complete	56.0	55.5	0.5	2.1
Looks at homework to see if it is complete	49.6	52.1	-2.6	4.3
Looks at homework to see if it is correct	52.1	57.2	-5.1	-0.8
Explains homework in a way that is easy to understand	55.1	54.4	0.8	6.7
Percentage of Students Who Report That Their Parent or an Adult Who Is Not Their Parent “Often” or “Always” Does the Following:				
Asks if homework is complete	89.0	87.1	1.9	0.7
Looks at homework to see if it is complete	80.1	81.8	-1.7	-1.4
Looks at homework to see if it is correct	80.4	82.1	-1.7	-3.8
Explains homework in a way that is easy to understand	80.5	83.3	-2.8	2.3
Percentage of Students Who Had the Following Individual Ask the Child to Correct Parts of Homework:				
Parent	89.5	91.6	-2.1	-1.8
An adult who is not their parent	75.0	76.2	-1.2	3.8
A parent or an adult who is not their parent	94.0	96.4	-2.4	-1.7
Sample Size ^a	769	647		

Source: Student Survey.

Note: The tables show two types of impact estimates: (1) “intent-to-treat” estimates (in the “Estimated Impact” column) use the full treatment and control groups, and (2) impacts on participants (in the “Estimated Impact on Participants” column) are the impacts after adjusting for the percentage of treatments who did not attend centers (“no-shows”) and the percentage of controls who attended centers (“cross-overs”). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included students’ demographic characteristics, students’ baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates.

^aSample sizes differ for some outcomes due to nonresponse. Sample sizes in this table are smaller than the other elementary impact tables because all outcomes in the table are from the student survey, which was not administered to students in grades K-2.

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

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

Table 4

Impacts on Academic Outcomes, Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Mean Grade:				
Math	79.9	80.6	-0.6	-0.6
English/language arts	81.2	80.8	0.3	0.1
Science	82.3	82.5	-0.1	-0.3
Social studies/history	81.2	82.2	-1.0	-0.8
Mean Reading Test Score	32.7	32.4	0.3	0.6
Percentage of Students Who Report That They “Often” or “Always” Complete the Homework Teachers Assign	77.8	79.4	-1.6	-4.5
Percentage of Students Whose Teachers Report That They “Often” Complete Their Homework	53.5	56.9	-3.4	-5.7
Percentage of Students Whose Teachers Report the Following:				
“Agree” or “Strongly Agree” That Student Completes Assignments to the Teacher’s Satisfaction	52.8	54.5	-1.7	-0.2
Student Achieves at “Above-Average” or “Very High” Level	22.2	28.1	-5.9**	-4.2
“Agree” or “Strongly Agree” That Student Comes to School Prepared and Ready to Learn	56.6	60.7	-4.0	-5.3
Student “Usually Tries Hard” in Reading or English	46.7	52.4	-5.7**	-6.0
Student “Often” Performs at or Above His or Her Ability	39.9	40.6	-0.7	-0.5
Level of Effort Composite ^a (Mean)	3.5	3.6	-0.1	-0.1
Sample Size ^b	1,055	880		

Source: Student Survey, School Records, Teacher Survey.

Note: The tables show two types of impact estimates: (1) “intent-to-treat” estimates (in the “Estimated Impact” column) use the full treatment and control groups, and (2) impacts on participants (in the “Estimated Impact on Participants” column) are the impacts after adjusting for the percentage of treatments who did not attend centers (“no-shows”) and the percentage of controls who attended centers (“cross-overs”). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included student characteristics such as indicators of students’ demographic characteristics, students’ baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates.

^aThe level of effort composite is based on five teacher-reported items regarding student (1) effort, (2) performance at ability level, (3) attentiveness, (4) participation, and (5) volunteering. Values on these items range from 1 to 5; a value of 1 on the composite indicates a low level, and a value of 5 indicates a high level.

^bSample sizes differ for some outcomes. For teacher-reported outcomes, the sample sizes are 967 treatment group members and 811 control group members; for student-reported outcomes, the sample sizes are 743 treatment group members and 619 control group members; for records outcomes, the sample sizes range from 819 to 1,044 for treatment group members and from 732 to 860 for control group members; for test scores, sample sizes are 952 for treatments and 796 for controls.

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

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

Table 5

Impacts on Behavioral Outcomes, Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Mean Number of Days Student Was:				
Absent	8.1	8.4	-0.2	0.1
Late	4.4	4.3	0.1	-0.3
Percentage of Students Whose Teachers Report Doing the Following “Two or More Times”:				
Disciplining the child for misbehaving	47.0	43.2	3.8	1.1
Sending child to the office for misbehaving	13.6	12.4	1.2	3.3
Giving child detention	20.8	17.5	3.4	3.1
Calling parents about child’s behavior	28.1	23.1	5.1**	4.5
Percentage of Students Who Report the Following Happens “Some” or “A Lot”:				
Student has to miss recess or sit in the hall	22.4	16.9	5.5**	3.7
Parents have to come to school about problem	22.3	16.8	5.6**	3.6
Student-Reported Discipline Problem Composite ^a (Mean)	1.7	1.6	0.1**	0.1
Teacher-Reported Discipline Problem Composite ^b (Mean)	1.8	1.7	0.1**	0.1
Negative Behavior Composite ^c (Mean)	1.7	1.7	0.0	0.0
Percentage of Students Who Were Suspended During Most Recent School Year	11.5	7.5	4.1**	5.3***
Sample Size ^d	1,055	880		

Source: Student Survey, School Records, Teacher Survey.

Note: The tables show two types of impact estimates: (1) “intent-to-treat” estimates (in the “Estimated Impact” column) use the full treatment and control groups, and (2) impacts on participants (in the “Estimated Impact on Participants” column) are the impacts after adjusting for the percentage of treatments who did not attend centers (“no-shows”) and the percentage of controls who attended centers (“cross-overs”). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included student characteristics such as indicators of students’ demographic characteristics, students’ baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates.

^aThe student-reported discipline composite is based on three responses: (1) how often the student is sent to the office for doing something wrong, (2) how often the student misses recess or sits in the hall, and (3) how often parents have to come to school about a problem. A value of 1 on the composite means a low occurrence of student-reported discipline problems, and a value of 3 means a high occurrence.

^bThe teacher-reported discipline composite is based on four teacher responses regarding the student’s behavior in the past month: (1) how often the student is disciplined for misbehaving, (2) how often the teacher has given the student detention, (3) how often the teacher has sent the student to the office for misbehaving, and (4) how often the teacher has contacted the student’s parents regarding behavior. A value of 1 on the composite means a low occurrence of teacher-reported discipline problems, and a value of 4 means a high occurrence.

^cThe negative behavior composite is based on student responses to five questions regarding how often they do the following: (1) break something on purpose, (2) punch or hit someone, (3) argue with their parents, (4) lie to their parents, and (5) give a teacher a “hard time.” Values on these items range from 1 to 4; a value of 1 on the composite indicates a low level, while a value of 4 indicates a high level.

^dSample sizes differ for some outcomes. For teacher-reported outcomes, the sample sizes are 967 treatment group members and 811 control group members; for student-reported outcomes, the sample sizes are 743 treatment group members and 619 control group members; and for records outcomes, the sample sizes range from 819 to 1,044 for treatment group members and from 732 to 860 for control group members.

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

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

Table 6

Impacts on Parent Involvement, Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Percentage of Students Whose Parents Report Doing the Following:				
Helped their child with homework at least three times last week	62.0	59.0	3.1	2.5
Checked on their child's homework completion at least three times last week	90.1	91.0	-0.9	-1.1
Asked their child about things they were doing in class at least seven times last month	66.7	68.4	-1.6	-1.5
Percentage of Students Whose Parents Did the Following at Least Three Times Last Year:				
Attended an open house at the school	40.3	39.0	1.2	0.9
Attended parent-teacher organization meetings	48.7	44.3	4.5	9.6***
Attended an after-school event	40.0	45.3	-5.3	0.1
Volunteered to help out at school	29.0	32.8	-3.8	-2.8
Sample Size ^a	784	658		

Source: Parent Survey.

Note: The tables show two types of impact estimates: (1) "intent-to-treat" estimates (in the "Estimated Impact" column) use the full treatment and control groups, and (2) impacts on participants (in the "Estimated Impact on Participants" column) are the impacts after adjusting for the percentage of treatments who did not attend centers ("no-shows") and the percentage of controls who attended centers ("cross-overs"). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included students' demographic characteristics, students' baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates.

^aSample sizes differ for some outcomes depending on nonresponse.

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

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

Table 7

Impacts on Other Outcomes, Elementary School Centers, Year 2

Outcome	Treatment Group	Control Group	Estimated Impact	Estimated Impact on Participants
Percentage of Students Who Report Feeling the Following Levels of Safety After School up Until 6 p.m.:				
Very safe	79.1	76.4	2.7	4.4
Somewhat safe	18.4	16.5	1.9	-1.5
Not at all safe	2.5	7.1	-4.6***	-2.9
Percentage of Students Who Rate Themselves as “Good” or “Excellent” on Working with Others on a Team or Group	77.7	84.7	-7.0**	-6.1**
Percentage of Students Who Report It to Be “Somewhat True” or “Very True” That They Get Along with Others their Age	85.9	85.8	0.1	4.5
Percentage of Students Whose Teachers Report That the Student Gets Along Well with Others	69.8	76.1	-6.3**	-8.1***
Percentage of Students Who Do the Following “Some” or “A Lot”:				
Help another student in school	75.0	73.4	1.6	0.7
Help another student after school	58.6	56.6	2.0	9.3**
Percentage of Students Who Rate Themselves as “Excellent” on Sticking to What They Believe in, Even if Their Friends Don’t Agree	50.0	52.1	-2.1	0.5
Sample Size ^a	784	658		

Source: Student Survey, Teacher Survey.

Note: The tables show two types of impact estimates: (1) “intent-to-treat” estimates (in the “Estimated Impact” column) use the full treatment and control groups, and (2) impacts on participants (in the “Estimated Impact on Participants” column) are the impacts after adjusting for the percentage of treatments who did not attend centers (“no-shows”) and the percentage of controls who attended centers (“cross-overs”). The percentages and mean values of outcomes for treatment and control students have been regression-adjusted for baseline differences between the groups. The control variables in the regression included students’ demographic characteristics, students’ baseline test scores, and school attendance. Weights are used to adjust impact estimates for nonresponse. Impacts on participants are estimated using an instrumental variables method, and the significance levels may differ from significance levels of the intent-to-treat estimates.

^aSample sizes differ for outcomes depending on the source. For student-reported outcomes, the sample sizes are 743 treatment-group members and 619 control-group members.

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

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

Table 8
Impacts on Academic Achievement by Subgroup, Elementary School Centers, Year 2

Outcome	Low Baseline Test Scores ^a			High Baseline Test Scores ^a		
	Treatment Mean	Control Mean	Estimated Impact	Treatment Mean	Control Mean	Estimated Impact
Mean Student-Reported Reading Confidence Composite	2.8	3.0	-0.2**	3.3	3.2	0.0
Percentage of Students Whose Teachers Report That They Achieve at an “Above-Average” or “Very High” Level	0.1	0.1	1.0	0.3	0.4	-7.8
Mean Class Grade						
Math	78.9	77.3	1.6	82.5	83.3	-0.8
English	79.9	77.5	2.4**	83.5	84.5	-1.0
Science	80.5	79.1	1.4	83.9	85.5	-1.6
Social Studies	79.1	78.6	0.5	83.7	85.9	-2.2**
Mean Reading Test Score	31.1	34.4	-4.4	48.8	45.9	5.4
Number of Observations:						
Student-reported outcomes		642			471	
Teacher-reported outcomes		686			653	
School records outcomes (Grades)		579			599	
School records outcomes (Reading scores)		631			693	

Source: Student Survey, School Records, Teacher Survey.

Note: Subgroup impacts reported in bold indicate that the estimated impact for one subgroup differed significantly from the estimated subgroup impact for the other related subgroup(s) at the .05 level or higher. Weights are used to adjust estimates for nonresponse.

^aStudents are defined as having low (high) scores if they scored below (above) the median reading test score for the study sample.

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

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

Table 9
Impacts on Behavioral Outcomes by Subgroup, Elementary School Centers, Year 2

Outcome	Male			Female		
	Treatment Mean	Control Mean	Estimated Impact	Treatment Mean	Control Mean	Estimated Impact
Percentage of Students Whose Teachers Report Doing the Following "Two or More Times":						
Disciplining the child for misbehaving	60.0	49.5	10.4**	33.9	34.9	-1.0
Sending child to the office for misbehaving	20.8	16.4	4.4	5.9	7.4	-1.5
Giving child detention	25.9	21.5	4.3	17.0	12.7	4.4
Calling parents about child's behavior	39.8	29.5	10.3**	15.6	15.5	0.0
Percentage of Students Who Report the Following Happens "Some" or "A Lot":						
Student has to miss recess or sit in the hall	31.2	21.0	10.2**	14.6	12.7	1.9
Parents have to come to school about problem	28.6	20.2	8.4	16.7	14.5	2.1
Percentage of Students Who Were Suspended	15.6	9.4	6.2**	6.9	6.6	0.3
Number of Observations:						
Teacher-reported outcomes		887			924	
School records outcomes (Suspensions)		846			879	
Student-reported outcomes		630			708	

Source: Student Survey, Teacher Survey, School Records.

Note: Subgroup impacts reported in bold indicate that the estimated impact for one subgroup differed significantly from the estimated subgroup impact for the other related subgroup(s) at the .05 level or higher. Weights are used to adjust estimates for nonresponse.

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

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

Table 10
Impacts on Behavioral Outcomes by Subgroup, Elementary School Centers, Year 2

Outcome	Low Baseline Disciplinary Problems Composite ^a			High Baseline Disciplinary Problems Composite ^a		
	Treatment Mean	Control Mean	Estimated Impact	Treatment Mean	Control Mean	Estimated Impact
Percentage of Students Whose Teachers Report Doing the Following “Two or More Times”:						
Disciplining the child for misbehaving	36.2	31.3	4.9	58.9	46.8	12.1
Sending child to the office for misbehaving	9.9	9.9	-0.1	20.5	21.3	-0.8
Giving child detention	15.8	13.3	2.6	32.3	18.2	14.2**
Calling parents about child’s behavior	20.6	17.4	3.2	36.1	24.7	11.4
Percentage of Students Who Report the Following Happens “Some” or “A Lot”:						
Student has to miss recess or sit in the hall	7.6	12.0	-4.4	29.6	16.7	12.9**
Parents have to come to school about problem	11.3	8.3	3.0	30.2	21.0	9.2
Percentage of Students Who Were Suspended	12.1	7.2	4.9	21.6	19.6	2.0
Number of Observations:						
Teacher-reported outcomes		614			303	
School records outcomes (Suspensions)		581			279	
Student-reported outcomes		636			318	

Source: Student Survey, Teacher Survey, School Records.

Note: Subgroup impacts reported in bold indicate that the estimated impact for one subgroup differed significantly from the estimated subgroup impact for the other related subgroup(s) at the .05 level or higher. Weights are used to adjust estimates for nonresponse.

^a The baseline student discipline composite was based on students' responses to how frequently the following three things happened to them: (1) sent to the office for doing something wrong, (2) have to miss recess or sit in the hall, and (3) parents have to come to school about a problem they're having. Students are defined as having low (high) levels of discipline problems if the composite falls below (above) the median of the composite for the study sample.

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

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

Footnotes

¹The amount of self-care observed in the evaluation is roughly consistent with estimates from the National Household Education Survey, which reported that two percent of K-2 grade students and eight percent of 3-5 grade students are in self-care (Kleiner et al. 2004).

²Attendance records indicate that 82 percent of treatment students attended 21st Century programs at some point during the study. Parent reports of after-school program attendance are lower than the estimates provided by the attendance records because the surveys asked about after-school program participation in the spring, whereas attendance records measured attendance over the duration of the study.

³The program's impacts on participation in tutoring are discussed in the section on homework, below.

⁴Tabulations from the National Household Education Survey for 2001 indicate that 59 percent of students in grades 1 to 5 get homework help three days a week or more (which can be interpreted as analogous to "often" or "always" getting help). Table 6 shows levels of assistance that are comparable to these estimates. For example, 59 percent of control-group parents reported helping their child with homework at least three times in the past week. The frequencies shown in Table 3 are not directly comparable because they address specific types of homework assistance provided.

⁵Baseline reading scores were imputed by calculating the mean baseline reading score among students with a baseline score and assigning the mean score to students who were missing the baseline score. Handling missing baseline reading scores in other ways, such as estimating impacts only for students with nonmissing baseline reading scores and excluding the baseline reading score from the list of regressors, did not change the findings.

⁶The pattern of impacts on participants for students trying hard in reading and students achieving at a high level was similar (p-values of .07 and .12, respectively).

⁷The pattern of impacts on participants on other behavior outcomes, including the student-reported discipline composite (p-value of .07) and the teacher-reported discipline composite (p-value of .06), was similar.

⁸These results contrast with the first-year findings, which found statistically significant, positive impacts on homework help, asking about what the child was doing in class, and attending an after-school event (Dynarski et al. 2004).

⁹The impact on participants (p-value of .08) was consistent with the intent-to-treat impact.

¹⁰In the first year, treatment-group students were more likely than control-group students to report helping other students after school. This was not found in the second year. However, participants were more likely than nonparticipants to report helping other students after school.

¹¹Students are defined as having low (high) scores if they scored below (above) the median reading test score for the evaluation sample. Students are defined as having low (high) behavior problems if their behavior problem composite variable is below (above) the median level of the behavior composite for the evaluation sample.

¹²The text focuses on impacts that differ significantly between subgroups (shown in tables as bold text) and that differ significantly from zero (marked by asterisks). Some caution should be exercised in interpreting the statistical test results because with the large number of outcomes and subgroups considered here, some statistical tests will be positive by chance.

¹³The correlation between being in the boy subgroup and the high-discipline subgroup was 0.16, which suggests that the subgroups overlap to some degree, but not substantially.