Prev Sci (2010) 11:142–154
DOI 10.1007/s11121-009-0156-7

An Experimental Evaluation of the All Stars Prevention Curriculum in a Community After School Setting

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

Published online: 27 October 2009

Abstract This study tested the effectiveness of a prevention curriculum, All Stars, as implemented in a year-long school-based after school program and provides an independent replication of the effects of All Stars on targeted mediators and problem behaviors using an experimental methodology. Middle school students (N=447) who registered for the after school program were randomly assigned to the experimental or control condition. The sample included approximately equal proportions of males and females, was 70% African American, and 59% of the students received subsidized meals at school. All Stars was delivered with reasonable integrity to the program design, although with lower quality than reported in earlier efficacy trials. However, actual student exposure to the program was lower than expected due to low levels of attendance in the after school program. Students who ever attended received an average of 16 h of All Stars instruction. Results showed no differences between the treatment and control students at post-test on any of the outcomes or mediators. Further, no positive effects were found for youths receiving higher dosage, higher quality program delivery, or both. Insufficient time to achieve high quality implementation in the after school context and potential deviancy training are suggested as reasons for the failure to replicate positive program effects.

Keywords All Stars · Effectiveness trial · Randomized experiment · After school program

By eighth grade, approximately 20% of U.S. students have smoked cigarettes, been drunk, or tried illegal drugs. About half of high school seniors have engaged in these risky, unhealthful behaviors (www.monitoringthefuture.org). Because the prevalence and intensity of substance use increases so rapidly after the beginning of high school, middle school is a logical time in which to deliver drug prevention programming. Over the past half-century Americans have placed increasing responsibility on schools to teach skills and provide information youth need to avoid drugs and other problem behaviors (Catalano et al. 2004). In fact, as of 1999, 76% of schools in the U.S. reported using at least one prevention curriculum (Gottfredson and Gottfredson 2001), 85% of which focused on drug prevention (Gottfredson et al. 2000).

However, the feasibility of offering substance use prevention curricula during the school day is on the decline. With the Bush administration's No Child Left Behind legislation, schools and school districts are under increased pressure to focus their efforts on increasing academic achievement (Hamilton et al. 2007). Although it is reasonable to expect that programming focused on reducing substance use will improve learning outcomes as well,

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 thank Elise Andrews of the Baltimore County Local Management Board and Beahta Davis and the staff of the Baltimore County Dept. of Recreation and Parks for implementing and managing the after school programs. Thanks also to Matthew Brigham, Stephanie DiPietro, Matt Gugino, Lynda Okeke, Freshta Rahimi, and Jaynie Trageser for research assistance and to William Hansen and three anonymous reviewers for helpful comments on an earlier draft.

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

N. Connell
Law and Justice Studies Department, Rowan University,
Glassboro, NJ, USA

many schools have decided that they can no longer devote valuable instructional time to such programming. Concurrent with the decline in the allocation of instructional time for prevention programming has been an upsurge in interest in after school programs (ASPs). Considerable federal, state, local, and private money is being invested in these programs. For example, the 21st Century Community Learning Center Program received approximately one billion dollars in federal funds annually from 2002–2007 to provide before- and after-school enrichment for students in low-performing schools. Estimates of total annual federal investment in out-of-school time reach as high as $3.6 billion (financeproject.org, 2007).

ASPs offer an alternative venue in which to provide high quality prevention programming. Reviews of evaluations of these programs, however, report inconsistent findings with respect to outcomes such as academic performance and social development. Several reviews suggest that ASPs can produce small but measurable improvements in academic performance (Durlak and Weissberg 2007; Lauer et al. 2006; Scott-Little et al. 2002), but Kane (2004) and Zief and Lauver (2006) find no such effects. Durlak and Weissberg (2007) suggest that ASPs also have beneficial effects on behavioral and attitudinal outcomes, but Zief and Lauver (2006) find no such effects. Disparate conclusions seem to depend upon the characteristics of the programs and on the methodological rigor of the studies included in the reviews. Very few studies of ASP effectiveness meet contemporary standards for scientific rigor in program evaluation (Flay et al. 2005), and conclusions from reviews differ depending upon where they have set the cut-off for scientific rigor in deciding which studies to include.

The limited research that has explored correlates of ASP effectiveness suggests that when structured and research-based content is incorporated into the programs, they are more effective for improving social competency skills and reducing a variety of problem behaviors, including substance use (Durlak and Weissberg 2007; Gottfredson et al. 2004, 2007). Also, improvements in outcomes seem to be conditional on program attendance (Kane 2004). Of course, achieving sufficiently high dosage of programming is likely to be more challenging in ASP settings than in school settings because of the voluntary nature of ASPs. In the evaluation of the Communities that Care prevention system, Fagan et al. (2008) reported that when a variety of evidence-based program models (including All Stars) were implemented in community settings, exposure to the evidence-based programming was lower in after school than in school settings.

The current study tested the effectiveness of a prevention curriculum, All Stars (www.allstarsprevention.com), when it was incorporated into a school-based ASP. This research provides an independent replication of the effects of All Stars on targeted mediators and problem behaviors in the context of a randomized trial of a year-long ASP.

All Stars

The All Stars "Core" includes 14 lessons intended to prevent substance use and to reduce bullying, violence, and other conduct problems. The developers identified five processes through which the program should affect these outcomes. First, the program builds awareness of *lifestyle incongruence* by helping youth identify their ideal futures and recognize that risky behaviors are inconsistent with these desired lifestyles. Second, youth clarify their beliefs about the prevalence and acceptability of risky behaviors by their peers *(normative beliefs)*. Third, the program challenges youth to make voluntary, public *commitments to abstain from drugs* as an indicator of maturity. Fourth, it promotes *bonding* with positive social organizations such as schools or community groups that support a positive norm. Fifth, the program encourages *positive parental attention* through parent–child communication, and parental monitoring of the student's activities. Teaching techniques are highly interactive and include games, small group activities, surveys, open discussions, homework assignments in which parents are encouraged to participate, hands-on group leader involvement, videotaped public commitments, and parental attendance at graduation ceremonies.

All Stars "Plus" includes 13 lessons designed to reinforce the changes in behaviors and attitudes realized through the Core program and teaches youth additional developmental skills. First, youth actively engage in lessons that teach *goal setting* skills and include realistic ways to achieve goals and encourage *persistence*. Second, youth learn a five-step process for *decision making* that discourages *impulsivity* and they practice it with hypothetical yet realistic situations. Last, youth learn *resistance skills* through games and role playing that include both verbal and nonverbal communication for resisting peer pressure. Figure 1 shows the All Stars model tested in this study.

In both the Core and Plus segments, key skills and techniques are often reinforced through review after they have been introduced. A special celebration at the conclusion of each program is designed to reinforce, reward, and remind students about the ideals they share, standards they agree on, and commitments they have made.

Evaluations of All Stars have generally found that the program produces positive effects on the targeted

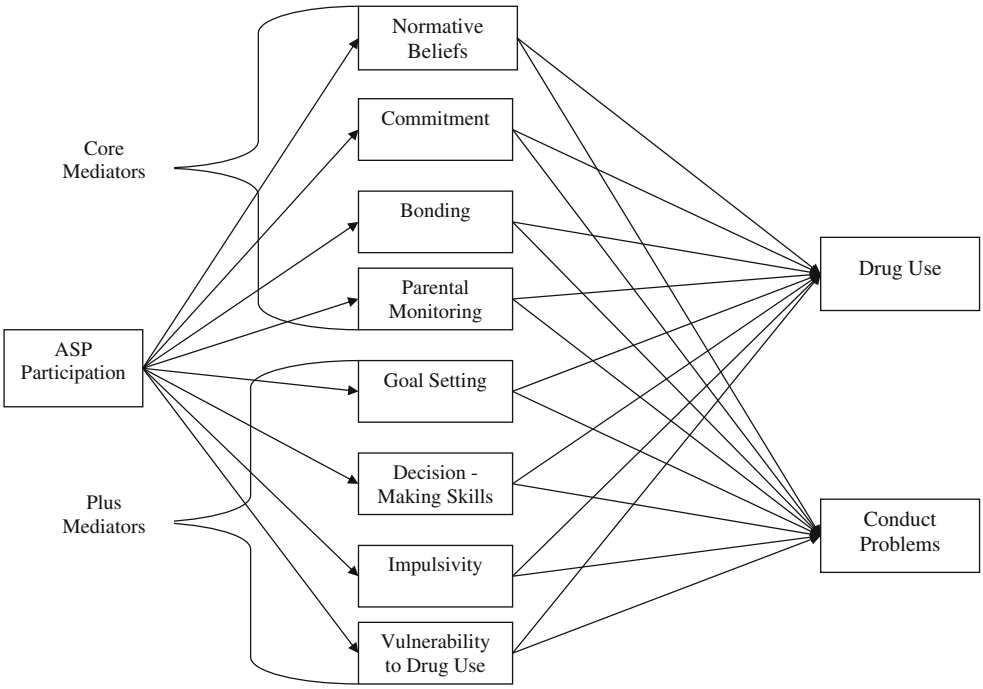

Fig. 1 Theoretical model underlying the All Stars program

mediators, but results have been mixed for substance use and violence outcomes.[1] A series of studies conducted by or in collaboration with the developer of the program have found moderately positive effects of the program. Hansen (1996) found All Stars was more effective than DARE in improving the Core mediators among seventh-graders.

Using a sample of 14 middle schools, 11 of which were randomly assigned to receive All Stars or a control condition, with 3 additional All Stars schools added post-randomization, Harrington et al. (2001) found positive effects for All Stars on bonding and commitment, but no effects on normative beliefs, substance use, or violence. However, another study of this sample indicated that All Stars increased perceptions of lifestyle incongruence and lowered alcohol, cigarette, and inhalant use (McNeal et al. 2004). This study demonstrated that reductions in substance use were accounted for by increases in three of the four proposed mediators (all except bonding). Additionally McNeal et al. (2004) found that All Stars was more effective when delivered by teachers as opposed to non-teachers.

Hansen and Dusenbury (2004) examined mediators targeted by the Core and Plus as well as substance use outcomes. They reported that both programs improved perceptions of lifestyle incongruence and parent–child communication, reduced alcohol use and drunkenness, and decelerated maturational decay in bonding and goal setting relative to the control condition. Plus was associated with reduced cigarette smoking.

Somewhat at odds with prior studies conducted by All Stars developers, Slater et al. (2006) found no effect of All Stars on substance use. This study tested the effect of All Stars in conjunction with a media campaign on alcohol, cigarette, and marijuana use. Independent of the media intervention, All Stars did not affect substance use. However, schools receiving both the media campaign and All Stars exhibited the best outcomes.

Overall, these evaluation studies suggest that All Stars positively influences the mediators it targets. Based on results of these preliminary evaluations, All Stars has been recognized as a promising program in the Office of Juvenile Justice and Delinquency Prevention's Model Programs Guide and the U.S. Department of Education's Expert Panel on Safe, Discipline, and Drug-Free Schools. Still, the available studies generally fall short of contemporary standards for scientific rigor required to establish intervention effectiveness (Flay et al. 2005) because they fail to fully randomize subjects to conditions and sometimes suffer from attrition problems.

Reviews of school-based drug prevention programs and programs aimed at reducing conduct problems and delinquent behavior more generally, however, have suggested that programs like All Stars are effective. For example, a meta-analysis of 178 school-based prevention programs (Gottfredson et al. 2002b; Wilson et al. 2001) found that

[1] Some research has examined All Stars effects on mediators not specifically targeted by All Stars and on sexual activity outcomes. We do not discuss these findings as they are beyond the scope of this research.

instructional programs that (a) emphasize self-control or social competency promotion using cognitive–behavioral instructional methods (e.g., methods that emphasize modeling, rehearsing, and coaching of new skills) and (b) establish norms or expectations for behavior are among the most effective strategies for reducing problem behaviors. These two categories of preventive activities have shown consistently positive results across a variety of different types of problem behavior outcomes, including delinquent behavior and substance use. The All Stars curriculum exemplifies both of these effective categories of prevention. The Core program focuses on increasing the salience of social norms regarding violence and substance use for participating youths. The Plus program teaches specific social competency skills needed to successfully avoid risky situations and make healthy choices. Both program components are taught using highly interactive methods such as those found to be related to program effectiveness in the meta-analysis just described. Thus, research on effective prevention practices in general coupled with the preliminary evaluations of All Stars suggest that this program should be effective for reducing substance use and related conduct problems.

Methods

Sample

This study of the effectiveness of the All Stars program was part of a larger study to test the effectiveness of an enhanced ASP model for improving learning and reducing problem behaviors. The larger study is described in Gottfredson et al. (2009). Five urban middle schools volunteered to participate in this study. All students who attended these five schools were invited to register for the ASP. Gottfredson et al. (2009) reported that students who registered for the ASP (who constituted between 8% and 21% of their school populations) were generally representative of the populations of their schools in terms of gender and SES. However, the ASP attracted a disproportionately higher percentage of minority youths.

Within each participating school, registered students had a 50% chance of being randomly assigned to the treatment group (i.e., invited to attend the after-school program) or to the control group. About half of the sample were males (54%), 70% were African Americans, 17% were Caucasian, 2% were Latino, 8% were multi-racial, and the remaining 2% were of another race. The average age of participants was 12.2, and 59% received subsidized meals at school. Students from all grade levels (six–eight) participated, but sixth graders were better represented than eighth graders (42% vs. 25%).

Treatment (N=224) and control (N=223) students did not differ in terms of either demographics (age, grade, family income, gender, race, single-parent household, receipt of subsidized meals, maternal education), pre-treatment academic indicators (school absences, suspensions, grades, two standardized test scores), and differed significantly on only 2 of 13 pre-treatment measures from youth surveys. The treatment youth scored higher in decision-making skills at pre-test than controls. One significant difference in 26 tests conducted is less than the number that would be expected by chance using a critical value of $p<0.05$. Site by experimental group interactions were also examined for the same 26 pre-treatment measures. No interactions by site were statistically significant at the $p<0.05$ level.

Students included in the outcome analysis were the 416[2] (93%) registered students who adequately completed a post-test survey at the end of the school year. Attrition analysis showed that registered youth who were excluded from the study[3] (N=31; 13 treatment, 18 control) were similar to those who were included, demographically and on a range of pre-treatment measures. Exceptions were age and attitudes favorable to drug use. The excluded cases scored in the more at-risk direction on these measures. Treatment by attrition interactions (reported in Gottfredson et al. 2009) were examined to test for differential attrition by treatment status. Of 28 interactions, 1 (MSA math score) was statistically significant at the $p<0.05$ level, suggesting that higher achievers may have been more likely to attrit from the treatment than the control group. No differential attrition was observed for any of the measures used in this report.

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

[2] We conducted power analyses for two-tailed independent t-tests, fixing the Type I error rate (alpha) at 0.05. These power analyses estimated the power available to detect a moderate-sized effect (0.30) and indicated that, using the entire sample, the power to detect differences between the treatment and control groups is 0.99. We calculated power for a number of different outcomes, including aggression (measured continuously) and vulnerability to drug use and last-month frequency drug use (measured as binary outcomes). These power analyses also estimated the minimum effect size detectable with a sample of our size and power fixed at 0.80. Analyses including both a dummy variable for school and the pre-test measure as a covariate will allow us to detect effects of about 0.17.

[3] These students were excluded from outcome analysis because of missing posttest data. They 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).

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. 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 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. Consequently, this study tests the effectiveness of the All Stars program in the context of a structured ASP relative to whatever alternative after school activities were available to youths in the study. These alternatives included a wide variety of leisure activities normally found in schools and communities (e.g., sports, performance/fine arts activities), but did not include All Stars.

Measures

Data on outcomes and mediators were taken primarily from a youth self-report survey. Teacher reports of youth social adjustment were also used. The student surveys contained 167 items and were administered to all youth who registered for the ASP following registration and again near the end of the school year. The response rate for post-test youth surveys was 93%.

Teacher ratings were sought from four of each youth's teachers near the end of the program. At least one survey was returned for 99% of students and two or more were returned for 88%. Responses were averaged to create one teacher rating score per student.

Outcomes and Mediators

Most All Stars outcomes and mediators were measured using multi-item scales. Higher scores on all scales indicate a higher level of the construct. Unless otherwise indicated, scales were averages based on all valid items. This study focuses on the outcomes of substance use and conduct problems as well as the mediators described by the All Stars developer, with the exception of *lifestyle incongruence*, which was not measured as a separate mediator in this study. We used 13 scales to assess last month drug use, disruptive classroom behavior, aggression, delinquency, and nine mediators targeted by All Stars Core and Plus curricula. We also measured tobacco, alcohol, marijuana, inhalant and other drug initiation using single-item measures that captured a student's first use of a substance during the year-long study period. Descriptive information about the number of scale items, range, and a sample item from each scale are presented in Table 1. Substance initiation items are not tabled.[4] See Table 2 for descriptive statistics and alpha reliabilities for measures used in this study

As is evident from Table 2, missing data were minimal for most measures used in the analyses. Missing data exceeded 5% of the available cases only for the decision-making measures. For all outcomes except decision-making, analyses simply excluded the small number of cases for which outcome data were missing. For decision-making (pre and post), we employed maximum likelihood methods for imputing missing data (following Allison 2002). We used the EM algorithm available in the SPSS Version 16.0 "missing value analysis" module to predict missing values based on several scales available in the pre- and post-test survey that were moderately correlated (0.3 or higher) with the decision-making scales. The item statistics for the imputed decision-making scales are nearly identical to those based on the raw scale scores.

Demographics

Demographic information on age, race, gender, parent income, family structure, and subsidized meals was collected through parent reports and youth self-reports.

Implementation

Graduate student observers and program staff contributed implementation fidelity data. Observers, who attended a 3-day training on All Stars implementation provided by Tanglewood Research (along with program staff), conducted 80 site visits; 14–18 visits per site. Observers attended 66 All Stars sessions during these visits. They and staff members completed All Stars fidelity checklists created by the developer (Tanglewood Research 2008a). A similar checklist was used in previous studies of All Stars implementation quality (Hansen 1996; Harrington et al. 2001). The checklists, tailored to individual lessons, asked for a rating of the overall quality of the lesson as taught, the level of student engagement in the lesson, and the extent to which each activity in each lesson achieved its stated objectives. Group leaders completed fidelity checklists covering every All Stars session. They also recorded session attendance and the activities covered in each lesson. The attendance data were used to calculate the total number of sessions attended and lessons received for each of the 224 treatment students. We report observer assessments of quality and engagement and staff reports of objective achievement. Observers were not able to accurately assess objective

[4] Participants who reported some use at pre-test or failed to report at one or both time periods for cigarettes (37, 9%), alcohol (75, 18%), marijuana (30, 7%), inhalants (40, 10%) and other illegal drugs (11, 3%) were excluded from analyses of initiation for each substance.

Table 1 Scale content

	N Items	Scale range	Sample item
Outcomes			
Last month drug use[a]	3	0–1	In the last month how often have you smoked marijuana?
Disruptive classroom behavior[b]	3	1–3	How often been sent out of a classroom by a teacher for bad behavior?
Aggression[b]	6	1–4	How often did you did you tease someone else your age?
Delinquent behavior[a]	7	0–7[f]	Have you stolen or tried to steal things worth less than $50?
Mediators			
Normative beliefs[a,b]	17	0–1	How wrong is it for someone your age to use marijuana?
Commitment[a]	3	0–1	I will never smoke cigarettes.
Bonding[b]	3	0–3[f]	I feel like I belong at this school.
Parental monitoring[a,c]	6	1–4	My parents know who I am with if I am not at home.
Goal setting[b]	6	1–4	Once I set a goal, I don't give up until I achieve it.
Decision-making skills[b]	4	1–4	How often do you stop to think about your options before you make a decision?
Impulsivity[d]	10	0–1	Do you often get in trouble because you do things without thinking?
Vulnerability to drug use[b]	3	0–1	Pretend your best friend offered you a drink of beer or wine and you did not want it. How hard would it be to say "no"?
Social competency[e] (TR)	29	1–4	The student removes him or herself from potential problem situations.

[a] Gottfredson and Gottfredson (1992)

[b] Tanglewood Research (2008b)

[c] Huizinga and Esbensen (1990)

[d] Eysenck et al. (1984)

[e] Gottfredson et al. (2002a)

[f] Count of number of endorsed items

Table 2 Mediator and outcome descriptives and reliabilities

	Pre-test				Post-test			
	Mean	SD	*N*	α	Mean	SD	*N*	α
Outcomes								
Last month drug use	0.09	0.29	412	0.69	0.18	0.39	412	0.74
Cigarette initiation	–	–	–	–	0.08	0.27	379	–
Alcohol initiation	–	–	–	–	0.16	0.37	341	–
Marijuana initiation	–	–	–	–	0.06	0.23	386	–
Inhalant initiation	–	–	–	–	0.09	0.28	376	–
Other drug initiation	–	–	–	–	0.03	0.16	405	–
Disruptive classroom behavior	1.38	0.49	413	0.64	1.63	0.58	409	0.64
Aggression	1.58	0.59	412	0.83	1.87	0.74	414	0.86
Delinquent behavior	0.43	0.99	413	0.72	0.86	1.55	412	0.81
Mediators								
Normative beliefs	0.73	0.25	414	0.88	0.64	0.29	415	0.90
Commitment	0.78	0.35	414	0.81	0.74	0.36	415	0.80
Bonding	2.49	0.86	414	0.68	1.99	1.13	414	0.73
Parental monitoring	3.37	0.50	415	0.61	3.26	0.56	414	0.70
Goal setting	3.11	0.51	414	0.70	3.02	0.58	415	0.76
Decision-making skills	2.83	0.76	398	0.83	2.72	0.82	381	0.86
Impulsivity	0.51	0.27	409	0.75	0.56	0.28	412	0.77
Vulnerability to drug use	0.34	0.47	407	0.91	0.34	0.47	402	0.90
Social competency (TR)	–	–	–	–	2.71	0.52	405	0.96

Scales are taken from the youth survey unless otherwise noted

TR teacher ratings

achievement as they often witnessed only portions of a given lesson.[5]

Observer Measures As noted above, observers used fidelity checklists to measure quality and engagement in each of the 66 All Stars lessons observed. Observers also measured youth engagement in All Stars and other program activities using a more general engagement observation form. Two-hundred two observations of youth engagement across different activities were used in this analysis. The engagement observation measured the degree of structure of each activity and systematically counted the number of students engaged during discrete program activities. At every 5-min interval the observer rated the level of structure from 1 to 5 (where low scores indicated little or no expectations for student conduct and high scores indicated clearly defined expectations of all students at all times) and tallied the number of students who were engaged or not engaged in the activity at that time. 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.

Analysis Strategy

After presenting descriptive information about the level of implementation of All Stars, we report on tests for program effects from a series of regression models that controlled for the pre-test level of each variable, age, gender, race (black v. non), and school site.[6] Two-tailed tests of significance using an alpha level of 0.05 were used. Continuous variables (disruptive classroom behavior, aggression, and all mediators other than bonding) were analyzed using OLS regression; delinquent behavior was analyzed as a count using negative binomial regression, and bonding was analyzed as a count using Poisson regression. Substance use outcomes, all of which were measured with binary dependent variables, were analyzed using logistic regression. Results from these regression analyses were used to produce covariate-adjusted post-test means, by experimental group, for the measures of the All Stars mediators and outcomes. We then tested for interactions of treatment by program quality. As will be described in more detail below, three of the five sites delivered All Stars with higher quality than the others. We tested for interactions of treatment status and quality to explore the possibility that more positive outcomes were observed among students who attended higher quality programs. Finally, we present the results of dosage analyses in which the mediators and outcomes were correlated with the number of lessons received. These simple dosage analyses were augmented with regression analysis controlling on pre-treatment variables related to lessons received to help to rule out selection artifacts as an alternative explanation for the pattern of correlations observed.

Results

Implementation

As planned, the ASP that included the All Stars intervention operated for 3 h per day, 3 days per week, for a total of 96 days during the 2006–2007 school year at each of the five sites. All five sites followed the same daily schedule: The program day began with a brief (approximately 30-min) period during which students enjoyed a snack while staff handled administrative tasks. On Tuesdays and Wednesdays, this snack period was followed by an hour and a half block during which students attended academic assistance sessions (usually homework) with staff assistance and All Stars sessions each for 45 min. Students were typically divided into two groups during this period and alternated between the academic and All Stars activities. The last hour of the day was spent on leisure activities, usually sports, crafts, board games, movies, or computer games. On Thursdays, the entire 3 h of programming was dedicated to snack time and leisure activities.

The average daily attendance at each site was a function of the total number of students enrolled, the drop-out rate, and the average daily attendance. Although the study design called for enrolling 50 treatment students per site, the number of registered youths varied somewhat across sites, so that the number of treatment students actually

[5] Each All Stars lesson is designed to be delivered start-to-finish during one session. However, the ASP sites delivered one lesson over 2 days in two 45-min sessions because the program structure provided extra time.

[6] Subjects were clustered within site. Clustering often violates the independence assumption in OLS regression and may introduce negative bias in standard errors, therefore inflating alpha levels in statistical tests. This negative bias increases in severity as the degree of correlation among clustered observations increases. When the proportion of variance in outcomes that is between groups (the intra-class correlation, or ICC) approaches zero, the consequences of clustering are nil. In this study, very little of the variance in the time 1 measures was between schools. The ICCs ranged from 0.000 to 0.025 and for the most part did not reach nominal levels of statistical significance. The largest proportions of variance between schools were found for disruptive classroom behavior and commitment to abstain from drugs. This suggests that clustering is unlikely to have inflated significance tests statistics that assumed a simple random sample. Nevertheless, we follow the advice of Snijders and Bosker (1999) for handling clustered data when the number of clusters is small (<10), and include dummy variables for schools in all outcome analyses to correct for intercept differences across schools.

enrolled ranged from 36 to 60 (M=45). As reported in Gottfredson et al. (2009), 54% of the treatment sample withdrew from the ASP before the end of the year, and students attended an average of only 36 days of the possible 96. Drop-out occurred in every month during which the program operated, but was highest during the middle of the year, from November through January. The drop-out and attendance rates also varied by site. Site B had much less attrition than other sites with only 32% of students withdrawing before the end of the year. All other sites lost 50% or more of their participants to drop-out. Site A lost 67% while Site C lost 70%. Site B students also attended considerably more days (45.6) than students at other sites. Sites A and C had the lowest attendance, approximately 30 days. The average number of students present per day ranged from 11.6 to 24.2 (M=17.0), and the average number of staff present at each site varied from 2.9 to 3.5, (M=3.2). All of these factors combined to produce an average student to staff ratio ranging from 4.0 to 7.7 (M=5.4).

The sites delivered an average of 26 of the 27 possible All Stars lessons over an average of 102 sessions. The number of sessions is large because students were often split into two or three groups to create smaller All Stars classes[7] and because lessons were regularly delivered over two sessions on consecutive days. Most of the time staff delivered sessions for 45 min, as planned. Low program participation decreased student exposure to the All Stars program. Most of the youth who ever attended the ASP (N=205), received at least one All Stars session (91%). Of those youth who ever participated in All Stars (N=187), the average student received 11.3 lessons and 15.7 h of instruction (Table 3). Variation in the quantity of All Stars programming did not differ significantly across site. Also, although considerable variation across students was observed for the number of sessions received, attendance appeared to be almost random. That is, among the 224 treatment students, the number of All Stars sessions received was not correlated with any of the demographic or pre-test measures from the youth survey. It was significantly correlated with only two of the pre-test measures from school records: Students in the higher grade levels (p<0.05) and students who had more days absent from school in the year prior to the ASP (p<0.01) attended fewer All Stars sessions. Although staff turnover made it impossible for every lesson to be delivered by a trained staff member, 92% of sessions were led by trained staff. Furthermore, 35% of sessions were taught by a group leader who was a certified teacher.

[7] The grouping for All Stars was haphazard, depending on the number of students present on each day. The All Stars groups were not systematically recorded.

Table 3 Quantity of All Stars received

	N	Range	Mean	SD
Sessions attended—all treatment	224	0–49	17.46	14.24
Sessions attended—youth who ever attended	187	1–49	20.92	13.06
Lessons received—all treatment	224	0–24	9.47	7.27
Lessons received—youth who ever attended	187	1–24	11.34	6.48
Total hours of All Stars received—all treatment	224	0–38.67	13.14	10.77
Total hours of All Stars received—youth who ever attended	187	0.67–38.67	15.73	9.90

Variation by site was not statistically significant

As described in the procedures section, we obtained four measures of implementation quality from observer or staff reports. Unlike the quantity of All Stars delivery and student dose, quality of All Stars implementation differed by site (see Table 4). The average observer quality rating was high (M=3.62; range 1–5); however, site B was rated significantly higher on session quality compared to all other schools (M=4.64). Observer ratings of student engagement also suggest that all sites engaged students in the curriculum (M=3.00; range 1–4). Each lesson contained a number of activities with unique objectives. Staff were asked to report which activities they taught and to rate the level of objective achievement for each. Staff reported a high percentage of activities taught per lesson (89%) and a high level of objective achievement (M=3.11; range 1–4). Sites B and C taught a significantly lower percentage of activities per lesson. Site C reported a significantly higher level of objective achievement for the activities that they taught. Site E reported a significantly lower level of objective achievement than other sites while also reporting the highest percentage of activities taught per lesson.

Group leaders were also asked to indicate if they modified the activities. In general, the sites reported a low level of modification (M=0.34) (range 0–2; where 0 meant the teacher did not modify the activity at all and 2 meant the activity was modified a lot); however, site A reported a significantly higher level of modification (M=0.67, p<0.01) than other sites.

Across all the implementation measures collected, site B's All Stars implementation was judged as the highest quality. Sites A and C also delivered All Stars with moderate quality, while sites D and E were ranked as having delivered All Stars with the worst quality. The dimensions of quality tended to co-vary with the exception of percentage of activities taught per lesson. Programs which showed low quality in other areas reported high

Table 4 Quality of All Stars delivery, by site

	Quality (OR)			Engagement (OR)			Objective achievement, per session (SR)			% of activities taught, per lesson (SR)		
	N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Site A	10	3.55	1.06	10	3.00	0.47	93	3.11	0.59	56	91.16	0.21
Site B	21	4.64[a]	0.53	22	3.18	0.55	86	3.17	0.51	45	77.11[a]	0.22
Site C	12	3.21	0.89	12	3.08	0.51	110	3.52[a]	0.65	47	82.16[a]	0.22
Site D	14	3.11	0.96	14	2.86	0.77	102	3.10	0.54	50	92.90	0.16
Site E	8	2.50	0.71	8	2.62	1.06	108	2.67[a]	0.53	53	97.26	0.10
Total	65	3.62	1.10	66	3.00	0.67	499	3.11	0.63	251	88.59	0.20

OR observer report, *SR* staff report

[a] Significantly different than at least one other site, $p<0.01$

levels of activity completion. Perhaps this discrepancy is attributable to the role of certified teachers at Sites A, B, and C. At these sites, some or most All Stars sessions were led by teachers, while no sessions were led by teachers at Site D or E. Teachers may feel more confident omitting parts of a curriculum that they do not consider essential.

Observations of activity structure and student engagement for all program activities allowed us to compare these dimensions across activity types. Compared to academics (typically homework assistance; *M* structure= 2.89, *M* engagement=0.52) and recreation activities (e.g., sports, games, computer time; *M* structure=2.91, *M* engagement=0.82), All Stars (*M* structure=3.63, *M* engagement=0.81) was the most structured and as engaging as recreation activities. Because recreation activities were included to entice students to attend, it is notable that the All Stars curriculum was comparable in terms of student engagement.

Finally, two existing evaluations of All Stars measured quality of implementation (as reported by trained observers) using the All Stars checklists provided by the developer. Direct comparisons of implementation quality in our study and these earlier studies can not be made because the response sets used differ slightly,[8] but it appears that the quality of implementation was lower than was reported in the earlier efficacy trials, as anticipated. Our overall quality rating was 3.62 on a scale from 1 to 5, and 55.4% of sessions were rated 4 or higher. Hansen (1996) reported overall quality ratings of excellent (52.7%) or good (46.7%) with few exceptions. Harrington et al. (2001) reported that 85% of sessions were rated as excellent or good.

Effects on Mediators and Outcomes

Table 5 shows the adjusted means, by group, from regression analyses testing the effect of treatment on All Stars mediators and outcomes. These analyses revealed that students in the treatment and control groups did not differ on any of these measures. In fact, the adjusted means were very similar across groups. The effect sizes were also small and did not display a pattern favoring one group over the other. Overall, All Stars does not appear to have been effective in preventing problem behavior or promoting healthy behavior in the treatment group youth.

Effects by Dosage and Quality

Implementation data reported earlier indicated considerable variability in both quality of delivery across sites and amount of the curriculum received across students. We investigated whether students who attended higher quality sites, received a higher dose of the curriculum, or both reported more positive outcomes. Regression equations similar to those used to test the effect of treatment were employed to test for differential effects by quality. In these equations, an interaction term combining treatment status and enrollment at sites A, B or C was included in the model.[9]

These analyses revealed 1 significant treatment by quality interaction out of the 18 examined: Treatment students attending the lower quality sites experienced significant deterioration in decision-making skills ($p<0.05$) relative to the control students from these sites. Treatment students attending the higher quality sites were not significantly different from their control peers in terms of decision-

[8] Hansen (1996) used a three-item response set for each question compared to our four- and five- item response sets. The response sets in Harrington et al. (2001) were not identified.

[9] We considered testing for interactions by the use of certified teachers, following McNeal et al. (2004) who showed that positive effects were found only for the teacher condition. However, in our study, such analyses were redundant with the analyses of interactions by quality because only in sites A, B, and C were lessons led by certified teachers.

Table 5 Adjusted post-test mean and effect sizes for targeted mediators and outcomes

	Control (*N*)	Treatment (*N*)	*p*-level	ES or OR
Mediators				
Normative beliefs	0.65 (205)	0.62 (210)	0.36	−0.10
Commitment	0.76 (205)	0.76 (210)	0.35	0.00
Bonding	1.99 (202)	2.00 (210)	0.85	0.01
Parental monitoring	3.27 (205)	3.26 (208)	0.95	−0.01
Goal setting	3.06 (203)	2.98 (210)	0.12	−0.14
Decision-making skills	2.72 (205)	2.70 (211)	0.77	−0.03
Impulsivity	0.56 (199)	0.55 (206)	0.92	−0.01
Vulnerability to drug use	0.36 (194)	0.32 (200)	0.85	1.04
Social competency	2.74 (198)	2.68 (207)	0.18	−0.13
Outcomes				
Last month drug use	0.18 (201)	0.19 (207)	0.91	1.03
Smoking initiation	0.09 (194)	0.06 (185)	0.90	0.95
Drinking initiation	0.16 (171)	0.16 (170)	0.75	1.10
Marijuana initiation	0.11 (191)	0.01 (195)	0.47	0.72
Inhalant initiation	0.09 (188)	0.09 (188)	0.17	0.73
Other initiation	0.02 (199)	0.04 (206)	0.40	1.73
Disruptive classroom behavior	1.66 (200)	1.59 (206)	0.14	−0.12
Aggression	1.88 (202)	1.86 (208)	0.70	−0.03
Delinquent behavior	0.82 (202)	0.90 (210)	0.61	0.05

Means adjusted for pre-test level of each variable, age, gender, race and school site. For the mediators, positive effect sizes and odds ratios greater than "1" favor the treatment group.

making skills ($p>0.05$). Given the large number of tests conducted, the one interaction that reached conventional levels of statistical significance must be regarded as potentially due to chance. The dosage analyses revealed that the number of All Stars lessons actually received was not significantly correlated with any of the outcomes included in our analysis, either for the treatment group only or for the entire sample, and either overall or in the higher quality sites. In order to rule out the possibility that these small zero-order correlations were confounding larger dosage effects with selection effects (which might occur if higher risk youths attended more lessons), we also conducted regression analyses that controlled for all covariates included in the intent-to-treat analysis as well as pre-test school truancy, which was the only pre-treatment study variable (other than grade level) related to number of lessons attended. None of these analyses changed the conclusion that dosage was not related to the outcomes examined in this study.

Conclusion and Discussion

This research explored the use of ASPs as an alternative venue in which to provide prevention programming. Specifically, we tested the effectiveness of the All Stars prevention curriculum, as implemented in a year-long school-based ASP. This is the second independent replication of the effects of All Stars on targeted mediators and problem behaviors, and the first study of All Stars that used random assignment of all students to treatment and control conditions.

We found that, although the quality of All Stars delivered in the context of this effectiveness trial was not as high as the level reported in previous efficacy trials conducted by the developer of the program, it was delivered with reasonable integrity to the program design in the five ASP sites. The sites delivered on average 26 of the 27 possible lessons with relatively little modification, 89% of the activities included in the curriculum were taught, and the observer and staff-ratings of the quality of delivery were high (e.g., observers' rating of overall quality was 3.6 on a five-point scale). This level of adherence to the program is comparable to the levels reported from the evaluation of the Communities that Care prevention system in which a variety of evidence-based program models (including All Stars) were implemented in community settings (Fagan et al. 2008). In that study, the percentage of objectives met according to implementers ranged from 73% to 99%, with a median of 90%. The percentage of expected activities taught in the typical lesson in our study was 89% (see Table 4). The amount of the program *delivered* in our study exceeded the typical level reported in Fagan et al. (2008).

Despite this reasonable level of fidelity in program delivery, student exposure to the All Stars lessons was lower than expected due to low levels of attendance in the ASP. Although considerable variability in dosage was observed across students, the typical student who attended the program at all received 16 h of All Stars instruction and participated in 11 All Stars lessons. The average hours received was approximately three-fourths (77.5%) of the 20.25 h of total exposure time recommended by the developer (i.e., 27 45-min sessions), and the number of lessons received was 41% of the total 27 possible. Exposure levels were even lower when non-attenders were also included. Results showed no differences between the treatment and control students at post-test on any of the outcomes or mediators. The results suggest that assignment to a group that receives All Stars in the context of a year-long ASP does not produce different outcomes than does participation in whatever other after school activities are typically available to youths.

Analyses taking variability in the quantity or quality of programming into consideration also showed no positive effects of the program for youths receiving higher dosage, higher quality delivery, or both. These disappointing results suggest that ASPs may not be an ideal venue in which to provide effective prevention programming. Although it may be possible to design ASPs that are more attractive to youths, the typical voluntary ASP does not provide the kind of consistent attendance required to achieve program effects. Furthermore, resources beyond what are typically available for ASPs in community settings will be required to achieve high quality implementation. We tapped into an existing network of ASP providers for this study, keeping the pay scale and management structure the same as for other ASPs run by the organization. Although the measures of the quality of program delivery captured in our study suggest that the program was delivered with reasonable integrity, we note that the youth service workers who delivered All Stars in this study were not accustomed to delivering prevention curricula. It is possible that additional resources in the form of more highly skilled personnel, more extensive training, or even more time devoted to program development may have improved the quality of program delivery in ways not directly assessed in our study and produced more positive outcomes.

Yet, although there was room for improvement in the quality of delivery and the level of exposure to the program, our results imply that more consistent attendance and higher quality implementation, at least within the range of quality available in our study, would not have resulted in positive outcomes. Many explanations for the null findings are possible, and our data can not adjudicate among them. Poor fit of the All Stars program to the study population is one possible explanation. Although our sample was similar to those used in prior evaluations in terms of age, it differed in terms of race, socioeconomic disadvantage, and gender. Our sample had more males than females (54%), while all other samples studies contained more females than males. Our study contained an unusually low percentage of Caucasian students (17%), while other studies included between 46% and 83% Caucasian students. No other study had as high a percentage of African American students as we did, and the level of socioeconomic disadvantage reported for our students (59%) is higher than for any the samples reported in the prior studies. Future studies might explore differences across different demographic groups in their response to the All Stars program.

Below we speculate about one other possible explanation for the null results.

Deviancy Training? An Exploratory Analysis

Prior research has suggested that interventions delivered in group settings may produce unintended negative consequences by increasing opportunities for "deviancy training," which occurs when peers reinforce deviant comments or conduct by responding with approval and attention. This can occur in interventions targeting general populations as well as in those targeting high-risk youths. Dishion and colleagues (Dishion et al. 1996, 1999; Patterson et al. 2000) have shown that increased exposure to deviancy training is associated with increased substance use, delinquency, and violence. Although we did not observe negative All Stars effects, it is possible that the anticipated positive effects were offset by a simultaneous negative effect due to deviancy training.

While certainly unintended, All Stars and similar programs provide an unusual opportunity for youths to openly discuss drug use and other negative behaviors with adults without the expectation of reprisal. Simply opening the topic of drug use for discussion may prompt adolescents to tell entertaining or personal stories about drugs. Although group leaders are cautioned against reinforcing pro-drug expressions, some may encourage youths to self-disclose, offering verbal support and sympathy. Laughter and caring from peers and teachers could reinforce perceptions that having personal experiences with drug use leads to positive attention from others.

The possible detrimental effects of deviancy training in the context of preventive interventions have received scant attention in the research literature. However, Hallfors et al. (2007), seeking to understand the mechanisms that might have explained previously-reported (Cho et al. 2005) negative effects of the Reconnecting Youth program, found that teacher encouragement of deviant attitudes was positively related to a measure of program fidelity such that teachers who delivered the program with higher fidelity also provided the most encouragement for deviancy. These same high-fidelity teachers were responsible for producing the most negative student outcomes. The authors suggested

that these negative effects resulted when the high-deviancy training teachers reinforced deviant youths with warmth and attention. This acceptance of deviant behavior by an authority figure may have normalized it or even made such behavior more appealing.

Although not the focus of our study, we collected observation data on deviancy training in the context of activities delivered in the ASPs. Observers measured deviancy training by coding the frequency of deviant talk and behavior and attendant responses of peers and group leaders (on a spectrum ranging from encouraging to chastising). Rorie et al. (2009) provide details on these observations, which were collected for each 5-min segment during 398 discrete activities. These observations, although too few to support a rigorous study of this topic,[10] allow us to comment on the extent to which deviancy training might have occurred more often in the context of the All Stars intervention and whether it was particularly pronounced in high-implementation sites. We examined the observation data for evidence that deviant talk would be reinforced more frequently in All Stars than in other activities and that, following Hallfors et al. (2007), this relationship would be strongest in the high implementation schools. We found, first, that instances of deviant talk were more likely to be observed during All Stars sessions than other activities ($p<0.05$). The number and percentage of 5-min segments observed that contained deviant talk was 228 (28% of observed segments) for All Stars sessions and 495 (24%) for all other activities combined. Further, All Stars (vs. all other activities) interacted with implementation quality[11] to produce the highest levels of deviant talk at the sites with the highest quality implementation ($p<0.05$). At the highest quality sites, the percentage of 5-min segments of All Stars sessions that contained deviant talk was 30%, while at the lower quality implementation sites it was 22%. The percentage of observed time segments containing deviant talk did not differ by implementation quality for non-All Stars activities. Finally, implementation quality interacted significantly with All Stars to produce more group leader reinforcement of deviant talk in All Stars at the higher implementation sites. Group leader responses to deviant talk were most positive during All Stars at the more well-implemented site ($p<0.01$). Specifically, the average group leader response to deviant talk[12] in All Stars in the best sites ($M=2.06$) was more reinforcing than the average group leader response at these same sites in other activities ($M=2.17$), and it was more reinforcing than the average group leader response in the lower quality sites either during All Stars ($M=2.40$) or during other activities ($M=2.17$). Expressed instead as percentage of group leader responses that were reinforcing, 26% of instances of deviant talk in All Stars sessions received a positive response from group leaders in the highest quality sites, compared with 4% to 6% in the other three conditions.

Our results, although tentative, suggest that prevention curricula such as All Stars increase opportunities for deviant talk to occur. They also support findings reported in Hallfors et al. (2007) showing that teachers who implemented Reconnecting Youth with the most skill also inadvertently encouraged deviance by normalizing it. Although we have limited statistical power to explore site-specific program effects of deviancy training, exploratory analyses indicated that the effect size for treatment on normative beliefs (the mediator that would be expected to be most responsive to deviancy training) is small but positive (0.09) for students attending lower-implementation schools, while it is of moderate magnitude in the unanticipated direction (−0.25) for students in the higher implementation sites. This pattern of findings suggests that the group leaders in the most well-implemented sites may have created a classroom environment in which students felt more comfortable expressing themselves, but wherein they also learned that drug use and antisocial conduct is commonplace. This finding, if replicated using more rigorous research, suggests that even programs that are designed to reduce problem behaviors and are implemented with high fidelity may produce subtle mechanisms of reinforcement which counteract the intended program effects on mediating attitudes and thus diminish the positive effects of the program. Future research should test this possibility.

In closing, we remind readers that the magnitude of effects for school-based prevention programs is generally in the small range (Gottfredson and Wilson 2003; Wilson et al. 2001), and can be expected to be even smaller when such programs are implemented in community settings under typical implementation conditions rather than under the watchful eye of a research staff (Lipsey and Wilson 1998). This research provides another in a growing body of studies documenting the challenges the prevention field faces as it attempts to move effective prevention practices from researcher-controlled settings into the real world.

[10] The 398 discrete activities included 101 All Stars sessions. However, the number of All Stars sessions in which responses to misbehavior were observed ($N = 74$) is too small to allow for fine-grained analysis of differences across the five sites.

[11] Several combinations of high v. low implementation schools resulted in significant interactions. The strongest effect appeared in analyses comparing the highest implementation site (Site B) to all others, in a pattern that replicated those reported in Hallfors et al. (2007). To maintain consistency with the analyses by program quality reported earlier, we report here the results of the All Stars by quality interaction test in which quality was measured using a dummy variable contrasting sites A, B, and C vs. D and E.

[12] Response is coded 1 = reinforcing, 2 = neutral, 3 = chastising.

References

Allison, P.D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Catalano, R.F., Berglund, M.L., Ryan, J.A.M., Lonczak, H., & Hawkins, J.D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Annals of the American Academy of Political and Social Sciences*, *591*, 98–124.

Cho, H., Hallfors, D., & Sanchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of Abnormal Child Psychology*, *33*, 363–374.

Dishion, T.J., Spracklen, K.M., & Andrews, D.W. (1996). Deviancy training in male adolescent friendships. *Behavior Therapy*, *27*, 373–390.

Dishion, T.J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, *54*, 755–764.

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.

Eysenck, S.B., Easting, G., & Pearson, P.R. (1984). Age norms for impulsiveness, venturesomeness and empathy in children. *Personality and Individual Differences*, *5*, 315–321.

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.

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.

Gottfredson, G.D., & Gottfredson, D.C. (1992). *What about you?*. Ellicott City, MD: Gottfredson Associates.

Gottfredson, G.D., & Gottfredson, D.C. (2001). What schools do to prevent delinquency and promote safe environments. *Journal of Educational and Psychological Consultation*, *12*, 313–344.

Gottfredson, D.C., & Wilson, D.B. (2003). Characteristics of effective school-based substance abuse prevention. *Prevention Science*, *4*, 27–38.

Gottfredson, G.D., Gottfredson, D.C., Czeh, E.R., Cantor, D., Crosse, S.B., & Hantman, I. (2000). *A national study of delinquency prevention in school final report*. Ellicott City, MD: Gottfredson Associates (ERIC No. ED 459 409) (available on line at http://www.gottfredson.com).

Gottfredson, G.D., Jones, E.M., & Gore, T.W. (2002a). Implementation and evaluation of a cognitive–behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science*, *3*, 43–56.

Gottfredson, D.C., Wilson, D.B., & Najaka, S.S. (2002b). School-based crime prevention. In L.W. Sherman, D.P. Farrington, B.C. Welsh, & D.L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 56–164). London: Routledge.

Gottfredson, D.C., Weisman, S.A., Soulé, D.A., Womer, S.C., & Lu, S. (2004). Do after school programs reduce delinquency? *Prevention Science*, *5*, 253–266.

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*, 289–318.

Gottfredson, D.C., Cross, A.B., Wilson, D.M., Connell, N., & Rorie, M. (2009). *A randomized trial of the effects of an enhanced after school program*. College Park: MD. Unpublished report available from the authors.

Hallfors, D., Cho, H., Dishion, T., Cuffee, J., & Hartman, S. (2007). *The role of teachers in school-based interventions for high risk adolescents: An alternative paradigm for predicting deviancy training*. Unpublished manuscript.

Hamilton, L.S., Stecher, B.M., Marsh, J.A., Sloan McCombs, J., Robyn, A., Russell, J., et al. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Corporation.

Hansen, W.B. (1996). Pilot test results comparing the All Stars program with seventh grade D.A.R.E.: Program integrity and mediating variable analysis. *Substance Use & 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., & Yungbluth, 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 & Behavior*, *28*, 533–546.

Huizinga, D., & Esbensen, F. (1990). *Scales and measures of the Denver Youth Survey*. Boulder, CO: Institute of Behavioral Science.

Kane, T.J. (2004). *The impact of after school programs: Interpreting the results of four recent evaluations*. The William T. Grant Foundation Working Paper.

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.

Lipsey, M.W., & Wilson, D.B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D.P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful intervention* (pp. 313–345). Thousand Oaks, CA: Sage.

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 & Behavior*, *31*, 165–178.

Patterson, G.R., Dishion, T.J., & Yoerger, K. (2000). Adolescent growth in new forms of problem behavior: Macro- and micro-peer dynamics. *Prevention Science*, *1*, 3–13.

Rorie, M., Gottfredson, D.C., Cross, A., Wilson, D., & Connell, N. (2009). *Structure and deviancy training in after-school programs*. Manuscript submitted for publication.

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 Evaluation*, *23*, 387–419.

Slater, M.D., Kelly, K.J., Edwards, R.W., Thurman, P.J., Plested, B.A., Keefe, T.J., et al. (2006). Combining in-school and community-based media efforts: Reducing marijuana and alcohol uptake among younger adolescents. *Health Education Research*, *21*, 157–167.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Tanglewood Research. (2008a). *All Stars quality assessments*. Retrieved July, 2006, from www.allstarsprevention.com/programs/programsEvaluationQA.asp.

Tanglewood Research. (2008b). *All Stars student surveys*. Retrieved July, 2006, from www.allstarsprevention.com/programs/programsEvaluationOA.asp.

Wilson, D.B., Gottfredson, D.C., & Najaka, S.S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology*, *17*, 247–272.

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