

Effects of Participation in After-School Programs for Middle School Students: A Randomized Trial

Denise Gottfredson, Amanda Brown Cross, Denise Wilson,
and Melissa Rorie

Department of Criminology & Criminal Justice, University of Maryland, College Park,
Maryland, USA

Nadine Connell

Department of Law and Justice Studies, Rowan University, Glassboro,
New Jersey, USA

Abstract: This study assessed the effects of attending an after-school program (ASP) on a range of outcomes for middle school youths. The program operated for 9 hr per week for 30 weeks and included attendance monitoring and reinforcement, academic assistance, a prevention curriculum, and recreational programming. Participants were 447 students randomly assigned either to the ASP or to after-school activities as usual. Program attendance was sporadic. Although treatment students experienced increased exposure to ASPs relative to controls, nearly all youth in both groups reported participating in some organized activity during the after-school hours, and the number of different activities in which youths reported being involved did not differ across groups. Participation in the treatment reduced time spent with friends with no adults present during the after-school hours. No differences between treatment and control youths were found on measures of conduct problems, academic performance, school attendance, or any of the intermediate behaviors and attitudes targeted. The study replicates findings from the national evaluation of the 21st Century Community Learning Centers Program for middle school students using a more rigorous research design.

Keywords: After-school programs, randomized trial, middle school

After-school programming has been increasing in the United States, with considerable federal, state, local, and private monies being invested in these programs. For example, the 21st Century Community Learning Center Program received approximately \$1 billion in federal funds annually from 2002

Address correspondence to Denise Gottfredson, Department of Criminology & Criminal Justice, University of Maryland, 2220D Lefrak Hall, College Park, MD 20742, USA.
E-mail: dgottfredson@crim.umd.edu

to 2007 to provide before- and after-school enrichment for students in low-performing schools. Estimates of total annual federal investment in out-of-school time have reached as high as \$3.6 billion (The Finance Project, 2007).

The rising popularity of after-school programs (ASPs) results primarily from new demands for accountability in education and the need for after-school care for children of working parents (Beckett, Hawken, & Jacknowitz, 2001; Gottfredson, Gertenblith, Soulé, Womer, & Lu, 2004; Kane, 2004; Lauer et al., 2006). Concerns about delinquency prevention are also linked to demand for ASPs, as the risk of juvenile arrest is highest during the after-school hours (Gottfredson, Gottfredson, & Weisman, 2001; Sickmund, Snyder, & Poe-Yamagata, 1997). The intuitive appeal of ASPs rests on the perception that unsupervised after-school time is either dangerous or simply wasted for adolescents. ASPs provide an opportunity to enhance academic learning and to introduce positive adult role models, and they may provide shelter from unsafe neighborhoods for low-income children in urban areas. ASPs are also a convenient platform on which to provide social and personal skills instruction that may be deemphasized in the classroom. Existing research, however, has found mixed results regarding the effectiveness of ASPs for achieving these outcomes.

PRIOR RESEARCH

Published Reviews

Although recent reviews generally agree that “ASPs are *capable* [italics added] of improving important youth outcomes” (Granger, Durlak, Yohalem, & Reisner, 2007, p. 3), very little can confidently be said about how they can achieve success. Further, many programs have been shown to have no effect on youth outcomes, and in some cases, ASP participants experience negative outcomes (e.g., conduct problems, increased substance use, and negative peer influence) in comparison with nonparticipants (Dynarski et al., 2003; Mahoney, 2000; Weisman et al., 2002). The uncertainty about the direction and magnitude of the effects of ASP participation results from the generally poor quality of ASP research. Very few studies of ASP effectiveness meet contemporary standards for scientific rigor in program evaluation (Flay et al., 2005). Conclusions from reviews differ depending on where they have set the cutoff for scientific rigor in deciding which studies to include. Similarly, reviews differ in their inclusion criteria for programs. Some reviews limit their data points to group-based activities that occur during the after-school hours and combine recreation and youth development activities with academic support activities (“typical” ASPs). Others include much more intensive and specialized programs that offer activities over the summer, on weekends, and during the school day in addition

to activities that take place during the after-school hours. Finally, evaluations of ASPs differ greatly in terms of the populations studied. Some include only elementary school age children, some only children at risk for academic failure, and some only socioeconomically disadvantaged children. Because the costs and benefits of participation in an ASP may differ for different populations, conclusions from reviews are expected to vary depending on the mixture of studies included in the review.

More than a dozen reviews of ASPs have been published in the past decade. Many of these reviews have focused broadly on "youth development" or "out-of-school" programs and, in so doing, have captured a broader set of programs than is of interest in this study. Eccles and Templeton (2002), for example, examined extracurricular activities, such as sports and leisure. Hollister (2003) included "out-of-school" programs that focused on youth development, many of which were not group-based programs but instead delivered tutoring or mentoring services to individual students. Fashola (1998) limited her review to group-based programs but included many programs that were not delivered during the after-school hours. Next we summarize findings from five recent reviews that focused on programs delivered mostly during the after-school hours to groups of youths. These reviews included only studies that used comparison groups and measured clearly defined outcomes. By summarizing results from a heterogeneous mixture of programs serving youths of different ages and different risk levels, these reviews focused on different slices of the ASP evaluation landscape and drew different conclusions about the effectiveness of ASPs. We follow this review of published reviews with a more focused summary of research on ASPs serving middle school youths.

Among the five recent reviews selected for summary here, three (Kane, 2004; Lauer et al., 2006; Scott-Little, Hamann, & Jurs, 2002) focused mainly on ASPs serving populations that were either socioeconomically disadvantaged or at risk for academic failure. Durlak and Weissberg (2007) included programs serving youths aged 5 to 18 years with no other limitations on youth characteristics, and Zief, Lauer, and Maynard (2006) excluded ASPs targeted specifically at youth with special needs such as learning disabilities, physical disabilities, emotional problems, or behavioral problems. The studies included in the Zief et al. review operated mostly in urban, school-based environments and served primarily low-income minority students in poor-performing schools. Most of the reviews included a mixture of studies of programs targeting elementary school students and middle school students, with a preponderance of the latter. Eighty percent of the studies included in Zief et al. served elementary-only populations.

The reviews also differed in the types of research designs included. As noted, all five limited their reviews to studies that included comparison groups and measured clearly defined outcomes. But they imposed different inclusion criteria related to the comparability of the treatment and comparison groups. Scott-Little et al. (2002), Lauer et al. (2006), and Durlak and Weissberg (2007)

all required a comparison group but imposed no requirements regarding comparability of the comparison groups or the application of statistical controls. The studies reported in Kane (2004) used mostly nonequivalent comparison group designs, but all studies that did not use random assignment statistically controlled for observed pretreatment differences between the ASP participants and the nonparticipants. Zief et al. (2006) limited the studies in their review to "well implemented experimental design studies" (p. 4). The probability is much higher in these studies that observed outcome differences between ASP participants and nonparticipants are not a result of unmeasured characteristics (such as motivation to attend).

Not surprisingly, the conclusions reached about the effectiveness of ASPs in these reviews vary. Zief et al. (2006) reported that most of their positive findings were on measures of time expenditure in the after-school hours: Participation positively influenced youths' participation in athletics and arts activities. Time spent in self-care was also lower for the ASP participants than for the controls. The review, however, found no significant effects of ASP participation on school attendance or behavioral outcomes. They noted "small but insignificant" (p. 22) effects of ASP participation on school grades.

Kane's (2004) review of four large ASPs designed to address school performance found that participants' grades and test scores were improved slightly by academic programs delivered in an ASP setting. However, these improvements often failed to reach standard levels of statistical significance. He concluded that it may be unrealistic to expect a relatively small amount of after-school academic support to have a large impact on achievement. Furthermore, the improvements in school outcomes observed by Kane were conditional on program attendance: The typical student attended the ASP programs only 1 or 2 days per week, but students who attended more frequently experienced better outcomes.

The reviews that employed less stringent criteria on the research designs arrived at more positive conclusions about the effectiveness of ASPs for influencing academic and behavioral outcomes. Lauer et al.'s (2006) meta-analysis of 35 after-school and summer-school programs for predominantly elementary-level, high-risk youths showed that such programs do, on average, have a measurable impact on students' academic performance. These programs provided a mixture of individual and group instruction methods. The average effect sizes for reading and math performance for ASPs in Lauer et al.'s study were small ($d = .07$ and $.16$, respectively) but statistically significant. Scott-Little et al. (2002) summarized results from 23 evaluations of ASPs that were not of a "drop-in" or "special activity" nature. They excluded from their review studies of tutoring and mentoring unless those activities were part of a broader program that was delivered in the after-school hours but included evaluations of programs that contained substantial in-school components. Scott-Little et al. concluded that the evaluations included in their review made "limited use of research designs that support causal conclusions and insufficient information to

allow for meta-analysis of program effects” (p. 387). However, their summary of evidence from four studies reporting on academic outcomes revealed small, positive effects ($d = .21$ and $.16$ for reading and math outcomes, respectively), and the authors tentatively concluded that ASPs can have positive impacts on participants.

Finally, Durlak and Weissberg’s (2007) meta-analysis summarized effects for programs promoting personal and social skills. Their analysis included 66 studies of ASPs for youth aged 5 to 18 years that had stated goals of promoting personal development in the areas of leadership, decision making, self-control, and so on. The studies included in their review evaluated interventions that occurred “outside of normal school hours,” but like Scott-Little et al. (2002) and Lauer et al. (2006), they also included summer programs and programs that contained a nontrivial in-school component. Durlak and Weissberg concluded on the basis of their meta-analysis that “after-school programs produced multiple benefits that pertain to youths’ personal, social and academic life” (p. 7). Specifically, their results showed that, on average, ASPs have a positive impact on school bonding, attitudes about self-efficacy and self-esteem, behavioral adjustment indicators (e.g., prosocial and antisocial behaviors as well as drug use), and school performance as measured by grades and achievement test scores (d s ranging from $.11$ to $.34$, with an average of $.22$). Of importance, these positive outcomes were detected only for programs that used evidenced-based skill training approaches. Programs that failed to include evidence-based approaches were unsuccessful in improving any outcome.

In summary, the existing reviews are inconsistent in their conclusions about the extent to which ASP participation influences important youth outcomes such as behavior and academic performance. Several reviews suggest that ASPs can produce small but measurable improvements in academic performance, and Durlak and Weissberg’s (2007) review suggests that benefits extend beyond academic performance to other behavioral and attitudinal outcomes. But the conclusions seem to depend on the characteristics of the programs and on the methodological rigor of the studies included in the reviews.¹ Most important, the methodological rigor of almost all studies of ASPs is below par. More

¹For example, the Durlak and Weissberg (2007) study employed a lower standard for methodological rigor in determining which studies to include in the review. This review included a larger number of studies than other reviews (66 compared with a range of 5 to 35 in the other reviews). Although the authors reported using a variety of strategies to rule out methodological factors as explanations for their findings, the mean effect sizes reported contain several extremely high effect sizes from methodologically weak studies. We examined the studies contributing the seven highest effect sizes (all with overall effect sizes greater than $.60$) and found that five of the seven studies suffered from severe attrition, differential attrition rates for the treatment and comparison groups, or obvious selection artifacts stemming from nonequivalent groups.

conclusive statements about ASP effectiveness must wait until more rigorous studies have been conducted.

Middle School Programs. The existing reviews of ASPs also contain many programs for elementary school age youths. As such, they are not directly relevant to our study of voluntary ASPs for middle school youths. We summarized 12 studies that targeted middle school youths. Table A1 in the appendix provides detailed information about the methods used, type of students targeted, and results from these studies.² Like the studies included in the reviews just summarized, many studies of middle school populations fall short of contemporary standards for scientific rigor required to establish intervention effectiveness because they fail to randomize subjects to conditions and sometimes suffer from attrition problems. Only two studies (Lauver, 2002; Smith & Kennedy, 1991) used a randomized design, and one of those studies included only girls. Only six (50%) studies could be considered reasonably rigorous in terms of their research designs, although none met the high standards described in Flay et al. (2005).

The stronger studies provided inconsistent findings regarding the benefits of ASP participation for middle school youths. Two of these studies reported on ASP effects on school attendance: Lauver (2002) found no effects, and Dyanarski et al. (2004) found beneficial effects. Three of these studies reported on ASP effects on academic performance: Lauver found no effects, Dyanarski et al. (2004) found beneficial effects on social studies but not other grades, and Weisman et al. (2002) found negative effects on grade point average. Several studies reported on ASP effects on measures of student misconduct or school dropout. Some found positive effects (Huang, Sung Kim, Marshall, & Perez, 2005; Smith & Kennedy, 1991), but others found negative effects (Dyanarski et al., 2004; Weisman et al., 2002). Clearly, we do not yet have a coherent picture of the effect of ASPs on the behaviors of middle school youths.

Two of the studies summarized in the appendix attempted to identify characteristics of ASPs related to their effectiveness. In a study of 14 ASPs in Maryland, Gottfredson et al. (2004) concluded that an emphasis on social and character development distinguished ASPs that reduced problem behaviors

²The studies included in Table A1 were identified through a bibliographic search of published and unpublished evaluations of ASPs that were similar to the one that is the subject of this evaluation in terms of population served and basic structure. We searched multiple online databases using the keywords “after school” and examined the reference lists of previous reviews of ASP research. Inclusion criteria were (a) the study reported analysis of student outcomes related to problem behavior or academic achievement, (b) the study included a comparison condition, (c) at least 50% of the study sample was middle school aged (11–13 years old or in Grades 6–8), (d) the program was delivered primarily during the after-school hours, and (e) the program included more than 10 sessions. This process yielded 12 qualifying studies.

from those that did not. Another study of 35 ASPs in Maryland found that the delinquent behavior of participants decreased significantly in comparison with nonattenders (Gottfredson, Cross, & Soulé, 2007). Program characteristics associated with positive outcomes for youth in these programs included smaller program size, use of published curricula, and more educated staff.

Among the studies summarized in Table A1, the national evaluation of the 21st Century Community Learning Centers (Dynarski et al., 2004) is most well known. This evaluation of ASPs funded by a large federal program used a comparison group design for a nationally representative sample of grantees that operated 21st Century programs in middle schools. The evaluation assessed programs run in 61 schools, including data from approximately 4,400 students collected over a 2-year period. The evaluated programs served an average of 60 youths per day and provided mostly homework assistance and recreational activities. The study found that participating students were no more likely to finish their homework or feel safe after school, despite these being two of the stated goals of the program. In fact, middle school participants were more likely to have had their property damaged, more likely to report they had used or sold drugs, and less likely to rate themselves positively at working out conflicts with others (Dynarski et al., 2003). However, the results from the evaluation of the middle school programs (which used a nonequivalent comparison group design) are not as conclusive as are the results from the evaluation of the elementary school programs (which used an experimental design). The current study provides a rigorous empirical evaluation of a middle school ASP that resembles those included in the 21st Century program study.

Of particular interest in the study reported here is an assessment of the extent to which youths who are registered for the ASP are actually exposed to potentially helpful services, and the extent to which the ASP provides an effective alternative to unsupervised time spent with peers during the after school hours. Prior evaluations of ASPs (just summarized) suggested that the voluntary nature of these programs may limit attendance and therefore exposure to program content. Also, the extent to which such programs succeed at reducing unsupervised time spent with peers will depend not only on program attendance but also on the nature of the experience youths would have during the after school hours if they did not have the opportunity to attend the ASP. That is, ASPs that in effect replace time spent at home with adult guardians or time spent in more closely supervised after-school activities with time spent in an ASP are unlikely to have an impact on unsupervised time usage. The effectiveness of ASPs for achieving desired outcomes will no doubt depend upon these critical features.

METHOD

This study randomly assigned students within each of five participating schools to an experimental ASP or to a "treatment as usual" control group. The program

operated for 9 hr per week for 30 weeks and offered attendance monitoring and reinforcement, academic assistance, a prevention curriculum, and recreational activities. The ASPs were located in public middle schools in Baltimore County, Maryland, that served high percentages of minority, socioeconomically disadvantaged youths. Youth surveys and school records were collected pre- and postprogram to measure the key outcomes targeted as well as the key intermediate outcomes. Teacher ratings were collected at the end of the program year. A survey measuring exposure to after-school activities was administered midway through the program year. Program observations were conducted twice per month, and implementation data were collected daily. Regression models for each dependent variable compared treatment and control group means on each outcome and mediator. Standardized mean difference effects size statistics were calculated using the adjusted posttest means. Next we describe the study setting, sample, intervention, counterfactual condition, measures, and analysis strategy.

Setting

The experimental ASP was implemented during the 2006–2007 school year via a partnership among four public agencies in Maryland.³ The five school sites selected were the first among all low-performing middle schools in the county to express interest and agree to cooperate with the research procedures. The participating schools had high populations of minority youth, large numbers of students receiving subsidized meals, and high mobility. The principals at all five sites expressed the need for ASPs in their schools, stating that no other comprehensive, school-based ASPs were available to their students.

³The University of Maryland (UMD), Baltimore County Local Management Board (LMB), Baltimore County Department of Recreation and Parks (BCRP), and the Baltimore County Public Schools (BCPS) worked together to implement the experimental program in the five schools; to provide space, supplies, and employees; and to evaluate the success of the program. UMD provided all material and personnel support for the formal evaluation of the effects of program participation on youth outcomes. The LMB is a small county government agency charged with facilitating collaboration across public and private child-serving agencies. This agency, which funds, monitors, and evaluates a variety of ASPs, facilitated involvement of BCPS and issued a request for proposals (RFP) to secure a vendor who would provide after-school services to youth. BCPS provided access to the student population, space for the program, and limited oversight to ensure compliance with school system regulations and coordination with the BCPS education program. BCRP was selected through a competitive process to run the programs. BCRP agreed to implement the program as described in the RFP, to conform to the requirements of the research design, and to assist with the process of data collection. A representative from the LMB managed and oversaw the contract performance of the vendor and provided continual coordination with the school system.

Table 1. Demographic characteristics of schools and number of students registered for the after-school program (ASP) at each participating school, by site, 2006–2007

School	Total Enrollment	No. of ASP Registrations	% Minority	% Subsidized Meals	% Mobility ^a
A	839	71	64.4	65.0	20.6
B	484	101	47.1	64.8	21.1
C	683	72	50.8	67.0	13.8
D	566	120	97.9	48.9	21.3
E	719	83	99.3	63.4	16.9
All BC middle schools ^b					
Average	879	—	51.5	37.2	12.3
Minimum	484	—	8.7	6.0	2.8
Maximum	1,490	—	99.3	67.0	22.0

^aThe percentage of students withdrawing for any reason during the school year.

^bAlternative schools are omitted.

Demographic characteristics of the five participating school sites and all Baltimore County middle schools are presented in Table 1.

Sample

All students who attended the five participating schools were eligible to register for the ASP. The study's recruitment goal was 100 students per school for a total of 500 students. Within each school, registered students had a 50% chance of being randomly assigned to the treatment group (i.e., invited to attend the ASP) or to the control group.

Participant recruitment began in the spring of 2006 as a joint effort by University of Maryland (UMD) and Baltimore County Department of Recreation and Parks (BCRP). Efforts included promoting the program at school and community events, including promotional fliers in every student's start-of-year-orientation packet, mailing multiple recruitment postcards to all students' homes, and placing an automated message on the home phones of eligible youth. Finally, school principals sent letters to academically or behaviorally "at-risk" students to encourage their enrollment in the ASP.

Recruitment goals were ultimately met or exceeded at two of the five school sites (see Table 1). Because recruitment lagged behind the ideal, principal referral and postcard recruitment efforts were continued into the fall 2006 semester after the programs had opened. When recruitment ended in January 2007, 447 students had registered and completed a pretest. Students who registered for the ASP were generally representative of the populations of their schools in terms of gender and socioeconomic status (as determined by receipt of subsidized meals), but the ASP attracted a disproportionate number of minority youths.

Table 2. Demographic characteristics of sample, by experimental group

Demographics	Total Sample ^a		Treatment ^b		Control ^c	
	<i>M</i> or %	<i>N</i>	<i>M</i> or %	<i>N</i>	<i>M</i> or %	<i>N</i>
Age	12.22 (.99)	447	12.30 (1.03)	224	12.15 (.94)	223
Family income (median)	\$32,040	403	\$32,894	204	\$32,000	199
% male	53.69	447	52.68	224	54.71	223
% Black	69.58	447	68.75	224	70.40	223
% 6th grade	41.83	447	42.41	224	41.26	223
% 7th grade	33.56	447	30.36	224	36.77	223
% 8th grade	24.61	447	27.23	224	21.97	223
% living with two parents	36.91	447	36.61	224	37.22	223
% subsidized meals	58.68	438	58.99	217	58.37	221
% mother is college graduate	12.56	438	13.57	221	11.52	217

Note. Standard deviations are in parentheses.

^a*N* = 447. ^b*n* = 224. ^c*n* = 223.

The 447 students were randomized into treatment and control conditions by the principal investigator using a random number generator in SPSS version 11.0 (SPSS Inc., Chicago, IL). Students were randomized into conditions within their schools, such that each student had a 50% chance of assignment to the treatment condition within his or her school. This method ensured that treatment and control groups would be of equivalent size at each school. Several rounds of randomization were conducted as new registrations were received throughout the recruitment period.

Randomization was successful in creating equivalent groups. Treatment and control students did not differ in terms of demographics (see Table 2) and, as reported in Gottfredson, Cross, Wilson, Rorie, and Connell (2010), differed significantly on only 1 of the 18 pretreatment measures. One difference out of 18 tests conducted is approximately what would be expected by chance using a critical value of $p < .05$.

Power Analysis

We conducted power analyses for two-tailed independent *t* tests, fixing the Type I error rate (alpha) at 0.05 using software for "Power and Sample Size Calculations" (PS, V2.1.31), based on work by Dupont and Plummer (1990, 1998). These power analyses estimated the power available, given the observed standard deviation of several outcomes, to detect an effect of 0.3 standard deviation units. These analyses indicated that, using the entire sample, the power to detect differences of the specified size between the treatment and control

groups is 0.99. We also estimated the minimum effect that would be detected in our study. Following procedures outlined in Lipsey (1998), we estimated that we are able to detect effects of about 0.17 with 80% power. We conclude that the study has ample power to detect effects as small as 0.17, and that it has ample power to detect effects in the 0.3 range in reasonably-sized subgroups.

Study Conditions

The program was offered on school grounds, 3 days per week (Tuesday, Wednesday, and Thursday), for 3 hr after the close of the regular school day. The number of youths served per day was restricted to a maximum of 50, but typically programs served far fewer than 50 students on a given day. Activities traditionally offered in ASPs (e.g., snack, sports, and crafts) occupied about two thirds of the program time. The program, free to participants, also included attendance monitoring and reinforcement, academic assistance, and the All Stars curriculum (<http://www.allstarsprevention.com/>). On Tuesdays and Wednesdays All Stars and academic assistance occupied 1.5 hr of the 3-hr program. Thursdays were devoted entirely to recreation and leisure activities.

Figure 1 depicts the program model for the ASP and the intended outcomes of each of the main components. Prior evaluations and meta-analyses of ASPs suggested that a combination of individual and group-based academic services provided in ASPs can have positive effects of grades and test scores (Lauer

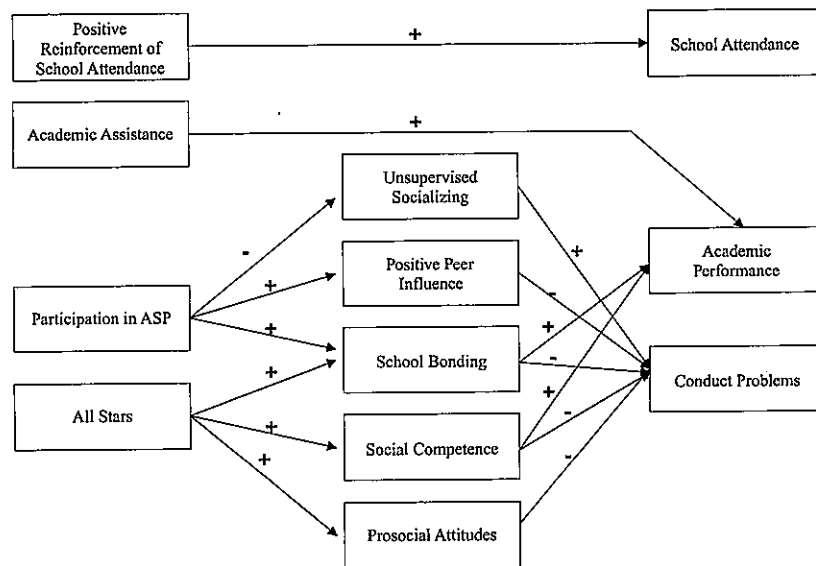


Figure 1. After-school program (ASP) logic model.

et al., 2006) but that positive effects are more likely to be observed for students who attend more often than 1 or 2 days per week (Kane, 2004). An attendance incentive component was therefore incorporated into the experimental ASP. This component, which provided rewards to individual students with good attendance in school, was expected to increase school attendance, which was in turn expected to increase participation in the ASP. Simply attending the ASP was expected to reduce unsupervised socializing, increase positive peer influences, and promote bonding to school, all of which were expected subsequently to influence conduct problems. The academic assistance component included homework assistance, workbooks, and independent reading. Consistent participation in these academic activities was expected to increase academic performance, as shown in Figure 1.

Prior evaluations and meta-analyses of ASPs suggested that ASPs were more effective for improving personal and social skills outcomes when they incorporated evidence-based social skills training components (Durlak & Weissberg, 2007), an emphasis on social and character development (Gottfredson et al., 2004), and use of published curricula (Gottfredson et al., 2007). The experimental ASP therefore incorporated the All Stars curriculum, which had been demonstrated in prior research to reduce substance use and aggressive behavior and to increase social competency skills (W. B. Hansen, 1996; W. B. Hansen & Dusenbury, 2004; Harrington, Giles, Hoyle, Feeney, & Youngbluth, 2001; McNeal, Hansen, Harrington, & Giles, 2004). This program focused on building attitudes and beliefs that are inconsistent with substance use and other risky behaviors, and on teaching skills necessary for healthy decision making. As Figure 1 shows, All Stars was intended to increase school bonding, social competence, and prosocial attitudes and beliefs, indirectly influencing academic performance and conduct problems via these mediators.

The control condition was "treatment as usual" except that members of the control group were invited to attend one after-school activity per month. Sites usually planned a special event or party for the days that control students were invited to attend. Eight such control group days occurred at the sites during the program year. These events were not well attended by control students. Fifty-two percent never attended; 29% attended once, twice, or three times; and 17% attended between four and eight times. However, five control students (2%) attended more frequently than the 8 days planned, one of whom attended nearly every day the program was open. When days of attendance for these five students were only counted when they attended the appropriate days to which they were technically invited (up to 8 days), the average attendance of the control group was 1.4 days.⁴

⁴Instrumental variables regression analyses that take into consideration the actual attendance patterns of study youths are discussed later. These analyses show that the control group attendance at the program did not influence the results of the study.

Although the level of participation in the experimental ASP by control group students was trivial, they were free to participate in whatever other after-school activities were available to them. Virtually all (96%) members of the control group participated in some organized after-school activity. Nearly 60% participated in an after-school activity at their schools, and most also participated in community-based, after-school activities. These alternative leisure activities are described in greater detail in the Results section.

Measures

Data from five sources are used for most analyses in this report: registration forms, a youth survey measuring primarily program outcomes, another youth survey measuring time expenditure during the after-school hours, teacher ratings, and school records. The percentage of subjects for whom data were obtained ranged from 87% to 100% for these data sources, with similar rates for treatment and control participants. Items from the first youth survey, teacher ratings, and school records were used to create outcome measures corresponding to the outcomes targeted by the ASP (Figure 1). Detail about the content of all outcome measures along with reliability coefficients are shown in Table A2 in the appendix, which is organized according to the outcomes shown in Figure 1. Gottfredson et al. (2010) provided the full item content of each scale and response formats used for each item as well as the published source for each scale. Each of the five data sources is described next.

Registration Form. Parents completed a registration form before their child began the program. This form was used to obtain demographic information as well as tracking information for those students who withdrew from their registered school during the course of the evaluation. Demographic information reported on the registration form included age, race, gender, grade, family income, and parental education.

Youth Surveys. Participating youths completed pretest and posttest youth surveys measuring primarily the outcomes targeted by the ASP. These surveys consisted of 167 items. Pretest surveys were administered to all treatment and control youth ($N = 447$) after receipt of registration materials and signed consent forms from their parents or caregivers. Posttests were administered near the end of the program. Surveys were typically administered during one school period in a large room such as the cafeteria or library. Students received a \$5 gift card for attending each survey administration. The response rate for posttest youth surveys was 96% ($N = 427$), although surveys from 11 of these respondents were later discarded because more than 40% of the items in the surveys were left blank, for a final response rate of 93%. Low study attrition is

at least partially attributed to a \$500 incentive offered to schools that achieved a 95% or higher response rate. All schools achieved this rate.

Youth Experiences Survey 2.0 (YES). Students completed the YES (D. M. Hansen & Larson, 2005) midway through the program year. This survey, which measured experiences during after-school activities, was administered in the same manner as the youth surveys except that no incentive was provided to the school. The YES response rate was 87% ($N = 389$).

School Records. School records were collected to measure student academic performance, attendance, and school suspensions for the year prior to the implementation of the program (2005–2006) as well as the year the program was implemented (2006–2007). At least one data element from these school records (both pre and post) was collected for all pretested youth. However, 2005–2006 GPA information was mostly unavailable for sixth graders, who were in elementary school during the 2005–2006 school year. GPA was not maintained electronically for these schools.

Teacher Ratings. During the spring of 2007, science, math, social studies, and English teachers were asked to rate 427 study participants.⁵ Teachers were offered \$5 for each survey completed. These surveys measured student classroom behavior, social adjustment, and academic competence. A total of 1,696 surveys were distributed to 192 school teachers. At the close of data collection, 65% of teachers ($N = 125$) had returned packets and 69% of student rating surveys ($N = 1,177$) were returned. At least one survey was returned for 99% of students, and two or more were returned for 88%.

Implementation measures. Data on the quantity and quality of services provided in the ASP was obtained using a Web-based management information system into which ASP staff entered information each day and via program observations conducted by UMD staff during 80 site visits. These data sources are described fully in Gottfredson et al. (2010).

Composite Scales. Our study includes multiple measures for several outcomes targeted by the program. Conducting multiple hypothesis tests for impacts at a given alpha level of significance increases the chance of Type I errors to greater than alpha unless adjustments are made for multiple comparisons. To guard against such chance findings, we follow the advice offered by an expert panel recently convened by U.S. Department of Education Institute of

⁵Twenty students had withdrawn from the BCPS system or had transferred to a new BCPS school too recently for the new teacher to rate the student reliably.

Educational Sciences to explore ways of appropriately handling multiple comparisons (Schochet, 2007). This panel recommended that the data be structured and outcomes be prioritized to reflect the design of the intervention and that confirmatory analyses be conducted to test global hypotheses within the main domains identified as central to the study's hypotheses. Accordingly, we developed scales to capture the eight outcomes identified in the program model (Figure 1).

Four outcomes (unsupervised socializing, positive peer influence, school bonding, and school attendance) were measured with a single indicator. The others (social competence, prosocial attitudes and beliefs, academic performance, and conduct problems) were measured with multiple indicators. These multiple indicators were combined to form four composite scales. Three of these scales: social competence, academic performance, and conduct problems (all ranging from -3 to $+3$) were computed by averaging the z scores of the component items and scales. The social competency composite scale included goal setting, decision-making skills, and impulsiveness. Prior to computing the average, the impulsiveness z score was reverse coded by multiplying the values by -1 . The average correlation among these scales was .28 at both pretest and posttest. The academic performance composite scale included teacher reports of academic competence, grade point average, Maryland School Assessment (MSA) reading, and MSA math scores (standardized test scores). The average correlation among these scales was .45 at pretest and .51 at posttest. The conduct problems composite scale included disruptive classroom behavior, aggression, delinquent behavior, victimization,⁶ last month drug use, number suspensions, and teacher reports of social competency. Prior to computing the average, the teacher reports of social competency z -score was reverse coded by multiplying the values by -1 . The average correlation among these scales was .30 at both pretest and posttest.⁷ Finally, the prosocial/antidrug attitudes composite scale (range = .00–1.00) was computed by averaging the attitudes unfavorable to drug use and belief in conventional rules scales. The correlation among these scales was .62 at pretest and .68 at posttest. Higher scores on all measures indicate a higher level of the outcome.

Attrition and Missing Data. Our primary source of outcome data, the posttest youth survey, was usable for 416 (93%) of the 447 registered students. The 31 students (13 treatment and 18 control) who were excluded from outcome

⁶Although not a measure of conduct problems per se, victimization is included in this composite because it is highly correlated with the other problem behavior measures (e.g., $r = .39$ and $.48$ for victimization and delinquent behavior at pretest and posttest, respectively). Evidence suggests that offenders and victims are often the same people.

⁷The pretest measures of the academic performance and the conduct problems composite scales exclude academic competence and social competency, respectively, because the scales from the teacher ratings survey were only measured at posttest.

analysis because of missing posttest data either refused to take the posttest ($n = 10$), had transferred out of Maryland schools ($n = 10$), or left more than 40% of the survey items blank ($n = 11$). An attrition analysis showed that youth who were excluded from the study ($n = 31$) did not generally differ from those who were included ($N = 416$), demographically or on a range of pretreatment measures. Exceptions were age and attitudes unfavorable to drug use. The excluded cases were older and had more favorable attitudes to drug use than those retained in the study.

Treatment by attrition interactions were conducted to test for differential attrition by treatment status that would bias the results of our study. Of 28 interactions, one (MSA math score) was statistically significant at the $p < .05$ level. This analysis suggested that higher achievers were more likely to attrit from the treatment than from the control group.

The amount of missing item-level data from the surveys is very low, in part because of our decision to compute scales based on all valid items. Missing data exceed 4% of the available cases only for posttest unsupervised socializing, pre- and post-decision-making skills measures, and the pretest data from school records. Therefore, most analyses simply excluded the small number of cases for which outcome data were missing. However, for the measures with more than 4% missing data, we employed maximum likelihood methods for imputing missing data (Allison, 2002). Outcome analyses involving these measures were conducted using both the imputed scores and listwise deletion. No substantive differences were observed across these two analyses, so we report only the results using the imputed scores.

Analysis Plan

Prior to comparing outcomes for the study groups, all outcome variables were examined to determine their best representation by identifying outliers and deviations from normality. Some variables were determined to be best represented as binary or count variables. When variables were transformed to reduce skew (for positive peer influence and school attendance), both the transformed and untransformed dependent variables were used in analyses testing for program effects. However, the results for the squared peer influence variable were similar to results using the untransformed variable, so the latter results are presented for the sake of simplification. In the case of school attendance, results using the logged variable are presented.

In all outcome analyses, two-tailed tests of statistical significance were employed, with an alpha level of .05. First, regression models were run. The model for each dependent variable included a dummy variable measuring assignment to the treatment condition (1 = treatment; 0 = control), a measure of the dependent variable taken at pretest, gender (1 = male; 0 = female), race (1 = Black; 0 = non-Black), age, and four dummy variables measuring

school site to correct for the clustering of individual cases within school. We used regression models appropriate for each dependent variable. Specifically, ordinary least squares models were used for normally distributed, continuous variables; negative binomial or Poisson regression was used for outcome variables that involved counts; and logistic regression was used for dichotomous outcome variables. Covariate-adjusted posttest means were calculated from these models. Standardized mean difference effects size statistics were calculated using the difference between the treatment and the control group adjusted posttest mean in the numerator and the pooled standard deviation for the corresponding unadjusted posttest measures.

RESULTS

Program Implementation

Gottfredson et al. (2010) and Cross, Gottfredson, Wilson, Rorie, and Connell (2010) report on the characteristics and training of the ASP staff, program management and social climate in the programs, and the quality and quantity of implementation of each program component. Briefly, 70% of staff had at least a BA degree, and the median years of experience working with youth was 5. One third of the staff were certified teachers. As was the case in the national evaluation of the 21st Century Community Learning Centers Program (Dynarski et al., 2003), staff turnover was common in this program. The typical staff member was employed for only 53% of the possible days. The original staff received more than 40 hr of training in all aspects of the program, but the replacement staff received less than 6 hr of training. Measures of program management and climate from program observations indicated considerable variability across sites. On measures of staffing and climate, two sites stood out as higher quality sites.

The All Stars prevention curriculum and academic assistance were each provided for approximately 1.5 hr per week, with the remaining time allocated to a variety of recreational activities (most often sports, board and card games, computer games, crafts, and dance). Almost half of the day was spent in life skills and academic activities on Tuesdays and Wednesdays. Thursdays consisted mainly of leisure activities. Overall, a little more than one third of the time was spent in life skills and academic activities, whereas the remaining two thirds consisted of leisure-based activities. Of the treatment youth who ever attended the ASP ($n = 205$), 72.2% participated in an academic activity. The average youth who ever participated in an academic activity ($n = 148$) received an average of 13.1 days of the academic assistance (range = 1–43). All Stars was well implemented by program staff, but student exposure was less than anticipated because of the dropout and low attendance (see next). The sites offered an average of 26 of the 27 available lessons (range = 23–27). Almost all of the

youth (91%) who ever attended the ASP ($n = 205$) participated in an All Stars session. Of those youth who ever participated in All Stars ($n = 187$), the average number of lessons received was 11.3 of the expected 27. The average hours received was 15.7 hr compared with 20.25 expected if All Stars were taught in twenty-seven 45-min sessions as recommended by the developer. Program content was similar to that reported for middle school programs that participated in the national evaluation of the 21st Century Community Learning Centers Program (Dynarski et al., 2003) except for the addition of the All Stars program.

Program Attendance

The ASP sites were open for 96 days beginning the 3rd week in September 2006 and running through May 2007. One hundred twenty youth, 54% of the treatment sample, withdrew from the ASP before the end of the year. When students withdrew, site staff noted the withdrawal date and the reason for withdrawal. The primary reason for dropout was voluntary withdrawal (63%). Another 20% were removed from enrollment by site staff because of very low or inconsistent attendance (usually after a month of unexplained absence). Two students were asked to leave because of behavior problems, 1 student moved, and the remaining 17 students withdrew for unknown reasons. Dropout varied significantly across site, and ranged from 32% to 70%.

Comparisons on all demographic and pretest measures showed that retained ($n = 104$) treatment students were more likely to be African American than withdrawn ($n = 120$) treatment students (77.0% vs. 62.0%; $p < .05$), and they were absent from school about 3 days less during the previous school year (6.0 and 8.4, respectively; $p < .01$).

Of the 96 possible days, the average days *enrolled* for treatment students was 54.2 days. Average days enrolled was (not surprisingly) higher for the retained students compared with withdrawn students (85.5 days and 27.0 days, respectively). The average days actually *attended* by all students was 35.6 days (37.1% of possible days and 55.0% of days enrolled). Days attended was also higher for retained students (61.0 days: 63.5% of possible days and 71.0% of days enrolled) compared with withdrawn students (13.5 days: 14.1% of possible days and 31.0% of days enrolled). Average days attended also varied by site, ranging from 29.2 to 45.6 ($p = .06$).

In summary, the level of withdrawal from the program was high and the rates of attendance were low, resulting in much lower levels of exposure than was intended. These figures are similar to attendance levels reported in other evaluations of ASPs. The evaluation of the 21st Century Community Learning Center programs, for example, found that 50% of students dropped out of the program and students attended an average of 32.5 days during the school year (Dynarski et al., 2003). An evaluation of the Communities that Care prevention system in which a variety of evidence-based program models were

implemented in community settings (Fagan, Hanson, Hawkins, & Arthur, 2008) also reported that exposure to the evidence-based programming was lower in after-school than in school settings. In that study, only 77% of participants receiving any programming received at least 60% of the sessions (compared with 96% in the school-based Communities that Care programs).

Control group contamination was low, but it did occur. As mentioned, five control students attended the ASP on more than 8 days. Program attendance for these students ranged between 11 and 89 days. Instrumental variables analysis presented later accounts for treatment exposure of these control students.

Estimated Impacts on Participation in ASPs and Activities

The ASP was intended to change the after-school activities of youth in the treatment group relative to youth in the control group who would have the "usual" after-school experience. Describing the activities of youth in the control group is important so that we can understand how their experiences differed from the treatment condition. The youth in the control condition were invited to the ASP on the last Thursday of every month, with the potential to attend the ASP eight times. On these days, youth participated only in recreational activities, as All Stars and academic activities were not offered. Of 223 control group youth, 48% ($n = 106$) attended the monthly ASP activities at least once. On average, control youth only attended 1.46 out of the possible 8 days (range = 0–8).

Of course, both treatment and control youths were free to participate in a variety of other after-school activities, both at school and elsewhere. The treatment group reported participating in ASPs at school more than the control group both in the YES (75% treatment vs. 56% control) and at posttest (67% treatment vs. 55% control). However, the percentage of control youth reporting attending an ASP in their school exceeded the 48% that could have been expected based on their attendance in the ASP previously reported. Therefore, at least some of the control youths must have attended an alternative ASP offered in their schools. Additional evidence that control youths participated in alternative programs is found in the youth's self-report at posttest on number of days spent in ASPs in the school. The treatment group reported spending about half a day more than control youth in ASPs at school (1.85 vs. 1.29). Although the intervention provided significantly more after-school programming to the treatment group, it seems that the control group was also able to find other activities at school in which to spend the after-school hours. What was the nature of these alternative activities?

Table 3 shows the proportion of youth reporting participation in all categories of activities as well as the number of activities reported, using responses to the YES survey. Overall, an equivalent proportion of youth in the treatment and control conditions reported participating in any activity during the after-school hours; 95% ($n = 187$) of treatment youth and 96% ($n = 185$) of control

Table 3. Participation in after-school activities, by experimental group

	Treatment ^a		Control ^b		<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Proportion reporting participation					
All activities	.95	.21	.96	.20	.83
Community/School-based activity	.23	.42	.25	.44	.66
Academic activity	.24	.43	.25	.43	.93
Performance/Fine arts activity	.45	.50	.48	.50	.52
Faith-based/Service activity	.34	.48	.33	.47	1.00
Sports activity	.72	.45	.73	.44	.72
No. of activities reported					
All activities (range = 0–22)	4.41	3.15	4.39	3.33	.98
Community/School-based activity (range = 0–6)	.30	.68	.31	.57	.94
Academic activity (range = 0–3)	.26	.48	.30	.57	.45
Performance/Fine arts activity (range = 0–5)	.63	.89	.70	.91	.37
Faith-based/Service activity (range = 0–4)	.40	.63	.36	.54	.68
Sports activity (range = 0–19)	2.08	2.03	2.16	2.64	.68

Source: YES survey.

^a $n = 196$. ^b $n = 193$.

youth reported participating in at least one activity after school. Treatment youth reported participating in an average of 4.41 different activities during the after-school hours compared with 4.39 for the control group. Independent-sample *t* tests showed no significant differences in treatment versus control reports of participation in any activity.

Youth in both groups reported participating in sports and performance/fine arts activities most often. Seventy-two percent of treatment youth and 73% of control youth reported participating in some form of sports after school (about 2.08 and 2.16 different sports on average, respectively). Forty-five percent of treatment youth reported participating in performance/fine arts activities compared with 48% of control youth (0.63 and 0.70 different performance activities, respectively). This analysis suggests that it is unlikely the ASP altered the variety of activities in which youths spent time during the after-school hours, although it did increase the amount of time spent in a school-based ASPs for treatment youths relative to controls.

Estimated Impacts on Student Outcomes

Table 4 presents the adjusted posttest means, significance level, and effect sizes comparing treatment and control youth for the five mediators and three outcome measures shown in the ASP program model (Figure 1). Posttest means were adjusted for the pretest measure of each variable, as well as for race, age, gender, and school site. The main effect for treatment reaches the $p < .05$

Table 4. Adjusted posttest means and effect sizes

Scale	Adjusted Posttest Mean		<i>p</i>	<i>d</i>
	Control (<i>n</i>)	Treatment (<i>n</i>)		
Intermediate outcomes				
Unsupervised socializing (days per week)	2.40(205)	1.89(211)	.01*	-.26
Positive peer influence	.79(202)	.78(210)	.54	-.05
School bonding	1.99(202)	2.00(210)	.85	.01
Social competence	-.01(194)	-.01(202)	.92	.01
Prosocial/Antidrug attitudes	.66(193)	.66(201)	.74	.03
Distal outcomes				
School attendance	.05(203)	.05(207)	.57	-.05
Academic performance	.01(205)	-.02(211)	.63	-.03
Conduct problems	-.01(205)	.00(211)	.84	.01

Note. Means are adjusted for pretest level of the dependent variable, race, age, gender, and school site.

**p* < .05.

level of statistical significance for only one measured outcome: unsupervised socializing. No significant differences between treatment and control youths were found on measures of conduct problems, academic performance, school attendance, prosocial/antidrug attitudes, social competence, school bonding, or positive peer influence. Effect sizes range from a high of $-.26$ for unsupervised socializing (indicating that the experimental ASP participants scored approximately one quarter of 1 *SD* lower, which is in the desired direction for this measure) to a low of $-.05$ for positive peer influence (indicating that the experimental ASP participants scored in the more negative or undesirable direction on this measure). The magnitude of the effect for the one difference that was statistically significant (unsupervised socializing) was small relative to our expectation: Youths attending the ASP reported being with their friends with no adults present for approximately one half day less per week than did control youths. We anticipated a larger difference on this outcome, given that the program ran 3 days per week. No Treatment \times School interactions were found for these eight outcomes.

Supplementary Analyses

Additional analyses were conducted to test for dosage and interaction effects. Gottfredson et al. (2010) reported results from instrumental variable regressions that estimated the effect of actual days of attendance (for treatment and control students) on the outcomes using random assignment as an instrument. Of course, given the overall null effects reported from the ITT analysis, the positive effects resulting from more days attended would imply that youths

who attended infrequently were harmed. Although unlikely, this pattern of effects is possible, especially given findings from earlier evaluations of relatively unstructured programs of a drop-in nature that have shown negative outcomes for ASP participants in comparison with nonparticipants (Dynarski et al., 2003; Mahoney, 2000; Weisman et al., 2002). These analyses showed that more days of actual attendance is not significantly related to any of the outcomes measured.

We explored the possibility that the program might have been more beneficial for certain subgroups of youths: latchkey, lower socioeconomic status, more at risk, and moderately at risk. We also investigated whether youth age interacted with program effectiveness, and we tested for conditional effects by level of program implementation measured in two different ways. In total, we conducted 56 (eight outcomes by seven potential moderator variables) tests for moderator effects. Only 1 produced a significant interaction, fewer than the number that would be expected by chance. Tests for interaction by program implementation quality yielded no significant differences on the eight outcome variables examined. Therefore, we conclude that although we observed variability across the sites in implementation quality, this variability was not related to program effectiveness.

DISCUSSION AND CONCLUSION

In this study, we tested the effectiveness of a reasonably well-structured, school-based ASP similar to the routine programs delivered by BCRP except for the addition of a research-based prevention curriculum. The ASP delivered attendance incentives, academic assistance, recreation, and a prevention curriculum. Staffing for the program was inconsistent, and student attendance was sporadic. The program as implemented resembled the middle school ASPs offered in the 21st Century Community Learning Center Programs. As such, the study adds to the existing research on ASPs (summarized earlier) by providing a much needed, well-implemented randomized trial of a typical ASP for middle school students.

Assignment to the treatment condition resulted in a substantial increase in the level of participation in both the experimental ASP and school-based ASPs more generally, relative to the control students. However, nearly all youth in both the treatment and control conditions reported participating in some organized activity during the after-school hours, and the number of different activities in which treatment and control youths reported being involved did not differ. Participation in the treatment reduced time spent with friends with no adults present during the after-school hours. The magnitude of this effect was small relative to our expectation: Youths attending the ASP reported being with their friends with no adults present for approximately half a day per week less than control youths. No significant posttest differences between

treatment and control youths were found on measures of conduct problems, academic performance, school attendance, prosocial/antidrug attitudes, social competence, school bonding, or positive peer influence. Frequent attenders did not have different outcomes than infrequent attenders, and the analysis of moderators showed that differential effects for subgroups of youth were detected in fewer instances than predicted by chance. Although the quality of implementation varied across the five implementing sites, we found no evidence that exposure to higher quality programs resulted in more beneficial outcomes for participating youths.

The results from this experiment are similar to results from the national evaluation of the 21st Century Community Learning Centers Program for middle school students (Dynarski et al., 2003, 2004). That study reported no effects of participation on self-care after school, homework completion, or feelings of safety after school. Middle school treatment group students had lower rates of school absenteeism than comparison group students, but their school grades were similar for the most part. They were also more likely to have had their property damaged, more likely to report they had used or sold drugs, and less likely to rate themselves positively at working out conflicts with others (Dynarski et al., 2003). However, that evaluation used a matched comparison group design and therefore could not rule out the possibility that the study outcomes were due at least in part to selection. This study provides a more rigorous empirical test of a middle school ASP that resembled those included in the 21st Century program study.

Results from this experiment are also similar to results from other experimental trials of ASPs. Zief et al. (2006), for example, limiting their meta-analysis to "well-implemented experimental design studies," found that of the 97 impacts measured by the five studies included in their review, 84% showed no significant differences between the program and control youth. As was the case in our study, Zief et al.'s positive findings were on measures of time expenditure in the after-school hours. Our only significant program effect was on unsupervised time spent with peers in the after-school hours. Zief et al. found no significant effects on academic or behavioral outcomes.

Our results are not consistent with prior research that has suggested that ASPs are effective under certain conditions. The research summarized earlier suggested that more positive effects would be observed for students who attended more (Kane, 2004), and in programs that were more structured, smaller, and staffed by highly educated staffs (Gottfredson et al., 2007). We did not observe stronger effects for students exposed to more of the program, for more at-risk students, or in sites characterized by higher quality implementation. Within the range of implementation quality observed in this study, none of the programs could be regarded as effective.

Recognizing that the results from our study do not generalize beyond the five participating sites, our conclusion is that programs like this are not strong enough to increase academic performance, reduce problem behavior or

school nonattendance, or influence any of the targeted intermediate behaviors and attitudes other than time expenditure. We therefore believe that it will be beneficial to explore alternative models for middle school ASPs. Qualitative impressions of the programs summarized in Cross et al. (2010) suggest that the most consistent attendance was achieved in the sites in which staff were more effective at creating emotional bonds with the youth participants. These observations are consistent not only with criminological theory that links social bonding with several prosocial outcomes (e.g., Hirschi, 1969) but also with prior reviews that have found that the most effective programs are those in which staff have more positive relationships with youth (Beckett et al., 2001; McComb & Scott-Little, 2003). This suggests that middle school youths may respond better to after-school activities that focus on developing bonds with prosocial adults. Such models might be organized more like mentoring activities such that a small number of youths might be connected to an adult who would help youths develop a particular skill or ability. The structure for these programs might be more fluid and flexible than the typical comprehensive ASP model, allowing youths to participate in a variety of competing activities as well. This model could be organized around much more focused activities (e.g., photography, acting and math) keyed to specific youth interests and could incorporate content shown to produce desired outcomes in prior research. But they would involve much closer relationships between the youths and the adults than is typical in a comprehensive ASP. Any one experience might last for a shorter duration, and youths might opt to participate in more than one throughout the school year. One of the activities offered in this format might be structured tutoring, which has been shown to improve literacy and math skills (Cawelti, 1999; Cohen, Kulik, & Kulik, 1982; Wasik, 1998), even in after-school settings (Lauer et al., 2006).

The idea behind our research—that incorporating more evidence-based programming into existing comprehensive ASPs for middle school youths will improve their effectiveness—continues to make sense but only if the programs can be delivered in such a way as to hold youths' interest. In this study, the main evidence-based program component, the All Stars prevention curriculum, was implemented in a reasonably high-quality fashion at all five sites. Staff at all five sites were trained to implement All Stars, and a high proportion of lessons at all sites were implemented by trained staff. But the typical student received only slightly more than half of the program lessons because of nonattendance, and the outcomes most directly targeted by the All Stars curriculum were no different for students who were and were not exposed to the program. This of course does not imply that All Stars and similar evidence-based prevention programs are not effective but rather that voluntary ASPs are not ideal settings in which to attempt to deliver such a program. Unless more stable staffing and more regular attendance can be achieved in the traditional, comprehensive ASP model most commonly provided, public dollars may be better spent improving services delivered during the regular school day where greater exposure and implementation quality are more likely (Fagan et al., 2008).

REFERENCES

- Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.
- Beckett, M., Hawken, A., & Jackowitz, A. (2001). *Accountability for after-school care: Devising standards and measuring adherence to them*. Santa Monica, CA: Rand Corporation.
- Cawelti, G. (Ed.). (1999). *Handbook of research on improving student achievement*. Arlington, VA: Educational Research Service.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237–248.
- Cross, A. B., Gottfredson, D. C., Wilson, D. M., Rorie, M. R., & Connell, N. M. (2010). Implementation quality and positive experiences in after-school programs. *American Journal of Community Psychology*. (Published Online March, 2010: DOI: 10.1007/s10464-010-9295-z.)
- Dupont, W. D., & Plummer, W. D. (1990). Power and sample size calculations: A review and computer program. *Controlled Clinical Trials*, 11, 116–128.
- Dupont, W. D., & Plummer, W. D. (1998). Power and sample size calculations for studies involving linear regression. *Controlled Clinical Trials*, 19, 589–601.
- Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago: Collaborative for Academic, Social, and Emotional Learning.
- Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., & Mansfield, W. (2003). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program: Final report* (U.S. Department of Education, National Center for Education Evaluation and Regional Assistance). Washington, DC: U.S. Government Printing Office.
- Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., & Mansfield, W. (2004). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program: New findings* (U.S. Department of Education, National Center for Education Evaluation and Regional Assistance). Washington, DC: U.S. Government Printing Office.
- Eccles, J. S., & Templeton, J. (2002). Extracurricular and other after-school activities for youth. *Review of Research in Education*, 26, 113–180.
- Fabiano, L., Pearson, L. M., & Williams, J. J. (2005). *Putting students on a pathway to academic and social success: Phase III findings of the Citizen Schools evaluation*. Washington, DC: Policy Studies Associates.
- Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the Community Youth Development study. *American Journal of Community Psychology*, 41, 235–249.
- Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk.
- The Finance Project. (2007). *Estimated federal investment in out-of-school time*. Washington, DC: Author. Retrieved April 2008 from http://www.financeproject.org/publications/estimatedfederalOST_TFPflyer.pdf
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D. C., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151–176.
- Girod, M., Martineau, J., & Zhao, Y. (2004). After-school computer clubhouses and at risk teens. *American Secondary Education*, 32, 63–76.
- Gottfredson, D. C., Cross, A. B., & Soulé, D. A. (2007). Distinguishing characteristics of effective and ineffective after-school programs to prevent delinquency and victimization. *Criminology & Public Policy*, 6, 601–631.
- Gottfredson, D. C., Cross, A. B., Wilson, D. M., Rorie, M., & Connell, N. (2010). *A randomized trial of the effects of an enhanced after-school program for middle school students*. Final report submitted to the U.S. Department of Education Institute for Educational Sciences.
- Gottfredson, D. C., Gerstenblith, S. A., Soulé, D. A., Womer, S. C., & Lu, S. (2004). Do ASPs reduce delinquency? *Prevention Science*, 5, 253–266.
- Gottfredson, D. C., Gottfredson, G. D., & Weisman, S. A. (2001). The timing of delinquent behavior and its implications for ASPs. *Criminology & Public Policy*, 1(1), 61–80.
- Granger, R. C., Durlak, J. A., Yohalem, N., & Reisner E. (2007). *Improving after-school program quality* (Working paper). New York: William T. Grant Foundation.
- Hansen, D. M., & Larson, R. (2005). *The Youth Experience Survey 2.0: Instrument revisions and validity testing*. Unpublished manuscript. Retrieved January 2008 from [http://web.aces.uiuc.edu/youthdev/UnpublishedManuscriptonYES2111%20\(2\).doc](http://web.aces.uiuc.edu/youthdev/UnpublishedManuscriptonYES2111%20(2).doc)
- Hansen, W. B. (1996). Pilot test results comparing the All Stars program with seventh grade DARE: Program integrity and mediating variable analysis. *Substance Use and Misuse*, 31, 1359–1377.
- Hansen, W. B., & Dusenbury, L. (2004). All Stars Plus: A competence and motivation enhancement approach to prevention. *Health Education*, 104, 371–381.
- Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J., & Youngbluth, S. C. (2001). Evaluation of the All Stars™ character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. *Health Education Research*, 28, 533–546.
- Hirschi, T. (1969). *Causes of delinquency*. Berkeley: University of California Press.
- Hollister, R. G. (2003). *The growth in after-school programs and their impact*. Washington, DC: Brookings Institution.
- Huang, D., Sung Kim, K., Marshall, A., & Perez, P. (2005). *Keeping kids in school: An LA's BEST example*. National Center for Research on Evaluation, Standards, and Student Testing Center, Center for the Study of Evaluation, Graduate School of Education and Information Studies, University of California, Los Angeles.
- Kane, T. J. (2004). *The impact of after school programs: Interpreting the results of four recent evaluations* (Working paper). New York: William T. Grant Foundation.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76, 275–313.
- Lauver, S. C. (2002). *Assessing the benefits of an after-school program for urban youth: An impact and process evaluation* (Doctoral dissertation, University of Pennsylvania, 2002). *Dissertation Abstracts International*, 63(02), 553A.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social science research methods*. (pp. 44-76). Thousand Oaks, CA: Sage.

Mahoney, J. L. (2000). School extracurricular activity participation as a moderator in the development of antisocial patterns. *Child Development, 71*, 502-516.

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

McNeal, R. B., Hansen, W. B., Harrington, N. G., & Giles, S. M. (2004). How All Stars™ works: An examination of program effects on mediating variables. *Health Education and Behavior, 31*, 165-178.

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

Schochet, P. Z. (2007). *Guidelines for multiple testing in experimental evaluations of educational interventions*. Princeton, NJ: Mathematica Policy Research, Inc.

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

Shelton, D. (2008). Translating theory into practice: Results of a 2-year trial for the LEAD programme. *Journal of Psychiatric and Mental Health Nursing, 15*, 313-321.

Sickmund, M., Snyder, H. N., & Poe-Yamagata, E. (1997). *Juvenile offenders and victims: 1997 update on violence*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Smith, C., & Kennedy S. D. (1991). *Final impact evaluation of the Friendly PEERSuasion program for Girls Incorporated*. New York: Girls Incorporated.

St. Pierre, T. L., Mark, M. M., Kaltreider, D. L., & Aikin, K. J. (1997). Involving parents of high risk youth in drug prevention: A three-year longitudinal study in Boys & Girls clubs. *Journal of Early Adolescence, 17*, 21-50.

Wasik, B. A. (1998). Using volunteers as reading tutors: Guidelines for successful practices. *The Reading Teacher, 51*, 562-570.

Weisman, S. A., Womer, S. C., Kellstrom, M. A., Bryner, S. L., Kahler, A., Slocum, L., et al (2002). *Maryland after school community grant program, report on the 2001-2002 school year evaluation of the phase 3 after school programs* (Technical Rep.). (Available from dgottfredson@crim.umd.edu)

Zief, S. G., Lauver, S., & Maynard, R. (2006). *The impacts of after-school programs on student outcomes: A systematic review for the Campbell Collaboration*. Retrieved April 2008 from http://www.campbellcollaboration.org/doc-pdf/zief_afterschool_review.pdf

APPENDIX

Table A1. Summary of previous research on after-school programs serving middle school youth

Author	Intervention Duration	Design	Sample	% Attrition/ Potential Differential Attrition Bias	Outcomes Reported	Results	Favorable?
Dynarski et al (2004)	21st Century Community Learning Centers, 2 years	Pre-Post, NECC, adequate pretest controls	1,782 TX 2,482 C	91% TX 90% C/no	Self-care, location after school, days staying after school for activities, activity participation, teacher reports of effort, school discipline records, homework habits, educational aspirations, social and emotional outcomes, feelings of safety, negative behaviors, victimization	Treatment effects for days after school for activities, participation in lessons and clubs, less likely to be with sibling after school, better school attendance, social studies grade. Also increase in negative behavior composite scale for the treatment group.	D, N, A
Fabiano, Pearson, and Williams (2004)	Citizen Schools Program, 3 years	Pre-Post, NECC, adequate pretest controls	855 TX 855 C	no data	Attendance, suspension, promotion to next grade, math and English grades, standardized test score in English and math	6th & 7th graders: Treatment effect attendance, suspension, promotion to next grade, test score in English 8th graders: Treatment effect attendance, promotion to next grade	D, N
Girod, Marineau, and Zhao (2004)	CLICK! After-school technology club, 2 years	Pre-Post, NECC, lacking adequate pretest controls	231 total at pretest; at posttest 81 TX, 78 C	31% / yes	Student-teacher relations, readiness for classroom instruction, perceived parental involvement in school, overall school value, and experience using computer technologies	Treatment effect on valuing school and experience with computers	N

Table A1. Summary of previous research on after-school programs serving middle school youth (*Continued*)

Author	Intervention Duration	Design	Sample	% Attrition/ Potential Differential Attrition Bias	Outcomes Reported	Results	Favorable?
Gottfredson et al. (2004)	Maryland After School Community Grant Program, programs in operation for various durations	Pre-Post, NECG, adequate pretest controls	Older Sample— 239 TX 201 C	Overall older and younger samples— 11% TX 13% C/no	Delinquent behavior, rebellious behavior, last-year drug use, intentions not to use drugs, hours/week in self-care, involvement in constructive activities, social skills, positive peers, peer drug models	Treatment effects for constructive activities, drug-using peers, and last year drug use—Structural equations model finds treatment effect for latent “delinquent behavior”	D, N, A
Gottfredson et al. (2007)	Maryland After School Opportunities Fund Program, programs in operation for various durations	Pre-Post, NECG, adequate pretest controls	389 TX 108 C	41% TX 31% C/yes	Delinquency, victimization, and substance use	Treatment effect for delinquency	D, N
Huang et al. (2005)	LA’s Best Program, three years	Pre-Post, NECG, adequate pretest controls	5,827 TX 5,816 C	0%	School dropout	Treatment effect for dropout	D, N, A
Lauver (2002)	After-school recreation program, three years	Pre-post, RCT	126 TX 101 C	3%/ yes	Constructive activities, self-care, time spent on homework, educational aspirations, attendance, grades, standardized test scores	Treatment effects for participation in fitness activities, time spent on homework, educational aspirations	D, N, A
Prenovost (2001)	After-school learning program, first year of program operation	Pre-Post, NECG, lacking adequate pretest controls	300 High-Dose 304 Low-Dose 828 C	No data	Standardized reading and math test scores, study effort, school attendance, feelings of safety at school	High-dose treatment group improved more in attendance	N
Shelton (2008)	LEAD expressive art program, 14 weeks	Pre-Post, NECG, adequate pretest controls	46 TX 43 C	0%	Self-esteem, resilience, behavioral self-control, and protective factors	No treatment effects	D, A
Smith and Kennedy (1991)	Friendly PEERsuasion program, 14 weeks	RCT, Pre-Post	All girls. 152 TX 202 C	17% TX 19% C/no	Avoiding substance use, leaving situations in which substances are being used	Study used critical value of $p < .10$. Treatment effects on drinking, drinking initiation, and leaving situations where drinking was occurring	D, N, A
St. Pierre, Mark, Kaltreider, and Aikin (1997)	Boys and Girls Clubs implementing a drug prevention program, 3 years	Pre-Post, NECG, adequate pretest controls	411 TX 105 C	39% TX 46% C/yes	Basic social skills, drug knowledge, attitudes about drugs, drug use, and drug refusal skills	Treatment effects for drug refusal skills, drug knowledge and drug attitudes. No effects on drug use outcomes	D, N
Weisman et al. (2002)	Maryland After School Community Grant Program, programs in operation for various durations	Pre-Post, NECG, adequate pretest controls	594 TX 476 C	21% TX 23% C/no	Social skills, GPA, rebelliousness, commitment, intentions not to use drugs	Iatrogenic effects for social skills, GPA, rebelliousness, commitment, but positive treatment effect for Intentions not to use drugs	D, N, A

Note. Includes only studies of programs delivered primarily during after-school hours and that included more than 10 sessions. NECG = nonequivalent comparison group; TX = treatment group; C = comparison group; D = meets criteria for sound research design having either a randomized design or using sufficient controls on identified pretest differences between groups; N = meets criteria for sufficient sample size having more than 100 participants per experimental group; A = no unaddressed problems with attrition, coded when attrition was < 20%, or evidence was presented demonstrating that differential attrition was not introducing bias; RCT = randomized, controlled trial; GPA = grade point average.