

Implementation Quality and Positive Experiences in After-School Programs

Amanda Brown Cross · Denise C. Gottfredson ·
Denise M. Wilson · Melissa Rorie · Nadine Connell

Published online: 14 March 2010
© Society for Community Research and Action 2010

Abstract Data collected during an evaluation of a multi-site trial of an enhanced after-school program were used to relate quality of program implementation to student experiences after school. The enhanced after-school program incorporated a drug use and violence prevention component that was shown to be effective in previous research. Building on Durlak and Dupre's (Am J Community Psychol 41:327–350, 2008) dimensions of implementation, we assessed the level of dosage, quality of management and climate, participant responsiveness, and staffing quality achieved at the five program sites. We evaluated how these characteristics co-varied with self-reported positive experiences after-school. The study illustrates how multiple dimensions of program implementation can be measured, and shows that some but not all dimensions of implementation are related to the quality of student after-school experiences. Measures of quality of management and climate, participant responsiveness, and staffing stability were most clearly associated with youth experiences. The importance of measuring multiple dimensions of program implementation in intervention research is discussed.

Keywords After-school programs · Implementation · Evidence based practice

Introduction

After-school programming has been increasing in the U.S. Considerable federal, state, local, and private monies are being invested in these programs. For example, the Twenty-First Century Community Learning Center Program received approximately one billion dollars in federal funds annually from 2002 to 2008 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 (financeproject.org, 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 et al. 2001; Gottfredson et al. 2004; Kane 2004; Lauer et al. 2006). Concerns about delinquency prevention are also linked to demand for ASPs, as the after-school hours present the highest risk of arrest for juveniles (Gottfredson et al. 2001; Sickmund et al. 1997). The intuitive appeal of ASPs rests on the perception that unsupervised after-school time is either dangerous or simply wasted time for adolescents. ASPs may provide an opportunity to enhance learning, to introduce positive role models, and to provide shelter from unsafe neighborhoods, especially for low-income children in urban areas. ASPs are also a convenient platform on which to provide social and personal skills instruction that may not be provided during the school day.

Addressing these objectives via ASPs presents the same hurdles faced by all community- and school-based intervention strategies, such as recruitment and retention of participants, determining the needs of the target group and setting reasonable goals for change, hiring, training and maintaining well-qualified staff, formulating and

A. B. Cross (✉) · D. C. Gottfredson · D. M. Wilson · M. Rorie
Department of Criminology and Criminal Justice, University
of Maryland, 2220 LeFrak Hall, College Park,
MD 20742-8235, USA
e-mail: across@crim.umd.edu

N. Connell
Rowan University, Glassboro, NJ, USA

implementing a successful curriculum or tailoring an existing curriculum to suit the specific population and goals of the program, and gaining the support of community and governmental agencies. This process is more efficient when best practice recommendations are available, but best practices research on ASPs is still in its infancy. The picture painted by existing research on ASPs is one of tremendous heterogeneity, both in terms of programming and outcomes.

Recent reviews on effectiveness of after-school programming generally agree that, “ASPs are capable of improving important youth outcomes” (Granger et al. 2007, p. 3), but very little can confidently be said about specific program content associated with success. Additionally, many programs have had no effect on youth outcomes and in some cases, ASP participants experienced negative outcomes (e.g., conduct problems, increased substance use, and negative peer influence) in comparison to non-participants (Dynarski et al. 2003; Mahoney 2000; Weisman et al. 2002). We know that ASPs can contribute to positive development but many programs have failed to do so.

Despite lack of specific recommendations for content, several studies have suggested that incorporation of evidence-based practices increases program effectiveness. In a research initiative spanning several years of ASP evaluation in Maryland, the authors found a consistent pattern where ASPs that emphasized social skills instruction were more successful in improving a variety of youth outcomes than those that did not (Gerstenblith et al. 2005; Gottfredson et al. 2004; Weisman et al. 2002). A study of a different statewide after-school initiative in Maryland also found that programs which incorporated published curricula were more effective in reducing youth substance use (Gottfredson et al. 2007). A meta-analytic study of ASPs targeting personal or social skills found that, on average, ASPs had a positive impact on school bonding, attitudes about self-efficacy and self-esteem, behavioral adjustment indicators (pro-social and anti-social behaviors as well as drug use) and school performance. But this was only true in programs that used evidence-based skill training approaches. Programs that did not include evidenced-based approaches were unsuccessful in improving any outcome (Durlak and Weissberg 2007).

Variability in implementation fidelity may be another key to understanding why some programs work and others do not. Implementation quality is necessarily a characteristic of all interventions, but it is underutilized as an explanatory variable when assessing outcomes (Dusenbury et al. 2003). In a review of literature on prevention and promotion interventions that measured the effect of implementation characteristics on outcomes, Durlak and Dupre (2008) found that factors related to implementation

had a consistent effect on outcomes. The authors identified eight aspects of implementation and summarized the findings from 59 empirical studies that quantitatively examined the link between one or more of these aspects and program outcomes. They found significant, positive relationships between level of implementation and outcome in 76% of cases. In the majority of cases where no effect of implementation was detected, the authors noted that lack of variability in implementation could explain the null findings.

It is clear that implementation quality is an important determinant of the effectiveness of interventions. Knowing that poor to moderate implementation quality is the norm in intervention innovation (Gottfredson and Gottfredson 2002), researchers should take full advantage of the interventions planned for the near future or those now underway by seizing the opportunity to carefully document multiple dimensions of implementation quality. Variability in implementation is especially likely to influence outcomes in studies designed to assess program effectiveness under “real world” conditions (e.g., “effectiveness trials”) as opposed to those in which the researcher has tighter control over implementation conditions (e.g., “efficacy trials”). This study reports on a multi-site ASP intervention in which an “enhanced” program model was provided to practitioners who routinely delivered ASPs in the state of Maryland. The intent of the overall study was to assess the extent to which the routine practices of the implementing agency could be shifted in the direction of providing more research-based programming, and to measure the effects of doing so on a range of youth outcomes. Program effects on youth outcomes are reported elsewhere (Gottfredson et al., in press). This article focuses on the quality of program implementation and its association with youth reports of the quality of their experiences during the after-school hours. Although the goal of the study was to achieve standard implementation of the enhanced program model in all participating sites, we anticipated that actual implementation would vary both from the planned program model and across sites. Measuring this variability was a major focus of the overall study. The current report uses data from program logs, observations, and self-report surveys collected during this evaluation to examine variability in program implementation and to relate this variability to the after-school experiences of youth at each site.

The remainder of this article is organized as follows: First, we present the model for the enhanced after-school program. Second, following Durlak and Dupre’s (2008) articulation of dimensions of implementation, we summarize data on the level of implementation achieved at the five program sites. Third, we evaluate how this variability in implementation relates to the quality of student experiences after school. We hypothesize that programs which

were implemented well as reflected in high attendance, observed quality of management and climate, that offered activities which held students' attention, and that were staffed by a stable group of educated and trained adults would produce more positive experiences for youth. Analyses relating level of implementation with youth experiences after school are descriptive in nature. The small number of program sites renders statistical comparisons inadvisable.

Method

The Enhanced After-School Program Model

Following previous research on intervention effectiveness, we designed an enhanced after-school program for implementation in five low-performing middle schools in an urban, East coast school district during the 2006–2007 school year. The recruitment goal for the program was 100 participants per school, 50 of whom would be randomly assigned to the intervention group and 50 of whom would serve as controls, for a total of 500 research participants. Intervention students were invited to attend the enhanced ASP and control students were invited to a fun activity (typically a pizza party) at the program once per month. The program was to follow a traditional structure. It would be offered on school grounds, 3 days per week, for 3 h after the close of the regular school day. The enhanced program was intended to improve on traditional ASPs through the introduction of research-based program content, the *All Stars* curriculum, which was to occupy 1½ h of program time per week.

The All Stars prevention curriculum is designed to delay the onset of and prevent substance use and other high-risk behaviors among adolescents. It targets mediators which are known to correlate with drug use such as normative beliefs about drug use, incongruence of drug use and goal achievement, commitment to abstain from drugs, and bonding to school. Program designers recommend delivering the program in school or community settings. Previous tests have shown All Stars to be effective in reducing tobacco, alcohol, and inhalant use and improving related attitudinal outcomes for middle school students (Harrington et al. 2001; McNeal et al. 2004; Hansen and Dusenbury 2004). The U.S. Department of Education (<http://www.ed.gov/admins/lead/safety/exemplary01/exemplary01.pdf>) and the Substance Abuse and Mental Health Services Administration recognize All Stars as a model program (<http://www.modelprograms.samhsa.gov/model.htm>).

The implementation plan called for All Stars lessons to be delivered each week in two 45-min sessions. Twenty-seven separate All Stars lessons were available to site staff.

Each lesson was divided so that one lesson could be delivered in two sessions. One or more staff members from each site participated in a 3-day training delivered by the developer of the All Stars curriculum.

The enhanced ASP also included an attendance incentive system based on a token economy model in attempt to boost student attendance. High levels of attendance in elective ASPs are historically difficult to achieve (Grossman et al. 2007; Weiss et al. 2005). As planned, the attendance incentive system would award students weekly points contingent on school and ASP attendance. These points could be exchanged for a variety of prizes.

The sites also offered academic support for 1½ h a week which consisted primarily of supervised homework assistance. A more advanced model for academic assistance which included one-on-one tutoring was planned, but this model was not implemented. Finally, sites offered a range of leisure activities such as fitness activities, board games, arts and crafts, field trips, computer projects or computer free time, service learning, workforce skills and holiday or other special event celebrations. However, the bulk of leisure activities regularly available at sites consisted primarily of fitness activities, board games and computer free time.

The intervention was implemented by a contracted vendor which was a county-level government agency that specialized in providing recreation and leisure activities for youths. The vendor was responsible for managing the day-to-day operations of the program sites, hiring and supervising all program staff, carrying out the All Stars program, academic assistance and attendance incentives, and providing leisure activities. The researchers arranged staff training in All Stars and attendance incentives while the vendor provided staff training for all other aspects of the program which were implemented. Participant recruitment efforts were undertaken by both researchers and the vendor. Data collection responsibilities were also shared. Researchers conducted on-site observations and administered surveys to youth. Vendor staff entered process data into a web-based management information system each day the programs operated.

Dimensions of Implementation

Durlak and Dupre (2008), drawing largely on an earlier analysis by Dane and Schneider (1998), described eight dimensions of implementation. Three of these are relevant to our study: (1) *Dosage*, how much of the program has been delivered, (2) *Quality*, how well different program components have been conducted, and (3) *Participant responsiveness*, the degree to which the program stimulated the interest or held the attention of participants. We also

assessed a fourth area of implementation not addressed by Durlak and Dupre. We measured three characteristics of staff quality, each having been identified in previous research as salient to program effectiveness: *staff turnover* (Armstrong and Armstrong 2004), *education* (Gottfredson et al. 2007; Rosenthal and Vandell 1996) and *training* (Armstrong and Armstrong 2004; Fashola 1998). In this article, we relied on the four dimensions introduced above as an organizing framework.

Sample

All under-performing middle schools in the district were consulted about the opportunity to run the free program in their schools. Schools invited to participate had low math and reading standardized test scores relative to the rest of the county and the state. The five school sites selected were the first to express interest and agree to cooperate with the research procedures. Student recruitment efforts began in the spring of 2006. Registration was open to all students who attended the participating schools but principals were asked to encourage youths who they considered especially “at-risk” to register. The participating schools served high percentages of minority youth (47–99% minority population) and large numbers of students who received subsidized meals (64–67% receiving free or reduced-price lunch).

When recruitment ended in January of 2007, 447 students were registered for the program, 224 of whom were randomly assigned to the intervention group. Because this article examines how variability in program implementation across sites co-varies with student experiences after school, only members of the intervention group are included in the analysis. About half of this sample were males (53%), 71% were African Americans, 17% were Caucasian, 8% were multi-racial and the remaining 4% were of another race. The average age for participants was 12.3 (SD = 1.0), more than half of students received free or reduced meals at school (58%), and 6th graders were the most likely to register (42%) while 8th graders were least likely (25%).

Measures

Implementation

Data were collected from attendance records, a general program observation, a student engagement observation, and employment records. Attendance records, which measured program *dose*, were entered into a web-based management information system daily at each site. A check of management information system records against a randomly selected set of paper attendance records showed high agreement ($r = .98$) across sources.

Observation measures were used to assess quality of programming (in terms of *program management and climate*) and *student engagement*. A team of research assistants conducted observations during 80 site visits between October 2006 and April 2007. Observers generally traveled to the sites in pairs, rotating among sites, but occasionally a site visit was conducted by one person. Two observers were present at 64 of the 80 observations (80%). Site E was observed 14 times, sites A, C, and D were observed 16 times and B was observed 18 times.

The observation protocol directed observers to complete one overall program observation and two engagement observations per visit. One engagement observation was to be completed during All Stars or academics, and the second during a leisure activity. In this way, we hoped to capture a balance of high and low structure activities.

The program observation instrument addressed quality of program management and social climate by measuring structure, supervision, social climate, behavior management, and leader skill. These measures were developed based on observations used in an earlier study of an after-school initiative (Gerstenblith et al. 2005), where they were shown to be related to program effectiveness. The domains of management and climate were intercorrelated so we studied them globally. Observation team members were instructed to confer about their observations at the end of a site visit and complete the program observation together, resolving any disagreements through discussion. Nineteen items from this instrument were dichotomized and averaged to create a management and climate scale with a mean of .60 (SD = .26) and alpha reliability of .87. The scale contained four items measuring supervision (e.g., “Are there ever opportunities for youths to leave the program activities and go to an unsupervised area?”), five items measuring social climate (e.g., “Do you see any evidence of friction between youth and program staff?”), eight items measuring structure (e.g., “Activities seem to be planned well in advance, with very little improvisation.”), one item assessed sound behavior management (misbehavior was observed infrequently), and the final item assessed the observers’ impression of the skillfulness of content delivery overall. Scores on the scale ranged from 0 to 1 and reflected the proportion of items assessed favorably.

The engagement observation counted the number of students engaged and not engaged during each 5 min interval in discrete program activities. Students were considered engaged when they were attending to the assigned activity instead of unrelated tasks or socializing. Engagement rates for each activity were calculated based on the sum of students judged to be engaged across intervals divided by the total number of student observations. These student engagement rates were averaged across all observations to create a mean level of student engagement for

each site. Inter-rater reliability (IRR; cross-observer agreement) for the engagement rates was established during the first month of observations. Analyzing 20 pairs of observations, IRRs for the engagement rates were .88. Data on *staff turnover*, *training*, and *education* were gathered from employment records. Items include days worked, hours of training, and education level.

Youth Self-Reported Experiences

Youth self-report of quality of experiences was measured by the Youth Experiences Survey (YES; Hansen and Larson 2005). This survey was administered to students in January and intended to describe the type and quality of after-school activities and addressed enjoyment and positive experiences in the ASP. The instrument measured six dimensions of positive development experiences: identity, initiative, basic competencies, teamwork and social skills, positive relationships, and adult networks and social capital. Items on the YES directed students to indicate the extent to which they experienced a variety of situations in their leisure activities (e.g., “I had the opportunity to be in charge of a group”, “I practiced self-discipline”, “This activity helped prepare me for college”). The survey contained 66 questions, 53 of which were used to assess positive experiences. All questions had a four-item response set in which a “1” indicated that the youth did not have the experience at all and a “4” indicated that the youth definitely had the experience. Responses to these items were averaged to create a YES positive experiences scale. Alpha reliability for the scale was high at .97. See Hansen and Larson (2005) for a detailed description of the instrument. The response rate for the YES was 85% ($N = 189$) of the intervention youth, 92% ($N = 173$) of whom attended the after-school program at least once. The YES positive experiences scale average is based on responses from these 173 youths.

Procedures

In order to describe the success of implementation we must compare the level of implementation at each site to some type of standard. Unfortunately, no established standards for ASP implementation yet exist. For example, reviewing evidence about how attendance relates to success in out-of-school programs, the Harvard Family Research Project (Simpkins-Chaput et al. 2004) concluded that it is impossible to make statements about how much attendance is required to improve outcomes. The authors noted that existing program evaluations do not measure attendance with a common metric, and that the large variety of out-of-school programs available to youth operate in very different timeframes. While some studies found that high attendance is associated with the best outcomes, other studies found that

students exposed to a moderate amount of programming have the best outcomes (Simpkins-Chaput et al. 2004).

In the absence of agreed upon standards, it is difficult to define “good” or “poor” implementation in the abstract. One alternative is to compare the level of achieved implementation to the ideal (e.g., 100% attendance, zero staff turn-over). Doing so would paint an unrealistically grim picture of the success of implementation because we know that ideal implementation is virtually never achieved (Durlak and Dupre 2008). Another alternative is to compare the levels of implementation to that of a fairly typical program, such as the Twenty-First Century Community Learning Centers. Doing this would paint an unrealistically sunny picture because such “run-of-the-mill” programs have not been particularly ambitious in pursuing high quality implementation. Another possibility is to compare the sites to each other. While failing to provide an absolute evaluation of the quality of program implementation, this approach at least provides a relative ranking that can be correlated in our sample with relative rankings on our measure of youth experiences in the after-school hours. We opted for this relative approach, and will also comment on absolute level of implementation quality when there is a reasonable basis for doing so. We demonstrate the level of implementation achieved on each of the dimensions and assess whether each site achieved high, moderate or low success relative to the remaining sites. This is indicated in the tables which follow by “+” for high, “0” for medium, and “-” for low success. We then compare the average YES score at each site to the assessment of success on dimensions of implementation.

Results

Dosage

For the analysis of implementation dosage, we used the mean number of days that students actually attended the program. Days of attendance for individual students ranged from 0 to 94. Table 1 displays the average number of days that students attended each site and shows that the typical student attended 36.7 days ($SD = 29.4$) (out of 96 possible days). As illustrated in Table 1, Site B students attended considerably more days (47.0) than students at other sites. Sites A and C had the lowest attendance. Students at these sites attended the program approximately 30 days.

Program Management and Climate

Results for program management and climate (Table 1) showed Sites B, C and E were rated positively, with favorable ratings higher than 70%. Site A performed poorly in

Table 1 Levels of implementation dose, engagement, and management and climate, by site and overall

Site	Days attended	Rank	Engagement rate	Rank	Management and climate	Rank
A	31.5	–	.71	0	.50	0
B	47.0	+	.81	+	.71	+
C	29.6	–	.79	+	.69	+
D	36.3	0	.73	0	.39	–
E	35.4	0	.80	+	.70	+
Overall	36.7		.77		.60	

Note: “+” indicates high success, “0” indicates moderate success, and “–” indicates low success

management and climate, with only half of the scale’s items assessed favorably. Site D’s performance was particularly weak; it received favorable ratings on only 39% of the items.

Student Engagement

The overall engagement rate across all sites was .77 (see Table 1). Sites B, C and E had very similar average rates of engagement ranging from .79 to .81. Programming at Sites A and D was less engaging to youth, where students focused on activities offered only at a rate of about .70.

We looked more closely at engagement to determine if certain activities were more engaging than others. We compared the six activity types which were observed for engagement on more than five occasions: All Stars, academic assistance, fitness activities, arts and crafts, board games, and computer time. Academic assistance stood out prominently as the least engaging activity. While five of the six activities examined had engagement rates of .80 or higher, the rate for academics was only .52. These data make clear that, despite efforts to implement similar programs across the different sites, considerable heterogeneity in dosage, quality of management and social climate, and engagement existed.

Staff Turnover, Training and Education

The program design called for a site director and three assistants at each of the five sites. This level of staffing was not achieved. Only 14 of 20 direct services positions were filled when the programs opened. This initial level of staffing was not regarded as problematic by the vendor

because the student population was not yet at capacity at the start of the program. Thirteen individuals were hired after the beginning of the program either to fill vacancies or to replace lost staff. These new staff members did not receive the intensive start-up training that the original staff received. Six of the original fourteen staff members quit or were fired before the end of the year. Three staff members were relocated to new sites mid-year. Only six direct services staff worked at the site to which they were originally assigned for the entire program.

On average, the 27 staff members worked at programs on 56.1 days, 58% of the 96 days the programs operated. Staff at Sites B and E worked in their positions for more than 60 days on average while staff at Sites A and D worked fewer average days, 35 and 43 days respectively. Site C staff worked an average of 53 days (Table 2).

On average, staff across sites received 24.7 h of job training, but this figure was far higher for original staff. The 14 original staff members received more than 40 h of training on average, while the 13 replacement staff members received less than 6 h. Consequently, sites where turnover was higher tended to employ fewer highly-trained staff. Table 2 shows that staff at Site A received the least training and staff at Site E received the most, followed closely by Sites D and B. Site C again evidenced a moderate level of this characteristic. Although Site D experienced substantial staff turnover, average staff training was high. This is because two staff members at that site were original staff who attended the complete start-up training but initially worked at other sites. Due to turnover at Site D, these two staff members were removed from their original sites and reassigned to Site D.

Table 2 Staff turnover, training and education, by site and overall

Site	Days worked	Rank	Hours of training	Rank	% BA or higher	Rank
A	35.1	–	13.7	–	75.0	+
B	61.0	+	27.4	+	80.0	+
C	53.4	0	22.5	0	80.0	+
D	43.2	–	29.8	+	42.9	–
E	65.4	+	32.6	+	80.0	+
Overall	50.5		23.8		70.0	

Note: “+” indicates high success, “0” indicates moderate success, and “–” indicates low success

Staff members as a group were well-educated. All had completed high school and 70% were college graduates. The large majority of staff members at all Sites, except Site D, had earned a bachelor's degree. This was true for only 43% of Site D's staff.

Table 2 shows that staffing was particularly problematic at Sites A and D, where staffing was unstable as indicated by fewer days worked, and staff quality was low in terms of either the level of training or education. Sites B and E had the most stable and the most highly qualified staffs. Site C's staff were educated but staffing was moderately unstable and training hours were also in the moderate range.

Self-Reported Experiences

Average positive experiences reported on the YES by school are presented in Table 3. Students who attended Site B reported the highest average score, followed by Sites C and E with identical scores. Sites A and D had identical scores at the bottom of the ranking.

Overall, students scored 2.9 on a four-point scale, indicating that youth in our sample were exposed to generally positive experiences after school. Our program compares favorably with leisure activities assessed by youth who completed the YES in other studies. For example, the sample on whom the developers tested the instrument scored an average of 2.7 in positive experiences (Hansen and Larson 2005). Based on this more objective standard, sites in our sample were ranked "high" when the

average YES score exceeded 2.7 and moderate if they tied this score. No site averaged less than 2.7.

Comparison of Implementation Quality and Youth Experiences

A summary matrix of implementation and youth experiences measures is presented in Table 4. Although this comparison is challenged by the small number of sites available, Table 4 suggests an association between implementation and student experiences after school for all implementation variables with the exception of program dose. The results point to one site, Site B, in which implementation was consistently positive across dimensions and where students reported the most positive experiences. Site E also shows consistently high-quality implementation and positive youth outcomes, but student attendance was judged to be only moderate. Results also reveal two problematic sites, A and D, which consistently performed poorly relative to the other sites on all aspects of implementation studied with the exception of staff education for Site A and staff training for Site D. Youth at these sites also reported less positive experiences. The association between implementation and youth experiences at site C was not as straightforward. Site C was unable to gain consistent attendance from students or staff, yet students reported positive experiences after school on the YES.

Discussion

This study provided a preliminary look at how implementation in "real-world" settings may differ from the ideal and highlights the importance of measuring different aspects of implementation. It used data collected during an experimental evaluation of an ASP designed to incorporate evidence-based practices into the normal routine of after-school programming. The results revealed that program attendance, student engagement, program management and climate, staff days worked, staff training, and staff education varied across sites. Most of the dimensions of implementation we studied tended to co-vary with student self-reported experiences. Management and climate and

Table 3 Mean YES score, by site and overall

Site	YES positive experiences		
	Score	N	Rank
A	2.7	26	0
B	3.1	40	+
C	2.9	30	+
D	2.7	43	0
E	2.9	35	+
Overall	2.9	174	

Note: "+" indicates high success, "0" indicates moderate success

Table 4 Comparison of implementation levels and youth's self-reported experiences (YES)

Site	Days attended	Engagement rate	Management and climate	Days worked	Hours of training	% BA or higher	YES
A	–	0	0	–	–	+	0
B	+	+	+	+	+	+	+
C	–	+	+	0	0	+	+
D	0	0	–	–	+	–	0
E	0	+	+	+	+	+	+

Note: "+" indicates high success, "0" indicates moderate success, and "–" indicates low success

student engagement showed a consistent relationship to student experiences across sites. The combination of high staff training and education appeared to relate to high quality experiences, although one site (C) achieved positive YES scores despite only moderate staff training. Staff stability (as measured by average days worked) was also related to reported YES experiences, but again, Site C was judged as having only moderate staff stability. While the site with the highest dosage (B) had the highest YES score, the site with the lowest dosage (C) also had high YES scores.

This study suggests that levels of achieved implementation are related to youths' program experiences. The operative aspects of implementation in this study were quality of program management and climate (as measured by structure, supervision, social climate, behavior management, and leader skill), educated staff members who received sufficient training and remained in their positions over an extended period of time, and engaging program content. Programs with low levels of implementation on these dimensions were less successful in creating positive experiences for youth. These findings are in agreement with other intervention research which has stressed that the price of implementation failure is loss of program effects (Durlak and Dupre 2008; Dusenbury et al. 2003).

Interestingly, youth experiences were not highly related to our measure of dosage. This finding accords with prior warnings that high attendance does not always yield the best outcomes (Simpkins-Chaput et al. 2004). Our observations indicated that, although the ASPs were intended to be voluntary, many youths were being required to attend by their parents, who valued the free child care. This scenario was especially evident at site D, in which youths reported less positive outcomes despite having the second highest attendance rate. This finding implies that researchers should not assume that dosage is a measure of program quality. High attendance at a poorly implemented program may do more harm than good.

Qualitative Impressions

Qualitative impressions of each of the sites confirmed much of the quantitative data regarding program quality and suggested some additional features worthy of attention. The finding that Sites B and E received the best evaluations in terms of implementation and youth experiences did not surprise observers who attended the sites. Site B, in particular, had an undeniably positive atmosphere. The two consistent staffers at Site B worked effectively and cheerfully as a team. The site director and program assistant at this site provided a schedule for each day and announced it clearly at the start of the program, giving students a choice of several activities. They connected easily with and were trusted by the youth. Additionally, the charisma of Site B's

site director, a young, popular teacher at the school, certainly contributed to this site's success. He may have been the principal attraction of the program for many students.

Site E's success can also be linked to particularly effective staff. The site director and one program assistant at Site E also worked at the site for the duration of the program. As the year progressed, Site E experienced a relatively high rate of student drop-out, but the students who remained appeared to have bonded with each other and with the staff. Again, in parallel with Site B, staff members were generally cheerful and related to the youth warmly. This observation is in accord with previous research which found that positive emotional climate in an ASP was related to better outcomes (Pierce et al. 1999).

The observation staff was equally unsurprised that this evaluation reflected poorly on Sites A and D. Observers at Site A remarked on a very small program (it was the smallest, averaging 14 youth per day) consisting of disengaged, nearly disgruntled youth who did not appear to enjoy each other's company. Site A operated on a highly disorganized daily schedule; frequent staff turnover was doubtlessly a major contributor to this problem. In fact, maintaining a sufficient number of employees at Site A became so challenging that several replacement staff assigned to this program had competing commitments which interfered with their ability to arrive at the site on time or everyday that it was open.

Site D, which averaged 23 youths served each day, was rife with behavior problems to the extent that observers expressed concern to the vendor about the safety of youth. Students acted out with very little redirection from staff members. When discipline was exercised it appeared capricious and confusing to youth, consisting of extended periods of quiet time for all students including those who were not involved in misbehavior. While students at Site A appeared disgruntled, it was the staff at Site D who appeared irritated and apathetic. They sometimes seemed more content to socialize among themselves than to interact with youth or attempt to manage behavior.

One common element distinguished Sites A and D from the rest. These sites shared the experience of early turnover in the site director position. The original director at Site A worked in her position for only 11 days, while Site D's director left after 26 days. Vacancies in these positions were filled by lower-ranking staff members and necessitated reverberating staff reorganization. Instability in leadership and support staff at the two sites where program integrity was the most compromised likely affected these sites' ability to provide a positive environment for youth.

The evaluation indicates that Site C was a moderately successful program. Students reported positive experiences after school on the YES but implementation indicators were only moderately positive. This result was predictable

based on qualitative impressions as well. There was nothing exceptional to note about the atmosphere at Site C. It was run over the course of the full year by a seasoned educator who fluidly established standards for behavior and a reliable structure for the program. However, he was frequently absent from program activities, acting as more of a manager of other employees than as a direct-services provider. Other staff members were competent and kind in their interactions with students, but were not impressively warm or engaging. Over the course of the year, Site C became populated by a small but dependable group of students who appeared to get along very well with each other and easily followed program rules. An administrative employee described Site C as “its own little glee club.” Positive YES results at Site C may have been created by friendships among students which were in part facilitated by stable site management.

Conclusions

The combination of quantitative and qualitative data suggest that staff quality might be the single most important characteristic of program success because the quality of program staff seemed to affect other aspects of implementation. Staff members who were highly educated, well trained, and employed long-term appeared to observers to be more skilled in providing youth services. They appeared better able to establish sound management, create a positive social climate, and provide engaging content. Although the causal connection among factors cannot be ascertained in a descriptive study such as ours, this finding regarding the importance of staff resonates with findings for several other studies of ASPs (Gottfredson et al. 2007; Pierce et al. 1999; Rosenthal and Vandell 1996). Of course, high staff turnover is common in child care and ASPs settings (Granger 2008; Whitebook et al. 1998). Low wages, lack of fringe benefits, and part-time hours combine to make ASP employment undesirable for persons who are qualified for better jobs (recall that 70% of the staff members at the program discussed here were college graduates). Staff turnover remains a major challenge to high quality implementation in ASPs in general (Granger 2008). It can also be expected to remain a challenge in effectiveness trials such as this that attempt to deliver a program under the same conditions as one would expect in the real world.

Although we did not address program content in this paper at length, it is clear that academic assistance was the least engaging activity offered at the sites. This has major implications for practice because academic assistance, such as that delivered in the programs studied here, is a staple in many ASPs (Dynarski et al. 2003). ASPs which hope to impact academic outcomes should implement academic

enrichment activities that are engaging to youth to increase the likelihood of success in this area.

Finally, prior research suggests that the use of structured, evidence-based curricula are important to the success of ASPs (Durlak and Weissberg 2007; Gottfredson et al. 2004, 2007). Yet, the studies on which these findings are based have not generally assessed multiple dimensions of program quality. It is possible, and a question for future research, that the use of structured, evidence-based content is confounded with some or all of the dimensions of implementation discussed in this paper. Research that assesses multiple dimensions of implementation quality in a large sample of ASPs is required to begin to sort out the characteristics of effective ASPs.

Limitations

Our examination of the co-variation of implementation quality and youth experiences is imprecise due to the small number of sites available for cross-site analysis. Conclusions are also limited by lack of variability in dosage across programs. Only at Site B did students attend close to half of program days on average, limiting examination of implementation fidelity in relation to dosage. The conclusions of this paper are also limited by the fact that not all youth YES respondents were regular program attendees and we do not know how the youths at each site were spending their remaining out-of-school time in addition to the ASPs evaluated here.

Despite these limitations, we believe this study provides guidance for researchers who may wish to measure implementation in ASPs in the future, and underscores the importance of doing so. In particular, our study demonstrates that program implementation is multi-dimensional, and that, although many of the dimensions we measured co-vary, some do not. It therefore appears necessary to measure multiple aspects of implementation and to begin to build a stronger evidence base to support conclusions about the features of successful ASPs. As researchers undertake detailed, high-quality studies of intervention programming, the knowledge base on effective interventions will expand. Hopefully, future achievements will include the formation of reliable guidelines not only on which aspects of implementation are the most important to program success but also on how ASPs can achieve high standards of implementation in real world settings in which researchers wield little control.

Acknowledgments This work was supported through grant number R305F050069 from the U.S. Department of Education Institute of Educational Sciences to the University of Maryland. We acknowledge the support of the Baltimore County Local Management Board, the Baltimore County Public Schools, and the Baltimore County Department of Recreation and Parks for implementing and managing

the after school programs. We especially acknowledge the assistance of Elise Andrews of the Baltimore County Local Management Board and Beahtha Davis of the Baltimore County Department of Recreation and Parks. We would like to thank Gordon Bonham for collecting data from the school system and Stephanie DiPietro, Mathew Gugino, Lynda Okeke, Matthew Brigham, Freshta Rahimi, and Jaynie Trageser for research assistance. Finally, we thank three anonymous reviewers for helpful comments on an earlier draft, and Joe Durlak for major editorial assistance in improving the paper.

References

- Armstrong, T., & Armstrong, G. (2004). The organizational, community and programmatic characteristics that predict the effective implementation of after-school programs. *Journal of School Violence*, 3, 93–109.
- Beckett, M., Hawken, A., & Jackowitz, A. (2001). *Accountability for after-school care: Devising standards and measuring adherence to them*. Santa Monica, CA: Rand Corporation.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23–45.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
- Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. Retrieved February 12, 2007, from www.casel.org.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237–256.
- 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: New findings: US Department of Education, National Center for Education Evaluation and Regional Assistance*. Washington, DC: US Government Printing Office.
- Fashola, O. (1998). *Review of extended day and after school programs and their effectiveness*. Retrieved January, 2008, from www.csos.jhu.edu/crespar/techReports/Report24.pdf.
- Finance Project. (2007). *Estimated federal investment in out-of-school time*. Retrieved April, 2008, from www.financeproject.org/publications/estimatedfederalOST_TFPflyer.pdf.
- Gerstenblith, S. A., Soulé, D. A., Gottfredson, D. C., Lu, S., Kellstrom, M. A., Womer, S. C., et al. (2005). ASPs, antisocial behavior, and positive youth development: An exploration of the relationship between program implementation and changes in youth development. In J. L. Mahoney, R. W. Larson, & J. S. Eccles (Eds.), *Organized activities as contexts of development: Extracurricular activities, after-school and community programs* (pp. 457–478). Mahwah, NH: Lawrence Erlbaum & Associates.
- Gottfredson, D. C., Cross, A. B., & Soulé, D. A. (2007). Distinguishing characteristics of effective and ineffective after-school programs to prevention delinquency and victimization. *Criminology and Public Policy*, 6, 601–631.
- Gottfredson, D. C., Cross, A. B., Wilson, D. M., Rorie, M., & Connell, N. (in press). Effects of participation in after-school programs for middle school students: A randomized trial. *Journal of Research on Educational Effectiveness*.
- 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. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*, 39, 3–35.
- 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. (2008). After-school programs and academics: Implications for policy, practice, and research. *Social Policy Report*, 22, 3–19.
- Granger, R. C., Durlak, J. A., Yohalem, N., & Reisner, E. (2007). *Improving after-school program quality*. The William T. Grant foundation working paper. Retrieved May, 2008, from www.wtgrantfoundation.org/usr_doc/Improving_After-School_Program_Quality.pdf.
- Grossman, J., Campbell, M., & Raley, B. (2007). *Quality time after School: What instructors can do to enhance learning*. Philadelphia: Public/Private Ventures.
- Hansen, W. B., & Dusenbury, L. (2004). All Stars Plus: A competence and motivation enhancement approach to prevention. *Health Education*, 104, 371–381.
- Hansen, W. B., & Larson, R. (2005). *The youth experiences survey 2.0: Instrument revisions and validity testing*. Urbana-Champaign, IL: University of Illinois.
- 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.
- Kane, T. J. (2004). *The impact of after school programs: Interpreting the results of four recent evaluations*. The William T. Grant foundation working paper. Retrieved May, 2008, from http://www.wtgrantfoundation.org/usr_doc/After-school_paper.pdf.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Aphorp, 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.
- Mahoney, J. L. (2000). School extracurricular activity participation as a moderator in the development of antisocial patterns. *Child Development*, 71, 502–516.
- 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.
- Pierce, K. M., Hamm, J. V., & Vandell, D. L. (1999). Experiences in after-school programs and children's adjustment in first grade classrooms. *Child Development*, 70, 756–767.
- Rosenthal, R., & Vandell, D. L. (1996). Quality of care at school-aged child-care programs: Regulatable features, observed experiences, child perspectives, and parent perspectives. *Child Development*, 67, 2434–2445.
- 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.
- Simpkins-Chaput, S., Little, P. M. D., & Weiss, H. B. (2004). *Understanding and measuring attendance in out-of-school time programs. Issues and opportunities in out-of-school time evaluation brief no. 7*. Cambridge, MA: Harvard Family Research Project.
- 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 report.
- Weiss, H. B., Little, P. M., & Bouffard, S. M. (2005). More than just being there: Balancing the participation equation. *New Directions for Youth Development*, 2005, 15–31.
- Whitebook, M., Howes, C., & Phillips, D. (1998). *Worthy work, unlivable wages: The national child care staffing study, 1988–1997*. Washington, DC: National Center for the Early Childhood Work Force.