Journal of Quantitative Criminology, Vol. 17, No. 3, 2001

School-Based Prevention of Problem Behaviors: A Meta-Analysis

David B. Wilson,[1,2] Denise C. Gottfredson,[1] and Stacy S. Najaka[1]

This study examines the features of effective school-based prevention of crime, substance use, dropout/nonattendance, and other conduct problems. It summarizes, using meta-analytic techniques, results from 165 studies of school-based prevention activities that ranged from individual counseling or behavior modification programs through efforts to change the way schools are managed. The results highlight several inadequacies in the existing research for guiding policy and practice, the most notable of which is that many popular school-based prevention approaches have not been well studied to date. The study shows, however, that school-based prevention practices appear to be effective in reducing alcohol and drug use, dropout and nonattendance, and other conduct problems. The size of the average effect for each of the four outcomes was small and there was considerable heterogeneity across studies in the magnitude of effects, even within program type after adjusting for measured method and population differences. Non-cognitive-behavioral counseling, social work, and other therapeutic interventions show consistently negative effects, whereas self-control or social competency promotion instruction that makes use of cognitive-behavioral and behavioral instructional methods show consistently positive effects. Also effective are noninstructional cognitive-behavioral and behavioral methods programs. Environmentally focused interventions appear to be particularly effective for reducing delinquency and drug use.

KEY WORDS: prevention; drug use; problem behavior; delinquency; quantitative review.

1. INTRODUCTION

Conduct problems, substance use, and other forms of crime are common among teenagers. In 1999, 42% of twelfth-grade students in the United States reported that they had used illicit drugs in the past 12 months, and 74% reported having used alcohol (Johnston *et al*., 1999). The percentage

[1]Department of Criminology and Criminal Justice, University of Maryland, College Park, Maryland 20742-8235.
[2]To whom all correspondence should be addressed. E-mail: dwilson@crim.umd.edu

0748-4518/01/0900-0247$19.50/0

of seniors admitting minor criminal behaviors in the past year is quite high: Approximately 30% reported stealing something worth less than $50 and taking items from a store without paying for them. More serious violent crimes are less prevalent but are still common: 15% say that they got into a serious fight at school or work; 12% say that they hurt someone badly enough that the victim needed bandages or other medical attention; 4% say that they used a knife, gun, or other weapon to get something from a person; and 6% say that they carried a weapon such as a gun, knife, or club to school. Between 9 and 14% of youths reported crimes involving major theft and property damage in the past year (Johnson *et al*., 1997). These problem behaviors are harmful and costly to society.

Furthermore, individuals who engage in one form of the above behaviors are more likely to engage in others. Huizinga and Jakob-Chien (1998) showed, for example, that between 37 and 51% (depending on the sample) of youths aged 13 to 17 who committed serious crimes also used drugs, compared with between 1 and 3% among nondelinquent youths. These serious delinquents are also far more likely to experience a variety of mental health problems including aggression, obsessive-compulsive disorder, and depression than are nondelinquents. Not surprisingly, serious delinquents are also three to four times more likely than nondelinquents to be receiving poor school grades, be truant, and be suspended from school. Evidence for the cooccurrence of these behaviors is indisputable (Huizinga and Jakob-Chien, 1998). This paper examines the effectiveness of the subset of prevention practices that occur in schools or are implemented by school staff and are designed to reduce the occurrence of these problem behaviors.

1.1. Previous Reviews of Prevention Programs

The most recent reviews of drug prevention efforts (Botvin, 1990; Botvin *et al*., 1995; Dryfoos, 1990; Durlak, 1995; Ennett *et al*., 1994; Gerstein and Green, 1993; Gorman, 1995; Gottfredson, 1997; Gottfredson, 2001; Gottfredson *et al*., 2001; Hansen, 1992; Hansen and O'Malley, 1996; Hawkins *et al*., 1995; Institute of Medicine, 1994; Norman and Turner, 1993; Tobler, 1992; Tobler and Stratton, 1997) generally concluded that substance abuse prevention efforts are effective for preventing substance use. Consistent with the general conclusions of the reviews on drug prevention efforts are the recent reviews and meta-analyses of studies aimed at reducing conduct problems and delinquent behavior (Catalano *et al*., 1998; Dryfoos, 1990; Durlak, 1995; Gottfredson, 1997, 2001; Gottfredson *et al*., 2001; Hawkins *et al*., 1998; Institute of Medicine, 1994; Lipsey, 1992; Lipsey and Derzon, 1998; Lipsey and Wilson, 1993; Samples and Aber, 1998; Stage and Quiroz, 1997; Tremblay and Craig, 1995). These reviews have established that at least some forms of prevention work to reduce delinquency,

substance use, and other forms of problem behaviors and have led to a spirit of optimism about the effectiveness of prevention.

Beyond the general agreement that something works, however, is much uncertainty about the magnitude of the effects of prevention programs, the characteristics of effective prevention activities, and the receptivity of different segments of the population to prevention programming. The most useful studies for describing the magnitude of effects are meta-analytic studies that report post program outcome differences between the experimental and the comparison groups. One of the earliest meta-analyses of school-based prevention programs was conducted by Tobler (1986; see also Tobler and Stratton, 1997), who synthesized effect sizes derived from 98 research studies. The mean effect across studies indicated that young people who had received prevention services scored lower on measures of substance use after the program than those youths who had not received such services. This effect, although significantly different from zero, is modest.

The meta-analytic studies in this area document wide variability in the magnitude of effects from study to study. Gottfredson *et al.* (2001), for example, report effect sizes on measures of problem behaviors ranging from -0.86 to 3.09 across 178 studies of school-based prevention programs. Understanding which characteristics of prevention activities account for this variability is an important first step in designing more potent strategies.

Some of the observed variability in program effects is due to methodological differences across studies. These differences must be taken into account when examining substantive features of programs, for method characteristics are often correlated with the substantive characteristics of interest. Lipsey (1992) found that method variance accounted for 25% of the variance in effect sizes on recidivism outcomes. Once these factors were controlled, characteristics of the treatment accounted for an additional 22% of the effect size variability. Treatment characteristics associated with larger effects included programs targeting higher risk juveniles, programs providing larger amounts of meaningful contact with the youths, behavioral, skill-oriented, and multimodal programs, and programs judged to have a more sociological and less psychological orientation.

Reviews of efforts to reduce substance use have produced similar results. Hansen (1992) meta-analyzed studies of school-based substance abuse prevention curricula published during the 1980s and found that, after controlling for methodological features of the studies, social influence programs (e.g., those focusing on resistance skill training and often including norm-setting activities and pledges to remain drug-free) and comprehensive programs (e.g., those similar to the broadest social influence programs, but

also targeting other skills, such as more general decision-making skills) were most successful for reducing substance use.

Based on a meta-analysis of school-based drug prevention program, Tobler and Stratton (1997) concluded that "interactive" programs were more effective than "noninteractive" programs. They also showed that program content categories (e.g., social influence, information only, affective, etc.) were correlated with mode of delivery (interactive vs noninteractive) and suggested that some of the positive effect previously attributed to program content may in fact be due to the delivery method.

These meta-analyses converge in suggesting that focused approaches that teach specific behavioral and cognitive-behavioral skills are more effective than other strategies for reducing problem behaviors. But the reviews at the same time suggest that approaches that focus on normative change (Hansen, 1992) and other sociological variables (Lipsey, 1992) are also effective, as are multimodal or comprehensive programs (Durlak, 1995; Lipsey, 1992). The reviews also suggest that the methods may be as important or more important than the content (see also, Wilson, 1995).

This study is designed to increase knowledge about the features of effective school-based prevention. It focuses on school-based prevention because of the importance of the role of schools in providing prevention services (Gottfredson, 2001). It summarizes, using meta-analytic techniques, results from 165 studies of school-based prevention activities. These activities ranged from individual counseling or behavior modification programs through efforts to change the way schools are managed. The study is intended to determine, on the basis of currently available studies, what types of school-based prevention programs (in terms of both outcomes sought and activities implemented) are related to variability in the size of program effects. The study differs from previous meta-analytic reviews in the following ways: (a) it includes only school-based programs, e.g., those taking place in school buildings or run by school personnel; (b) it includes all types of school-based programs rather than only classroom-based instructional programs (e.g., Hansen, 1992) or universal programs (e.g., Tobler and Stratton, 1997); and (c) it includes studies measuring a wide range of problem behaviors rather than being focused only on substance use or delinquency as has been the case in many prior reviews. A previous review of this literature by the authors (Gottfredson *et al*., 2001) was more descriptive and less meta-analytic.

2. METHODS

2.1. Inclusion Criteria and Search for Relevant Studies

The studies included in this synthesis represent a subset of those detailed in the review by Gottfredson *et al*. (2001). To be included, a study

had to meet the following criteria: (a) it evaluated an intervention, that is, a distinct program or procedure intended to reduce problem behaviors among children and youth; (b) the intervention was school-based, that is, the intervention was operated in a school building, by school staff, or under school or school system auspices (e.g., classroom instruction or other classroom activities, schoolwide environmental changes, or modifications of teacher or school administrators behaviors and instructional practices); (c) it used a comparison group evaluation methodology, including nonequivalent comparison group research designs, and the comparison group was a no-treatment or minimal-treatment condition; and (d) it measured at least one outcome of interest to this review. Outcomes of interest included indicators of (a) crime, delinquency, theft, violence, and illegal acts of aggression; (b) alcohol and other drug use, excluding cigarette and smokeless tobacco use; (c) withdrawal from school, school dropout, nonattendance, or school tardiness; and, (d) rebellious behavior, antisocial behavior, aggressive behavior, defiance of authority, disrespect for others, suspension/expulsion, or other acting-out behavior.

Excluded from this synthesis were studies that did not report sufficient information to determine the direction of the effect for at least one of the four outcome categories specified above. Studies were included even if they did not report sufficient information to compute an effect size, contingent on our ability to determine the direction of the effect, that is, whether the intervention or comparison group was favored on at least one outcome of interest.

Potentially eligible studies for inclusion in this synthesis were identified through searches of computer bibliographic databases (e.g., PsychLit, ERIC, and Sociological Abstracts) and through the references of recent reviews of prevention programs (Botvin, 1990; Botvin *et al.*, 1995; Dryfoos, 1990; Durlak, 1995; Hansen, 1992; Hawkins *et al.*, 1995; Institute of Medicine, 1994; Tobler, 1986, 1992; Tremblay and Craig, 1995). In some instances, the search of recent reviews resulted in the identification and inclusion of a number of unpublished studies. This list was augmented with additional studies already known to the authors. Due to cost considerations, doctoral dissertations were excluded.

A total of 165 studies, representing 219 documents, met these criteria. A list of included studies is available from the authors. The vast majority of these studies were published in peer-reviewed journals (80%), with slightly over 10% published in some other form and slightly less than 10% unpublished.

2.2. Coding Unit: Treatment-Comparison Contrasts

It was common for studies to report on multiple treatment–comparison contrasts or to otherwise report data in a disaggregated way that repre-

sented a meaningful distinction for the purpose of this review. Examples include two distinct interventions compared to a single control group, the same intervention applied to distinct age groups, and an intervention examined after differing amounts of treatment (e.g., 1 year of a program, 2 years of a program, a full-program plus booster, etc.). For the present meta-analysis, treatment-comparison contrasts that represented the former two examples were included. Only the contrast representing the full intervention was included for studies of the latter type. In the example given, only the full-program plus booster was included. This is in contrast to Gottfredson *et al.* (2001) that included all of these possible contrasts. Thus, this meta-analysis included multiple treatment-comparison contrasts from some individual studies when the distinct contrasts represented unique interventions or when the program participants represented distinct groups of students. In some cases, these multiple treatment–comparison contrasts shared a comparison group. Although this introduced statistical dependencies in the data, this was balanced against the potential benefit of examining these programs in a more differentiated fashion. The statistical method for handling these dependencies is discussed below. In all, 216 contrasts were coded from the 165 studies included in the synthesis.

2.3. Coding of Study Characteristics

A code book similar to a survey form was developed to capture information regarding the specific nature of the intervention, characteristics of the student population, research methodology, measures of problem behaviors, and observed effects on these measures at all measurement points. Studies were coded by trained graduate students who meet weekly to discuss coding decisions. To improve reliability, all studies were coded by at least two coders and all coding discrepancies were discussed and resolved. A copy of the code book can be obtained from the authors.

2.3.1. Program Categories

A major challenge in reviewing a vast literature of this nature is grouping the interventions into conceptually meaningful categories. The program categories we developed emerged from our interaction with the studies, and we believe represent a reasonable and meaningful categorization of these interventions. Each treatment program was assessed for the presence or absence of 17 treatment components or activities (e.g., instruction, cognitive-behavioral or behavioral modeling, reorganization of grades) using a classification system developed for use in the National Study of Delinquency Prevention in Schools (Gottfredson *et al.*, 2000). Gottfredson (1997) and

Gottfredson *et al*. (2001) also describes the categories. If a treatment component was present, a judgment was made as to whether it was a major or minor component of that intervention. An iterative process was undertaken using these codes to group programs into mutually exclusive program categories, resulting in 11 program categories (see Table II). These 11 intervention categories are further grouped as being either environmentally focused or individually focused. Although other categorizations of these programs are possible, we believe that this categorization is both conceptually meaningful and consistent with the actual practices of school-based prevention programs.

2.3.2. Student Characteristics

The nature of the student population participating in the school-based prevention program was captured by a set of items addressing the age, grade range, gender, and racial distribution represented in the study. The coding protocol also captured a written description of the student sample, often taken verbatim from the written report or published document. We also coded whether the study represented the general school population or was restricted in some way to a high-risk group, such as youths with a criminal history or reported drug use.

2.4. Research Methodology

The soundness of the empirical evidence was assessed with seven items in the coding protocol. These items addressed assignment to conditions (e.g., random assignment to conditions), unit of assignment (e.g., student, class, school, etc.), unit of analysis, use of control variables in analyses to adjust for initial group differences, rating of initial group similarity, attrition, and an overall 5-point evaluation of methodological quality. The latter item, called the Scientific Methods Score (see Sherman *et al*., 1997), was informed by answers to the method rigor items and had the following anchors to assist the coders in making consistent ratings: (1) no reliance or confidence should be placed on the results of this evaluation because of the number and type of serious shortcoming(s) in the methodology employed; (3) methodology rigorous in some respects and weak in others; and (5) methodology rigorous in almost all respects. The double coding of these items by two graduate students and discussion of discrepancies helped improve the reliability of the final method score.

2.5. Program Effects

The effectiveness of the program on each available outcome was coded using the standardized mean difference effect size, a measure of the difference between the program and the comparison groups relative to the

standard deviation of the measure employed. Effect sizes were coded such that positive values always meant that the experimental group had a more desirable outcome than the comparison group, independent of the direction of the original scale reported in the study. Whenever possible, the post treatment or follow-up mean difference was adjusted for any baseline mean difference on that measure. The standardized mean difference effect size can be computed from a wide variety of data configurations reported by the primary studies (Lipsey and Wilson, 2001). The effect size (d) was defined as

$$d = \frac{\bar{X}_T - \bar{X}_C}{s_{\text{pooled}}} \tag{1}$$

where $\bar{X}_T$ and $\bar{X}_C$ are the program and comparison means, respectively, and s_{pooled} is the pooled within-groups standard deviation (see Hedges and Olkin, 1985). When baseline means were available, the effect size was computed as

$$d = \frac{\bar{\Delta}_T - \bar{\Delta}_C}{s_{\text{pooled}}} \tag{2}$$

where $\bar{\Delta}_T$ and $\bar{\Delta}_C$ are the posttest or follow-up means minus the baseline means for the program and comparison means, respectively, and s_{pooled} is the pooled within-groups posttest or follow-up standard deviation. Hedges (1981) showed that the above equation is upwardly biased when based on small sample size and provided a correction equation,

$$d' = \left(1 - \frac{3}{4N - 9}\right) d \tag{3}$$

where N is the combined sample size of the intervention and control groups. This correction was applied to all effect sizes based on Eqs. (1) and (2) above.

The standard deviation in the above equation was pooled across both the intervention and the comparison groups and must have been based on the variability across students, not the variability across classrooms or schools. A standard deviation for a level of analysis higher than the student level would tend to be smaller and thus would upwardly bias the effect size relative to an effect size based on student-level data.

Of the 551 effect sizes computed for this meta-analysis, 227 (41%) were based on the above formulas. An estimate of d can be computed from proportions based on dichotomous data, such as the proportion of students using marijuana, using the probit transformation. The effect size for an outcome reported as proportions was computed as

$$d = \Phi(p_T) - \Phi(p_C) \tag{4}$$

where $\Phi(p_T)$ and $\Phi(p_C)$ are the probits associated with the proportions (p) in each group with positive outcomes (see Rosenthal, 1994). When baseline proportions were available, the baseline probit for the program and comparison group was subtracted from the posttest or follow-up probit in the above formula. Roughly half (262, or 48%, of the 551) of the effect sizes computed for this meta-analysis were based on proportions. The remaining 11% of the effect sizes were estimated using other methods discussed by Lipsey and Wilson (2001).

Only 145 of the 216 intervention contrasts coded for this meta-analysis provided sufficient information to directly compute an effect size for at least one of the four outcome categories. The direction of effect for at least one outcome category, and in some cases whether or not the effect was statistically significant, was determined for all 216 intervention contrasts. The robustness of the effect size analyses based on the 145 contrasts will be examined by imputing a conservative effect size for all effects with a known direction.[3] This provides some assurance that the observed effects are not simply a function of researchers reporting more statistical detail for positive findings. Analyses were performed both with and without these imputed values.

The effect sizes were categorized into one of four broad outcome categories (criminal behavior, alcohol and other drug use, school dropout and nonattendance, and other problem behaviors). Measures of delinquency included a wide variety of acts for which individuals could be arrested, and in these studies, delinquency was frequently measured using youth self-reports. Common operationalizations of delinquency included age at first involvement, current criminal activity, and frequency of delinquent involvement. Alcohol and drug use also was most often measured using youth self-reports and operationalizations of use included status as having used alcohol or other drugs at least once, frequency of use, and amount typically

[3]If the effect was reported as statistically significant, then the effect size was imputed based on a t value associated with a two-tailed probability value of 0.05, with the degrees of freedom equal to the sum of the intervention and comparison groups minus two. This estimates what the smallest statistically significant effect size would be for that study on that outcome and will tend to underestimate the true effect, because a statistically significant effect rarely has a probability value exactly equal to 0.05. For effects that were statistically nonsignificant or where statistical significance was not reported, the effect was imputed based on the t value associated with a two-tailed probability value of 0.50. This imputed a small nonzero effect size in the direction reported by the study, and we believe that is preferable to imputing an effect size of zero in these cases, for the direction of the effect is known. The 25th and 75th percentiles for these imputed effect sizes were 0.01 and 0.14, with a median of 0.03. Thus, most imputed values were small. It is common practice in meta-analysis to estimate an effect size from incomplete data. To omit these studies would potentially upwardly bias the findings, for significant findings are more likely to be reported and reported in greater detail.

used. Dropout and nonattendance were most often measured using official school records. Other problem behaviors encompassed a variety of behaviors and were defined to include rebellious behavior, noncriminal antisocial behavior, aggressive behavior, defiance of authority, disrespect for others, school suspension, and school expulsion. While school suspension and school expulsion are most often measured using official school records, the remaining conduct problems (i.e., rebellious behavior, noncriminal antisocial behavior, aggressive behavior, defiance of authority, and disrespect for others) were most commonly measured using teacher and parent reports of behavior.

2.6. Statistical Analysis

This meta-analysis used the method of inverse variance weighting in the analysis of effect sizes, that is, each effect size was weighted by the inverse of its estimated variance. This has the effect of giving greater weight to studies based on larger samples. The inverse variance of effect sizes based on means and standard deviations was computed as

$$v_{\mathrm{d}} = \frac{n_{\mathrm{T}} + n_{\mathrm{C}}}{n_{\mathrm{T}} n_{\mathrm{C}}} + \frac{d'^2}{2(n_{\mathrm{T}} + n_{\mathrm{C}} - 2)} \tag{5}$$

where n_{T} and n_{C} are the intervention and comparison group sample sizes, respectively. The inverse variance of effect sizes based on proportions was computed as

$$v_{\mathrm{d}} = \frac{2\pi p_{\mathrm{T}}(1 - p_{\mathrm{T}})\, e^{\Phi(p_{\mathrm{T}})}}{n_{\mathrm{T}}} + \frac{2\pi p_{\mathrm{C}}(1 - p_{\mathrm{C}})\, e^{\Phi(p_{\mathrm{C}})}}{n_{\mathrm{C}}} \tag{6}$$

where the terms are defined as in Eqs. (4) and (5). The inverse variance weight is simply the inverse of these values (i.e., $1/v_{\mathrm{d}}$). Under the random effects assumption that the population effects estimated by distinct studies within program categories varied, a random effects variance component is estimated using maximum-likelihood methods and added to the above inverse variance weights (Lipsey and Wilson, 2001; Raudenbush, 1994). Our assumption of a random effects model was based in part on the large estimates of across study variability from other meta-analyses in this general research domain (e.g., Tobler and Stratton, 1997) and also on the broad nature of this synthesis.

The coding protocol for this meta-analysis allowed for the extraction of multiple effect sizes from an individual study, including both effect sizes for multiple outcomes per intervention–comparison contrast and intervention–comparison contrasts that shared a comparison group were coded. We handled the statistical dependencies among these multiple effect sizes in

two ways. First, for any given intervention–comparison contrast, we included only one effect size per major outcome category in an analysis and we analyzed each outcome category separately. When multiple effect sizes were available within an outcome category, the mean of these effect sizes was computed.

Second, for intervention–comparison contrasts that shared a comparison group, the statistical dependency was estimated and directly modeled in the analysis by incorporating an estimate of the covariance between any two dependent effect sizes into the inverse variance weight matrix used in the analysis of effect sizes. For effect sizes based on means and standard deviations, the covariance was computed as

$$\psi = \frac{1 + \frac{1}{2} d_1 d_2}{n_C} \tag{7}$$

where d_1 and d_2 are the two effect sizes that share a comparison group, and n_C is the sample size for the comparison group (see Gleser and Olkin, 1994). For effect sizes based on proportions, the covariance was computed as

$$\psi = \frac{2\pi p_C (1 - p_C)\, e^{\Phi(p_C)}}{n_C} \tag{8}$$

where the terms are defined as above for Eq. (6). This approach represents an incorporation of the methods developed by Gleser and Olkin (1994) for handling statistically dependent effect sizes and the random effects methods proposed by Raudenbush (1994) and is similar to the methods developed by Kalaian and Raudenbush (1996). The percentage of statistically dependent effect sizes in any given analysis was small. As such, the preliminary analyses performed without the added burden of modeling these dependencies were highly similar to the results from the final analyses that incorporated the covariances.

3. RESULTS

3.1. Description of Studies

Table I shows the sample descriptors for the 216 intervention–comparison contrasts. The most common grade ranges included in the studies were middle/junior high school students, followed by interventions involving late elementary students. Programs for high school students and early elementary students were less common. A few programs involved students from a broad range of grades, including two studies that evaluated an intervention presented to students of all grades.

Table I. Sample Descriptors for the 216 Intervention–Comparison Contrasts

Variable	Frequency	Percentage
School grades included		
Preschool kindergarten	5	2
Early elementary	19	9
Late elementary	56	26
All elementary	18	8
All elementary–middle	1	< 1
Late elementary/middle/junior	8	4
Elementary–senior high school	2	1
Middle/junior high school	68	31
Middle/junior–senior high school	7	3
Senior high school	32	15
Level of criminal involvement		
General school population	155	72
High-risk population	61	28

Most of the interventions evaluated were presented to a general student population (72%). Slightly over a quarter of the interventions were restricted to a student population identified as high-risk for problem behaviors or delinquency. These high-risk samples were predominantly male, with a weighted average proportion of males across samples of 78%. Of the 169 contrasts for which racial makeup was reported, the median proportion of Caucasians for the general population samples was 50%, with a very large range. For the high-risk populations, the median proportion of Caucasians was only slightly less than 50% (49%). Thus, the high-risk samples were quite similar to the general population samples in the proportion of Caucasians but tended to have more males than the general population.

The distribution of intervention types according to our categorization scheme is shown in Table II. Individually focused interventions have been studied to a much greater extent than environmentally focused interventions. The most common environmentally focused interventions were efforts to establish norms or expectations for behavior and classroom or instructional management programs. Of the individually focused interventions, most were instructional programs that included a self-control or social competency component (49% of all interventions), the majority of which used cognitive-behavioral or behavioral instructional methods. The cognitive-behavioral, behavioral modeling, or behavior modification category represented interventions that did not include an instructional component but rather involved teaching new behaviors through modeling, rehearsal, feedback on performance, and reinforcement. Counseling and other therapeutic-type prevention interventions and recreational and leisure activities have been sparsely evaluated.

Table II. Major Intervention Categories for the 216 Intervention–Comparison Contrasts

Variable	Frequency	Percentage
Environmentally focused interventions		
Establish norms or expectations for behavior	18	8
Classroom or instructional management	15	7
School and discipline management interventions	6	3
Reorganization of grades or classes	6	3
Individually focused interventions		
Self-control or social competency (instructional)		
With cognitive-behavioral or behavioral instructional methods	69	32
Without cognitive-behavioral or behavioral instructional methods	36	17
Cognitive behavioral, behavioral modeling, or behavior modification	30	14
Mentoring, tutoring, and work study	13	6
Other instructional	12	6
Counseling, social work, and other therapeutic interventions	7	3
Recreation, community service, enrichment, and leisure activities	4	2

The vast majority of the programs were delivered in a group setting (73%), generally the students' classroom. Many of the remaining programs included both a group and a one-on-one component. Few school-based prevention programs, at least based on this collection of studies, were delivered solely in a one-on-one format. Classroom teachers were the most common persons to administer the program to the students, with teacher involvement in 60% of the evaluated interventions. Peers, mental health professionals, research personnel, police officers, and college students were involved with between 8 and 14% of the interventions.

The coding protocol for the meta-analysis included numerous items to capture methodological variation across studies. Several of these variables are presented in Table III. Almost one-fifth of the intervention–comparison contrasts used random assignment to conditions. Of those contrasts that used nonrandom assignment to conditions, over a third (64 of 174) had program and comparison conditions that were highly similar on pretest data. Unfortunately, close to a quarter of the contrasts observed what we judged to be potentially important differences at pretest. These need not have been statistically significant but generally were.

It has been observed previously that many studies in the research area randomly or nonrandomly assign classes or schools to the treatment conditions yet analyzed the data as though individual students had been assigned (e.g., Gottfredson, 1997). This can clearly be seen in Table III, with only a third of the contrasts having the student as the unit of assignment

Table III. Methodology Descriptors for the 216 Intervention–Comparison Contrasts

Variable	Frequency	Percentage
Assignment to conditions		
Random	42	19
Nonrandom		
Only minor observed group differences	64	30
minor observed group differences	50	23
No group differences information	60	28
Rating of initial group similarities		
(1 = highly dissimilar; 7 = highly similar)		
1–3	63	29
4–5	94	44
6–7	59	27
Unit of assignment to conditions		
Student	69	32
Class	34	16
School	87	40
District or community	6	3
Mixed	14	7
Missing	6	3
Unit of analysis		
Student	196	91
Class	11	5
School	6	3
School and individual	3	1
Study-level selection of measurement		
No attention to measurement	0	0
Some attention to measurement	63	29
Some measures reliable/previously used	113	52
Careful selection of measures	40	19
Overall method rating		
Serious weaknesses	16	7
Moderate weaknesses	25	12
Some weaknesses/some strengths	47	22
Moderate strengths	104	48
Rigorous	24	11

yet over 90% performing inferential analyses at the student level. The effect of this is to overstate the statistical significance (an inflated α error rate) of observed effects. This does not create a problem at the meta-analytic level, for extracted from each study are the descriptive, not the inferential, statistics. Furthermore, all effect sizes based on means and standard deviations used standard deviations based on individuals, not higher units of analysis. As mentioned earlier, effect sizes computed using standard deviations based on classes or schools would produce values on a different and incommensurate metric.

In coding the methodological characteristics of the studies we also made judgments on the initial similarity of the intervention and comparison groups and the overall methodological quality. These judgments were made by two independent coders to improve reliability. As can be seen in Table III, close to two-thirds of the contrasts were placed in the top two categories of methodological quality and roughly half in the top three categories of initial group similarity. The relationship between these methodological variables and effect size is explored below to assess the robustness of the findings to methodological weaknesses in the primary research.

3.2. Effect Size Analyses

Overall, the prevention programs examined by this synthesis appear to have a small positive net effect on problem behaviors (see Table IV). The program effects on school problems (i.e., dropout and nonattendance) and other problem behaviors were roughly three times greater than for delinquency and alcohol and drug use. Although these overall means provide evidence that the school-based prevention programs evaluated were, on average, beneficial, the highly heterogeneous nature of the distributions suggests large differential effects across studies. The analyses below explore the relationship between the method, sample, and program features and the magnitude of the observed program effects.

3.2.1. Methodological Variation and Effect Size

Method differences are the first potential explanatory source of variation in effect size across studies to be examined. This protects against positing the importance of a substantive feature that is confounded across studies

Table IV. Random Effects Mean Effect Size by Outcome

Outcome	$\bar{d}$>	95% CI		k^a
		Lower $\bar{d}$>	Upper $\bar{d}$>	
Observed effect sizes				
Delinquency	0.04	−0.03	0.11	40
Alcohol/drug use	0.05	0.01	0.09	80
Dropout/nonattendance	0.16	0.05	0.27	39
Other problem behaviors	0.17	0.09	0.25	73
Observed/imputed effect sizes				
Delinquency	0.05	−0.01	0.11	47
Alcohol/drug use	0.04	0.02	0.07	103
Dropout/nonattendance	0.13	0.06	0.20	53
Other problem behaviors	0.15	0.10	0.19	122

[a]Number of effect sizes contributing to each analysis.

Table V. Random Effects Mean Effect Size by Method Features

Outcome	$\bar{d}^{>}$	95% CI		k^a
		Lower $\bar{d}^{>}$	Upper $\bar{d}^{>}$	
Assignment to conditions*				
Random	0.25	0.17	0.33	42
Nonrandom	0.08	0.05	0.10	174
Observed pretest differences				
Yes	0.11	0.05	0.17	50
No	0.09	0.06	0.12	166
Students assigned to conditions*				
Yes	0.18	0.12	0.24	69
No	0.07	0.04	0.10	147
Students unit of analysis				
Yes	0.10	0.07	0.13	196
No	0.08	0.01	0.16	20
Careful selection of measure*				
Yes	0.16	0.09	0.23	40
No	0.08	0.06	0.11	176
Overall method rating				
Serious weaknesses	0.07	−0.02	0.16	16
Moderate weaknesses	0.03	−0.04	0.10	25
Some weaknesses/some strengths	0.10	0.04	0.16	47
Moderate strengths	0.10	0.07	0.14	104
Rigorous	0.16	0.06	0.26	24

[a]Number of effect sizes contributing to each analysis.
*$p \leq 0.05$.

with a method feature. Only a single effect size representing problem behavior was used per intervention–comparison contrast in the analyses presented in Table V. It is interesting to note that the randomized designs yielded larger mean effects than the nonrandomized designs. Similarly, the trend for overall method quality with effect size is positive (weighted correlation, not shown in Table V, equals 0.10; $p = 0.07$). This finding is not uncommon in meta-analysis (Lipsey and Wilson, 1993). Also not shown in Table V is the weighted correlation between the effect size and the rating of initial group similarity. Studies judged to have more highly similar intervention and comparison groups tended to observe slightly larger effect sizes (weighted correlation, 0.11; $p = 0.04$).

Although studies that assigned students to conditions observed larger effects than studies that assigned larger units, such as classes, this effect is confounded with the program format. Interventions with a one-on-one format almost exclusively assigned individuals to treatment conditions, whereas a minority (22%) of the group-level interventions assigned individuals to conditions. Interventions with a one-on-one format may be more

effective than group-based approaches. As expected, whether or not the data were analyzed at the student or group level had little effect on the observed effect size.

Two implications follow from these findings. First, study design appears to be related to observed effects and, therefore, needs to be modeled when examining the relationship between substantive features, such as program type, and effect size. Second, the inclusion of weak designs in this study does not appear to have upwardly biased the overall results. Thus, the positive overall findings in Table IV cannot be attributed to the inclusion of methodologically weak studies in this synthesis.

3.2.2. Sample Characteristics and Effect Size

Another source of variability across studies was the characteristics of the children and youth included in the sample. As can be seen in Table VI, studies that were restricted to a high-risk population observed larger effects than interventions directed at the general population. Many of the problems targeted by these programs have a low frequency of occurrence in the general population, constraining the upper bound of the observable effect. That is, it is difficult to decrease a behavior that has a low rate of occurrence prior to any intervention, thus restricting the effect for the general population. This finding is encouraging, illustrating that meaningful reductions in problem behaviors can be achieved with high-risk youths. It is not possible to assess, given this data, whether the overall level of problems at the school level decreases more from an intervention provided to the high-risk youth relative to an intervention provided to all students.

Table VI. Random Effects Mean Effect Size by Sample Character

Outcome	$\bar{d}^{>}$	95% CI		k^a
		Lower $\bar{d}^{>}$	Upper $\bar{d}^{>}$	
School grades				
Early elementary	0.05	−0.06	0.16	19
Late elementary	0.05	0.00	0.11	56
Middle/junior high school	0.09	0.04	0.13	68
Senior high school	0.14	0.06	0.22	32
Level of criminal involvement*				
General school population	0.07	0.04	0.10	155
High-risk population	0.20	0.14	0.21	61

[a]Number of effect sizes contributing to each analysis.
*$p \leq 0.05$.

3.2.3. *Method and Sample Equated Effect Sizes*

To assess the robustness of these findings across study differences in method and sample characteristics, a regression analysis was performed for each outcome. These analyses were used to produce covariate adjusted effect sizes by centering the residuals on the grand mean effect size for that outcome. The analyses of differences in effects across intervention characteristics are performed on both the raw effect size and the covariate adjusted, or method and sample equated, effect sizes. The results from these regression analyses are presented in Table VII.

Few individual method or sample characteristics are statistically significant predictors of effect size. Many of the observed unstandardized regression coefficients, however, are substantial, predicting large changes in the mean effect size associated with the various levels of the characteristics. For example, the regression coefficient for whether the study used random assignment shows that the effect sizes for randomized studies are predicted to be between 0.09 and 0.19 larger than for nonrandomized studies. A substantial portion of effect size variability was accounted for in the regression models for delinquency and dropout/nonattendance (0.33 and 0.18, respectively).

Table VII. Mixed Effects Regression Analysis for the Observed and Imputed Effect Sizes Regressed on Method and Sample Characteristics[a]

	Unstandardized regression coefficient			
Variable	Delinquency	Drug use	Dropout/ nonattendance	Other problem behavior
Early elementary	0.03	−0.08	0.13	0.01
Late elementary	0.04	0.04	−0.07	−0.01
Middle/junior high	0.08	0.02	−0.20*	0.03
Senior high school[b]				
High-risk population	0.23	−0.00	0.16	0.05
Random assignment	0.19	0.13	0.17	0.09
Observed pretest differences	0.33*	0.02	−0.04	0.03
Students assigned to conditions	−0.25	−0.06	−0.14	0.05
Students unit of analysis	−0.10	−0.03	—[c]	−0.03
Careful selection of measures	−0.08	−0.05	−0.00	0.12
Overall method rating	−0.07	0.01	0.06	−0.05
Initial group similarity	0.04	−0.01	−0.00	0.04
Intercept	−0.13	0.05	−0.12	0.07
R^2	0.33	0.04	0.18	0.08

[a] Sample sizes for the regression analyses are 47, 102, 53, and 122, respectively.
[b] Null category.
[c] No variability on this variable for this outcome.
*$p \leq 0.05$.

3.2.4. Intervention Types and Effect Size

A primary focus of this meta-analysis is an assessment of which school-based prevention programs appeared most promising, based on existing empirical evidence. Table VIII presents the weighted random effects mean effect sizes for each of the 11 intervention categories by each of the four outcomes of interest. This table presents both the unadjusted weighted random effects mean effect size and the method and sample equated effect size.

The overall impression from Table VIII is that environmentally focused interventions are generally effective, as are cognitive behaviorally and behaviorally based individually focused interventions, both with and without an instructional self-control or social competency component. Instructional strategies that do not use cognitive behavioral or behavioral instructional strategies; mentoring, tutoring, and work study programs; and recreational programs are not effective. Counseling, social work, and other therapeutic interventions (not elsewhere classified) in schools have negative effects. Note that these ineffective counseling strategies do not include cognitive-behavioral counseling.

For both the delinquency and the alcohol/drug use outcomes, the unadjusted effect size and methods equated effect size analyses are highly similar with no major differences. This suggests that the differences between the mean effect sizes across categories are not attributable to the measured method and sample differences. It is still plausible that other unobserved differences between the studies, other than the characteristics of the intervention, may confound these mean effects.

The two analyses for the dropout/nonattendance outcomes had differences worth noting. The mean effect size for classroom or instructional management programs was substantially reduced and became statistically nonsignificant in the methods and sample equated analysis, as did the effect for instructional self-control or social competency programs with cognitive-behavioral methods. The negative effect for non-cognitive-behavioral counseling and social work-type interventions was comparable in magnitude for both analyses, although it was not statistically significant in the method and sample equated analysis. Several other statistically nonsignificant and generally small effects were attenuated in the method and sample equated analysis.

The intervention effects on other problem behaviors were generally attenuated by the method and sample equated analysis. The positive effect of programs to establish norms and expectations for behavior drops by a third from a small effect to a very small effect. The mean effect size for the reorganization of grades or classes dropped to near zero, as did the small modest effect for other instruction programs. The statistically significant

Table VIII. Random Effects Mean Effect Size by Major Intervention Category and Outcome Category for all 216 Intervention–Comparison Contrasts Using the Observed and Imputed Effect Sizes

	Outcome category											
	Delinquency			Alcohol/drug use			Dropout truancy			Other problem behavior		
Variable	$\bar{d}_r$[a]	$\bar{d}_c$[b]	k	$\bar{d}_r$	$\bar{d}_c$	k	$\bar{d}_r$	$\bar{d}_c$	k	$\bar{d}_r$	$\bar{d}_c$	k
Environmentally focused interventions												
School and discipline management interventions	0.39*	0.16*	2	0.10	0.07	1	0.07	−0.14	2	0.11	0.04	4
Establish norms or expectations for behavior	0.02	−0.07	2	0.09*	0.08*	13	—[c]	—[c]		0.15*	0.05	6
Classroom or instructional management	0.13*	0.19*	5	0.11*	0.10*	6	0.20*	0.14	5	0.04	−0.01	12
Reorganization of grades or classes	0.23*	0.34*	2	0.40*	0.48*	1	0.09	−0.03	4	0.14	−0.01	4
Individually focused interventions												
Self-control or social competency (instructional)												
With cognitive-behavioral or behavioral instructional methods	0.10*	0.10*	9	0.04*	0.05*	32	0.29*	0.09	4	0.18*	0.08*	38
Without cognitive behavioral or behavioral instructional methods	−0.00	−0.00	12	0.03	0.03	34	0.02	0.03	11	0.06	0.01	15
Other instructional	−0.09	−0.08	3	−0.02	−0.01	7	0.11	−0.00	3	0.22	0.02	3
Cognitive behavioral, behavioral modeling, or behavior modification	0.12	0.06	4	0.18	0.22	3	0.37*	0.22*	8	0.32*	0.16*	25
Counseling, social work, and other therapeutic interventions	−0.41*	−0.17*	3	−0.19	−0.18	2	−0.20*	−0.18	6	0.04	−0.14	6
Mentoring, tutoring, and work study	0.06	−0.02	5	−0.12	−0.15	2	0.19	0.05	8	0.13	−0.03	6
Recreation, community service, enrichment, and leisure activities	—[c]	—[c]		−0.14	−0.11	2	0.11	0.13	2	0.03	−0.02	3

[a]The raw effect size.
[b]The covariance adjusted effect size (method and sample equated).
[c]No effect size.
*$p \leq 0.05$.

effects for instructional programs with cognitive-behavioral methods and noninstructional cognitive-behavioral or behavioral modeling programs were substantially attenuated in the equated analysis but remained statistically significant.

3.2.5. Publication Bias

It is widely recognized that omitting unpublished studies may lead to an upward bias in the findings from meta-analysis (Kraemer *et al.*, 1998; Lipsey and Wilson, 2001). We believe that the threat to our findings from publication bias is minimal for several reasons. First, this meta-analysis included some unpublished studies (slightly less than 10%) and studies published in book chapters or other forms (roughly 10%) that are less likely to be affected by the tendency to suppress studies without significant effects from peer-reviewed journals. Second, a scatterplot of the relationship between sample size and effect size shows the expected funnel shape that would occur if the effect size distribution is not censored (Light and Pillemer, 1984). Third, the difference in the mean effect size for published and unpublished studies was small, with the unpublished studies having the larger, not the smaller, mean effect. And, finally, the typical sample size of the studies included in this meta-analysis is large. Kraemer *et al.* (1998) argued and demonstrated statistically that restricting a meta-analysis to studies with large sample size protects against publication bias.

4. DISCUSSION

Many popular school-based prevention approaches have not been well studied. Gottfredson *et al.* (2000) showed that schools are implementing a broad range of prevention strategies. This range of activities now in practice contrasts sharply with the relatively narrow range of strategies for which we have a sufficient number of reasonably rigorous studies to justify solid conclusions regarding effectiveness. The number of relevant treatment–control contrasts for which effect sizes could be computed (see Table VIII) is small for all categories of programs except instructional programs and noninstructional programs using cognitive-behavioral or behavioral methods. The number of contrasts available for programs to establish norms and classroom organization and management activities exceeded 10 for selected outcomes. More studies of a broader range of school-based strategies are clearly needed.

The main finding of this research is that school-based prevention practices appear effective for reducing alcohol and drug use, dropout and nonattendance, and other conduct problems. The effect size for measures of

delinquency is positive, but its 95% confidence interval includes zero. The size of the average effect for each of the four outcomes is small. This average, however, masks a large amount of heterogeneity, due in part to differences across studies in the evaluation methods used, types of populations served, and type of prevention activities.

The evaluation methodology used was related to the magnitude of the effects observed. More rigorous studies produced higher effect sizes. This is good news, in a sense, because it means that the inclusion of studies that were less rigorous in this research synthesis did not inflate the estimates of overall effects. It also implies that it is necessary to control for characteristics of the methodology when examining the effects of variables of substantive interest, for methodology may be confounded with these relationships.

Characteristics of the target population were also related to observed effect sizes. Programs that targeted more at-risk populations had larger observed effects. Although conclusions about targeting must be tempered with a concern that the lower base rate of problem behavior for the general population places an artificial upper bound on the possible effect that could be observed, the result is encouraging given that a relatively small proportion of the population is responsible for a large proportion of crime. It appears that prevention strategies can be particularly effective with these higher risk populations.

Effect sizes varied considerably by type of program, even after statistically adjusting for measured characteristics of the population and methodological differences. That is, we observed a large range of effect sizes across program categories. For example, non-cognitive-behavioral counseling, social work, and other therapeutic interventions showed consistently negative effects across all four outcomes. Self-control or social competency promotion instruction using cognitive-behavioral and behavioral instructional methods, on the other hand, showed consistently positive results across all four outcomes, as did noninstructional programs using the same types of methods. Environmentally focused interventions were also particularly effective for reducing delinquency and drug use.

With the exception of non-cognitive-behavioral counseling interventions, for which evidence is consistently negative, we believe that it is premature to recommend against the use of any of the strategies included in the study because so few studies have been conducted in most areas. We can be reasonably confident in predicting that instructional prevention programs will be more effective when they are taught using methods based on sound learning principles. In the classification used in this study, cognitive-behavioral and behavioral modeling methods or training involved repeated exposure to new behaviors with rehearsal and feedback or extended use of cues to elicit behavior over long periods or in a variety of

settings. Purely behavioral programs were also included in this category. These interventions involved timely tracking of specific behaviors over time, behavioral goals, and use of feedback of positive or negative reinforcement to change behavior. Instructional programs that used these methods appeared to be more effective than instructional programs that were taught using more traditional methods, even if they covered social competency skill content. This may account for the limited effectiveness of the D.A.R.E. program (Gottfredson, 1997) relative to other programs (such as Botvin's Life Skills Training program) which have fared better in evaluations. This suggests that many of the existing curricula can be improved by incorporating new teaching techniques rather than replacing them.

We also recommend that schools make use of programs, such as Lochman's Anger Coping Program (Lochman, 1985, 1992; Lochman *et al.*, 1984, 1989; Lochman and Curry, 1986) and Bry's behaviorally based preventive intervention (Bry, 1982; Bry and George, 1979, 1980), that target high-risk youths with programs that incorporate cognitive retraining and behavioral methods, as described above. These programs appear to be among the most effective school-based programs.

Finally, it is apparent that any one school-based strategy, implemented in isolation, will not have a large effect, given that none of the evaluated program categories observed large effects. Schools seem to operate under this assumption already because they offer many different types of prevention programs simultaneously. It would appear that school-based prevention, in practice, is generally *not* a stand-alone curricular or other type of intervention. Rather, it is a *mix* of many different activities that schools implement. This suggests that at least as important as the question "Which program works?" are questions such as "Which combinations or sequences of strategies work best?" and "How can schools effectively design comprehensive packages of prevention strategies and implement them in a high-quality fashion?" Little is known about the potential additive and multiplicative effect of combinations of distinct programs. Researchers should study the relative effectiveness of sets of interventions and, eventually, develop a knowledge base to guide decisions about which combinations are most effective at which development stages.

ACKNOWLEDGMENTS

This work was supported in part by a grant from Jerry Lee. We gratefully acknowledge the research assistance of Todd Armstrong, Jennifer Castro, Elaine Eggleston, Paul Hirschfield, Veronica Puryear, John Ridgely, and Shannon Womer and the advice of Gary Gottfredson.

REFERENCES

Botvin, G. J. (1990). Substance abuse prevention: Theory, practice, and effectiveness. In Tonry, M., and Wilson, J. Q. (eds.), *Crime and Justice: A Review of the Research: Vol. 13. Drugs and Crime*, University of Chicago Press, Chicago, pp. 461–519.

Botvin, G. J., Schinke, S., and Orlandi, M. A. (1995). School-based health promotion: Substance abuse and sexual behavior. *Appl. Prev. Psychol.* 4: 167–184.

Bry, B. H. (1982). Reducing the incidence of adolescent problems through preventive intervention: One- and five-year follow-up. *Am. J. Commun. Psychol.* 10(3): 265–276.

Bry, B. H., and George, F. E. (1979). Evaluating and improving prevention programs: A strategy from drug abuse. *Eval. Prog. Planning* 2: 127–136.

Bry, B. H., and George, F. E. (1980). The preventive effects of early intervention on the attendance and grades of urban adolescents. *Professional Psychology* 11: 252–260.

Catalano, R. F., Arthur, M. W., Hawkins, J. D., Berglund, L., and Olson, J. J. (1998). Comprehensive community- and school-based interventions to prevent antisocial behavior. In Loeber, R., and Farrington, D. P. (eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*, Sage, Thousand Oaks, CA, pp. 248–283.

Dryfoos, J. G. (1990). *Adolescents at Risk: Prevalence and Prevention,* Oxford University Press, New York.

Durlak, J. A. (1995). *School-Based Prevention Programs for Children and Adolescents*, Sage, Thousand Oaks, CA.

Ennett, S. T., Tobler, N. S., Ringwalt, C. L., and Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project D.A.R.E. outcome evaluations, *Am. J. Public Health* 84: 1394–1401.

Gerstein, D. R., and Green, L. W. (eds.). (1993). *Preventing Drug Abuse: What Do We Know?* National Academy Press, Washington, DC.

Gleser, L. J., and Olkin, I. (1994). Stochastically dependent effect sizes. In Cooper, H., and Hedges, L. V. (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, pp. 339–355.

Gottfredson, D. C. (1997). School-based crime prevention. In Sherman, L. W., Gottfredson, D. C., MacKenzie, D., Eck, J., Reuter, P., and Bushway, S. (eds.), *Preventing Crime: What Works, What Doesn't, What's Promising: A Report to the United States Congress.* U.S. Department of Justice, Office of Justice Programs, Washington, DC, pp. 5.1–5.71.

Gottfredson, D. C. (2001). *Schools and Delinquency*, Cambridge University Press, New York.

Gottfredson, G. D., Gottfredson, D. C., Czeh, E. R., Cantor, D., Crosse, S., and Hantman, I. (2000). *A National Study of Delinquency Prevention in Schools*, Gottfredson Associates, Ellicott City, MD.

Gottfredson, D. C., Wilson, D. B., and Najaka. School-based crime prevention. In Sherman, L. W., Farrington, D. P., Welsh, B., and Mackenzie, D. C., New York (eds.), *Evidence-Based Crime Prevention*, Harwood Academic (in press).

Hansen, W. B. (1992). School-based substance abuse prevention: A review of the state of the art in curriculum: 1980–1990. *Health Educ. Res.* 7: 403–430.

Hansen, W. B., and O'Malley, P. M. (1996). Drug use. In DiClemente, R. J., Hansen, W. B., and Ponton, L. E. (eds.), *Handbook of Adolescent Health Risk Behavior*, Plenum Press, New York, pp. 161–192.

Hawkins, J. D., Arthur, M. W., and Catalano, R. F. (1995). Preventing substance abuse. In Tonry, M., and Farrington, D. P. (eds.), *Crime and Justice: A Review of the Research: Vol. 19. Building a Safer Society: Strategic Approaches to Crime Prevention*, University of Chicago Press, Chicago, pp. 343–427.

Hawkins, J. D., Farrington, D. P., and Catalano, R. F. (1998). Reducing violence through the schools. In Elliott, D. S., Hamburg, B. A., and Williams, K. R. (eds.), *Violence in American Schools*, Cambridge University Press, New York, pp. 188–216.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.* 6: 107–128.

Hedges, L. V., and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*, Academic Press, New York.

Huizinga, D., and Jakob-Chien, C. (1998). The contemporaneous co-occurrence of serious and violent juvenile offending and other problem behaviors. In Loeber, R., and Farrington, D. P. (eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*, Sage, Thousand Oaks, CA, pp. 47–67.

Institute of Medicine (1994). *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*, National Academy Press, Washington, DC.

Johnston, L. D., Bachman, J. G., and O'Malley, P. M. (1997). *Monitoring the Future: Questionnaire Responses from the Nation's High School Seniors, 1995*, Institute for Social Research, Ann Arbor, MI.

Johnston, L. D., O'Malley, P. M., and Bachman, J. G. (1999). *Drug Trends in 1999 Are Mixed*, University of Michigan News and Information Services, Ann Arbor, Dec. [on-line]. Available at www.monitoringthefuture.org; accessed 3/3/00.

Kalaian, H. A., and Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychol. Methods* 1: 227–235.

Kraemer, H. C., Gardner, C., Brooks, J. O., III, and Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychol. Methods* 3: 23–31.

Light, R. J., and Pillemer, D. B. (1984). *Summing Up: The Science of Reviewing Research*, Harvard University Press, Cambridge, MA.

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., and Mosteller, F. (eds.), *Meta-Analysis for Explanation*, Russell Sage Foundation, New York, pp. 83–127.

Lipsey, M. W., and Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In Loeber, R., and Farrington, D. P. (eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions,* Sage, Thousand Oaks, CA, pp. 86–105.

Lipsey, M. W., and Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *Am. Psychol.* 48: 1181–1209.

Lipsey, M. W., and Wilson, D. B. (2001). *Practical Meta-Analysis*, Sage, Thousand Oaks, CA.

Lochman, J. E. (1985). Effects of different treatment lengths in cognitive behavioral interventions with aggressive boys. *Child Psychiatry Hum. Dev.* 16: 45–56.

Lochman, J. E. (1992). Cognitive-behavioral intervention with aggressive boys: Three-year follow-up and preventive effects. *J. Consult. Clin. Psychol.* 60: 426–432.

Lochman, J. E., and Curry, J. F. (1986). Effects of social problem-solving training and self-instruction training with aggressive boys. *J. Clin. Child Psychol.* 15: 159–164.

Lochman, J. E., Burch, P. R., Curry, J. F., and Lampron, L. B. (1984). Treatment and generalization effects of cognitive-behavioral and goal-setting interventions with aggressive boys. *J. Consult. Clin. Psychol.* 52: 915–916.

Lochman, J. E., Lampron, L. B., Gemmer, T. C., Harris, S. R., and Wyckoff, G. M. (1989). Teacher consultation and cognitive-behavioral interventions with aggressive boys. *Psychol. Schools* 26: 179–188.

Norman, E., and Turner, S. (1993). Adolescent substance abuse prevention programs: Theories, models, and research in the encouraging 80's. *J. Primary Prev.* 14: 320.

Raudenbush, S. W. (1994). Random effects models. In Cooper, H., and Hedges, L. V. (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, pp. 301–321.

Rosenthal, R. (1994). Parametric measures of effect size. In Cooper, H., and Hedges, L. V. (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, pp. 231–244.

Samples, F. and Aber, L. (1998). Evaluations of school-based violence prevention programs. In Elliott, D. S., Hamburg, B. A., and Williams, K. R. (eds.), *Violence in American Schools*, Cambridge University Press, New York, pp. 217–252.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D., Eck, J., Reuter, P., and Bushway, S. (1997). *Preventing Crime: What Works*, *What Doesn't*, *What's Promising: A Report to the United States Congress*, U.S. Department of Justice, Office of Justice Programs, Washington, DC.

Stage, S. A., and Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychol. Rev.* 26: 333–368.

Tobler, N. S. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *J. Drug Issues* 16: 537–567.

Tobler, N. S. (1992). Drug prevention programs can work: Research findings. *J. Addict. Dis.* 11: 128.

Tobler, N. S., and Stratton, H. H. (1997). Effectiveness of school-based drug prevention programs: A meta-analysis of the research. *J. Primary Prev.* 18: 71–128.

Tremblay, R. E., and Craig, W. M. (1995). Developmental crime prevention. In Tonry, M., and Farrington, D. P. (eds.), *Crime and Justice: A Review of the Research: Vol. 19. Building a Safer Society; Strategic Approaches to Crime Prevention*, University of Chicago Press, Chicago, pp. 151–236.

Wilson, D. B. (1995). *The Role of Method in Treatment Effect Estimates: Evidence from Psychological, Behavioral, and Educational Treatment Intervention Meta-Analyses*, Doctoral dissertation, Claremont Graduate School, Claremont, CA.