Identifying Effective School-Based Substance Abuse Prevention Interventions

Michael Roona, Andrei Streke, Peter Ochshorn, Diana Marshall and Amy Palmer

Abstract

Previous meta-analytic studies of universal school-based drug education program evaluations have found that interactive programs are more effective than non-interactive programs and that within the group of more effective interactive programs, Comprehensive Life Skills programs are more effective overall than Social Influences programs. This study builds upon those earlier meta-analytic studies of universal school-based drug education program evaluations by examining the relative effectiveness of Comprehensive Life Skills programs and Social Influences programs on the use and/or abuse of different substances at different grade levels. We find the previous findings about the superiority of Comprehensive Life Skills programs appear to be true at the elementary and high school levels, but not at the middle school level. At the middle school level, we find no difference between the effectiveness of Comprehensive Life Skills programs and Social Influences programs for cigarettes, marijuana, all drugs combined, and all drugs excluding alcohol. For alcohol measures, we find the Social Influences programs are significantly more effective than the Comprehensive Life Skills programs. Furthermore, we find the "success" of the Social Influences programs is largely attributable to their ability to reduce alcohol abuse. The Social Influences programs have no effect on the prevalence of alcohol use. The superiority of the Social Influences programs relative to the Comprehensive Life Skills programs at reducing the prevalence of alcohol use at the middle school level is a consequence of the fact that Comprehensive Life Skills programs are counterproductive, i.e., they encourage alcohol use without reducing alcohol abuse. We conclude that promoting abstinence may not be a viable objective when substance use is normative in the culture, but preventing abuse and its attendant harms may be viable.

Introduction

Before we can implement effective interventions, we must know which types of interventions are effective. We also need to know if some interventions are substantially more effective than others and whether the benefits derived from the more effective interventions are worth the cost. Finally, we should know whether some interventions are more likely to be implemented fully and correctly

by those responsible for implementation and we should know about any adverse consequences that may derive from implementing different types of interventions. In addition, before we can implement effective substance abuse prevention interventions, we need to know what we are trying to prevent. Are we trying to prevent the use of cigarettes? Or the abuse of alcohol? When we talk about alcohol abuse, is binge drinking the primary concern or is driving while intoxicated the primary concern? An intervention may succeed in reducing underage drinking (or delaying the onset of alcohol use by underage youths) by reinforcing negative attitudes toward alcohol in some youths, yet among those youths who are predisposed to consume alcohol, that same intervention may encourage reckless behavior by inciting youths to play drinking games or drive while intoxicated. Conversely, an intervention may be highly effective at reducing binge drinking by encouraging moderation, but have no effect on the prevalence or frequency of alcohol consumption. How we define the measures of success for an intervention has a potentially profound effect on the "success" of the intervention. Promoting abstinence does not necessarily reduce harm, and reducing harm is not likely to encourage abstinence. Furthermore, because "the drug problem" in America is socially constructed in terms of the abuse of crack, crystal meth, heroin, or other "hard" drugs and the attendant criminal justice and public health problems associated with such abuse, reducing the prevalence of cigarette, alcohol, or marijuana abuse may have some effect on "the drug problem" but reducing the prevalence of use of cigarettes, alcohol, or even marijuana is unlikely to have any effect.

Studies examining the effectiveness of school-based interventions intended to prevent substance use or abuse have led to a divergence of opinion about how well different types of interventions work and for whom. Effective universal school-based drug education programs that target all students at a particular grade level or series of grade levels are seen by some as an ideal prevention intervention—despite the relatively modest effects that result from trying to prevent substance use or abuse in a heterogeneous population that includes many who are not predisposed to use or abuse substances—because such programs are relatively easy to implement and relatively inexpensive when opportunity costs are not taken into account. Others question how beneficial even the most effective universal programs can be if they fail to provide the counseling and other assistance needed by selective and indicated populations who are "at risk" of becoming substance users or abusers. Still others argue that specialist-intensive school-based add-on programs are too disruptive of school operations and stretch limited resources too far so we would be better off focusing on changing the school environment to protect against risk factors that are common to substance abuse, violence, teen pregnancy, and academic failure.

Even among those who advocate universal school-based drug education curricula, differences exist. Many argue that to be effective, drug education

programs must teach refusal skills. Others argue that time and labor intensive role-playing exercises that teach refusal skills are not essential because the key to effectively preventing substance abuse is correcting youth's misperceptions about how often their peers use various substances and the amounts they consume. While today many prevention experts claim teaching refusals skills and correcting normative misperceptions about substance use by peers are both important components of effective substance abuse prevention programs, critics argue that because evaluations demonstrating the effectiveness of these interventions measured substance use rather than abuse, there is no evidence that universal drug education programs promoting abstinence effectively reduce substance abuse and the myriad related problems often referred to generically as "the drug problem" in America.

Beyond the challenges posed by those who consider reducing the prevalence of substance abuse to be more important than reducing the prevalence of substance use and by the broader questions about whether transforming the school environment offers greater benefits or is more cost effective than providing individualized student assistance services or classroom-based drug education curricula are a range of challenges posed by differences in the quality of prevention program implementations and evaluations. Differences in experimental designs, thoroughness of implementation, measures of success, sample sizes, pedagogy, and program content all impact conclusions drawn about program effectiveness. To draw valid conclusions about program effectiveness, these differences must be taken into account to systematically synthesize findings from available studies. Because the traditional literature review tends to be subjective and use limited samples, meta-analytic statistical techniques have been employed to synthesize the ensemble of available studies to quantitatively assess the effectiveness of different types of universal school-based drug education programs.

Methods

The first step in the meta-analysis reported here was to locate a complete set of published and unpublished quantitatively evaluated studies and determine whether they met appropriate selection criteria. For current purposes, the selection criteria we employed specified that the studies all have (1) been reported or published between 1978 and 1998; (2) evaluated a universal school-based drug education curriculum implemented at any grade level from kindergarten through twelfth grade that did not specifically target ethnic groups, "high risk" or substance abusing youth, or other subpopulations; (3) sought to prevent (rather than treat) substance use or abuse; (4) employed experimental research designs or employed quasi-experimental research designs and reported both pretest and posttest measures; and (5) reported quantitative drug use measures. After the complete set of studies was located and screened for

inclusion, eligible studies were systematically coded and quantitative estimates of program impact were computed. Finally, correlates of program effectiveness were analyzed. This rigorous approach toward analyzing findings synthesized from the universe of program evaluations facilitates contextualizing the inconsistencies found in individual research studies to identify collective findings.

Previous meta-analytic studies of universal school-based drug education program evaluations have found that interactive programs are more effective than non-interactive programs and that within the group of more effective interactive programs, Comprehensive Life Skills programs are more effective overall than Social Influences programs (Tobler, Roona, Ochshorn, Marshall, Streke, & Stackpole, 2000; Tobler & Stratton, 1997). This report builds upon those earlier meta-analytic studies of universal school-based drug education program evaluations by focusing on evaluation findings for the 128 classroom-focused interactive interventions that were previously identified as being more effective than the non-interactive interventions. In addition, 36 variables have been recoded or created and coded and 109 effect sizes recomputed or newly computed for 36 evaluations that provided detailed information about the level of alcohol use measured.

Results

For the investigation reported here, we examined differential effects for Comprehensive Life Skills and Social Influences programs at the elementary, junior high,[1] and high school levels for nine overlapping first year samples,[2] differentiated by substance measure: (1) All Drug Use Measures (N=113); (2) Cigarette Use Measures (N=84); (3) Marijuana Use Measures (N=25); (4) All Non-Alcohol Use Measures (N=97); (5) All Alcohol Use Measures (N=49); (6) Subset of Alcohol Use Measures Corresponding to Not Heavy Alcohol Use Measures (N=36); (7) Not Heavy Alcohol Use Measures (N=36); (8) Subset of Not Heavy Alcohol Use Measures Corresponding to Heavy Alcohol Use (N=15); and (9) Heavy Alcohol Use Measures (N=15).

Tables 1 and 2 show intervention effects found for each non-alcohol and alcohol use outcome, respectively, by grade level for both of the classroom-focused interactive treatment modalities, namely, Social Influences (SI) and Comprehensive Life Skills (CLS). The effect sizes associated with interventions in

[1] We use junior high and middle school interchangeably in this report because we combined results of junior high school and middle school program evaluations when we conducted our analyses. This means in some cases data about sixth grade interventions provide information about elementary school interventions and in other cases they provide information about middle school interventions, depending upon the grade structure of the schools.

[2] Fifteen of the 128 interventions did not measure program effects within one year of pretest and have been excluded.

each construct category were weighted using the standard inverse variance formula and averaged to produce the aggregated effect sizes in each construct category. The tables show the total number of interventions and the cumulative weight associated with those interventions upon which each set of aggregated effect sizes is based. (Cumulative weight is a proxy for sample size.) The tables also report the weighted mean effect sizes in each construct category and the results of the following statistical tests on those effect sizes:

1. A t-test was employed to determine whether the weighted mean effect size is significantly different from zero. (The asterisk denotes significant results with the associated *p*-value less than .05.)

2. A z-test was used to determine if the weighted mean effect size for the Social Influences group differs significantly from the effect size for the Comprehensive Life Skills group. (This is shown in the column *effect difference*; the associated *p*-value is placed in the adjacent column.)

3. Because our previous research demonstrated that the effectiveness of interactive programs decreases as sample size increases and the Social Influences programs tend to be larger than the Comprehensive Life Skills programs, we controlled for the program size when testing for differences in program effectiveness (shown in the *adjusted difference* column with its associated *p*-value).

For discussion purposes, we will assume an effect size around 0 indicates ineffectiveness, an effect size around 0.1 indicates slight effectiveness, and an effect size around 0.2 indicates moderate effectiveness. For social programs, an effect size of 0.2 is generally considered small but clinically meaningful. Determining a cut point for clinical significance is somewhat arbitrary, but an effect size much less than 0.1 is clearly not clinically meaningful. Given the distribution of effect sizes in our analysis dataset, defining an effect size (or difference between effect sizes) around 0.1 as the smallest clinically meaningful estimate of program effect is pragmatic. Furthermore, it is consistent with the approach used by Sherman, Gottfredson, MacKenzie, Eck, Reuter, and Bushway (1998) in the report they prepared for the National Institute of Justice to submit to Congress about the effectiveness of crime prevention programs. An effect size of –0.1 is the smallest clinically meaningful indicator that a program is counterproductive. We will use –0.1, 0.0, 0.1 , and 0.2 as reference points in the discussion to follow.

The majority of the non-alcohol mean effect sizes are significantly larger than zero despite the small number of interventions in most of the construct categories. The effect sizes for the elementary school Social Influences interventions, however, are not significantly different from zero. The junior high school level is a

major focus of our analysis since it is the focus of the Safe and Drug Free Schools Program's Middle School Prevention Coordinator Initiative. The junior high school level provides the largest sample of studies (and subjects) and, therefore, a more grounded basis for statistically sound inferences. For junior high school, non-alcohol effects range from slightly effective (for all subsamples) to moderately effective (for Comprehensive Life Skills marijuana use: 0.217).

As shown in Table 1, significant effects favoring Comprehensive Life Skills programs were found at the elementary and senior high school levels for *All Drugs, Cigarettes, Marijuana, and All Drugs Excluding Alcohol*. (The only exception is the *Cigarettes* sample at the senior high school level, which had an effect difference greater than 0.1, but a *p*-value of .08.) This finding demonstrating that students in the Comprehensive Life Skills group showed significantly less increase in substance use relative to comparison students than students in the Social Influences group is consistent with the findings of Tobler, Roona, et al. Controlling for program size differences between Social Influences and Comprehensive Life Skills interventions had no effect on the statistical significance of the differences between treatment modalities at the elementary and high school levels, i.e., the magnitude of the statistically significant differences between the Comprehensive Life Skills and Social Influences programs was reduced when controlling for program size, but the statistical significance of the differences remained unchanged for $p < .05$.

Table 1. First Year Non-Alcohol Effect Sizes by Intervention Modality and Grade Level

All Drugs

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p*-value	Difference	*p*-value
Elementary	9 / 4302	-0.026		6 / 250	0.230	*	0.256	0.00	0.209	0.01
Junior High	45 / 9548	0.103	*	26 / 3399	0.129	*	0.026	0.19	0.001	0.95
Senior High	18 / 2711	0.078	*	9 / 590	0.469	*	0.391	0.00	0.298	0.00

Cigarettes

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p*-value	Difference	*p*-value
Elementary	6 / 3701	-0.032		6 / 250	0.230	*	0.262	0.00	0.219	0.01
Junior High	34 / 8864	0.095	*	22 / 3273	0.141	*	0.047	0.02	0.028	0.23
Senior High	11 / 1287	0.080	*	5 / 202	0.211	*	0.131	0.08	-0.081	0.38

Marijuana

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p*-value	Difference	*p*-value
Elementary	1 / 90									
Junior High	12 / 4661	0.146	*	8 / 443	0.217	*	0.072	0.15	-0.010	0.88
Senior High	2 / 649	-0.024		2 / 131	1.102	*	1.126	0.00	N/A	

All Drugs, Excluding Alcohol

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p*-value	Difference	*p*-value
Elementary	8 / 3944	-0.026		6 / 250	0.230	*	0.256	0.00	0.193	0.02
Junior High	37 / 8935	0.112	*	23 / 3316	0.135	*	0.023	0.24	-0.002	0.94
Senior High	16 / 2385	0.115	*	7 / 343	0.537	*	0.423	0.00	0.309	0.00

It is noteworthy, however, that at the junior high school level, the difference tests between Comprehensive Life Skills and Social Influences programs for most non-alcohol comparisons did not reach the minimally accepted level of significance despite having adequate statistical power to detect significant differences for $p < .05$. The clinically irrelevant effect size difference of 0.047 for the junior high school *Cigarettes* sample constitutes the only exception, but the difference ceased to be statistically significant when we controlled for program size

Table 2. First Year Alcohol Effect Sizes by Intervention Modality and Grade Level

All Alcohol Use Measures (excluding strategy #65)

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p* -value	Difference	*p* -value
Junior High	21 / 5622	0.093	*	11 / 530	0.000		-0.093	0.04	-0.237	0.00
Senior High	8 / 1895	0.002		4 / 415	0.491	*	0.489	0.00	N/A	

Subset of All Alcohol Use Corresponding to Not Heavy Alcohol Use

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p* -value	Difference	*p* -value
Junior High	17 / 2753	0.085	*	8 / 465	-0.045		-0.130	0.01	-0.224	0.00
Senior High	4 / 1184	0.002		2 / 84	0.274		0.272	0.08	N/A	

Not Heavy Alcohol Use

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p* -value	Difference	*p* -value
Junior High	17 / 2753	0.061	*	8 / 465	-0.100	*	-0.161	0.00	-0.238	0.00
Senior High	4 / 1184	0.003		2 / 84	0.233		0.230	0.11	N/A	

Subset of Not Heavy Alcohol Use Corresponding to Heavy Alcohol Use

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p* -value	Difference	*p* -value
Junior High	5 / 821	0.052		7 / 309	-0.140	*	-0.192	0.00	-0.495	0.00
Senior High	1 / 29			1 / 70						

Heavy Alcohol Use (excluding strategy #65)

	SI			CLS			Effect		Adjusted	
	N	Effect		N	Effect		Difference	*p* -value	Difference	*p* -value
Junior High	5 / 821	0.252	*	7 / 309	0.104		-0.148	0.03	-0.452	0.00
Senior High	1 / 29			1 / 70				0.08		

N/A: *Covariate difference is not applicable when sample/program size has no significant influence*

Table 2 shows separate analyses repeated for intervention evaluations that measured alcohol using different sets of studies (1) to assess the relative effectiveness of Social Influences and Comprehensive Life Skills interventions, (2) to examine the sensitivity of the findings about relative effectiveness to different subsets of studies, and (3) to contrast the relative effectiveness of Social Influences and Comprehensive Life Skills interventions on heavy alcohol use with

[3] Strategy #65 was excluded from our alcohol analyses because the evaluation only measured heavy alcohol use. By excluding this one anomalous case, we maintain comparability across heavy and non-heavy use samples.

effects on other measures of alcohol use.[3]

Although samples are smaller for the alcohol-based construct categories, we still have a sufficient number of interventions and subjects at the policy-relevant and clinically important junior high school level. (No Comprehensive Life Skills program evaluations at the elementary level measured alcohol use and only five Social Influences program evaluations at the elementary level measured alcohol use, so elementary level alcohol data have been ignored.) With regard to alcohol use, the relative effectiveness of the Social Influences and Comprehensive Life Skills interventions at the junior high school level was the reverse of what we found at other grade levels and for other substances. That is, z-tests revealed that effect sizes for the Social Influences groups are significantly higher than for the Comprehensive Life Skills groups for all alcohol-related junior high school samples. In half the cases, this is because the Comprehensive Life Skills programs are counterproductive and the Social Influences programs have no clinically meaningful effect.

The set of *All Alcohol Use Measures* at the top of Table 2 includes prevalence use measures, index measures of alcohol use, heavy alcohol use measures and other measures of alcohol use. The *Subset of All Alcohol Use Measures* immediately below shows little variation in effect sizes or significance tests when the data are subset to correspond to those programs that measure prevalence rates (and possibly other measures) except for the Comprehensive Life Skills effect size at the high school level, which appears to be skewed by an outlier.

The mean weighted effect sizes based on *Not Heavy Alcohol Use* (e.g., prevalence rates) are noticeably lower than the overall effect sizes for both intervention modalities, indicating that the presence of heavy use measures in the *Subset of All Alcohol Use Measures* appears to be largely responsible for the limited beneficial effects prevention programs have on alcohol use at the junior high school level.

The magnitudes of the effect sizes decrease somewhat when the evaluations that measured *Not Heavy Alcohol Use* are subset to include only those evaluations that also measured *Heavy Alcohol Use*, reinforcing the ineffectiveness of Social Influences programs and the clinically and statistically counterproductive nature of the Comprehensive Life Skills programs regarding "casual" alcohol use.

Tobler, Roona, *et al.* found that interactive programs are effective in reducing tobacco, alcohol, and other drug use. Contrasting effects on the use of different substances and different levels of alcohol use allows for more nuanced differentiation. Table 1 indicates that for non-alcohol use measures, the Comprehensive Life Skills programs are better than the Social Influences programs in elementary schools and high schools, but there is no difference

between the Social Influences and Comprehensive Life Skills programs at the middle school level, where most prevention programming takes place. For the 30-day prevalence rates and other general alcohol use measures in Table 2, however, the Social Influences programs are superior to the Comprehensive Life Skills programs at the junior high school level, although it is questionable whether the small statistically significant effects associated with the Social Influences programs are clinically significant. The effects of Social Influences programs on heavy alcohol use, though, clearly are clinically significant. The Social Influences programs appear to be moderately effective at reducing heavy drinking even though they are ineffective at reducing the overall prevalence of alcohol use.

Given the widespread belief in the prevention field that Comprehensive Life Skills programs are superior to the older Social Influences programs but empirical evidence to the contrary in our dataset for the junior high school students, we undertook follow-up analyses of the first-year junior high school samples to identify confounding factors that might account for this potential anomaly by examining differences among programs classified according to whether they target alcohol, cigarettes, or substance abuse in general. Five overlapping target-differentiated first-year samples were analyzed:

(1) All Alcohol Use Measures (N=32);

(2) Subset of All Alcohol Use Measures Corresponding to Not Heavy Alcohol Use Measures (N=25);

(3) Not Heavy Alcohol Use Measures (N=25);

(4) Subset of Not Heavy Alcohol Use Corresponding to Heavy Use Measures (N=12); and

(5) Heavy Alcohol Use Measures (N=12). The results are presented in Table 3.

Table 3. First Year Effect Sizes for Junior High School by Target Drug								
All Alcohol Use Measures (excluding strategy #65)								
	SI			CLS			Effect	
	N	Effect		N	Effect		Difference	*p* -value
Overall	21 / 5622	0.093	*	11 / 530	0.000		-0.093	0.04
Targets Alcohol	10 / 644	0.077	*	3 / 82	0.167		0.090	0.45
Generic Program	10 / 3414	0.111	*	8 / 448	-0.030		-0.141	0.01
Subset of All Alcohol Use Corresponding to Not Heavy Alcohol Use								
	SI			CLS			Effect	
	N	Effect		N	Effect		Difference	*p* -value
Overall	17 / 2753	0.085	*	8 / 465	-0.045		-0.130	0.01
Targets Alcohol	9 / 612	0.057		2 / 65	0.166		0.109	0.40
Generic Program	8 / 2142	0.094	*	6 / 400	-0.080		-0.173	0.00
Not Heavy Alcohol Use								
	SI			CLS			Effect	
	N	Effect		N	Effect		Difference	*p* -value
Overall	17 / 2753	0.061	*	8 / 465	-0.100	*	-0.161	0.00
Targets Alcohol	9 / 612	0.057		2 / 65	0.015		-0.042	0.75
Generic Program	8 / 2142	0.062	*	6 / 400	-0.119	*	-0.181	0.00
Subset of Not Heavy Alcohol Use Corresponding to Heavy Alcohol Use								
	SI			CLS			Effect	
	N	Effect		N	Effect		Difference	*p* -value
Overall	5 / 821	0.052		7 / 309	-0.140	*	-0.192	0.00
Targets Alcohol	1 / 25	0.460	*	2 / 64	0.015		-0.445	0.06
Generic Program	4 / 797	0.040		5 / 245	-0.182	*	-0.222	0.00
Alc-Generic		0.420	*		0.197			
Heavy Alcohol Use (excluding strategy #65)								
	SI			CLS			Effect	
	N	Effect		N	Effect		Difference	*p* -value
Overall	5 / 821	0.252	*	7 / 309	0.104		-0.148	0.03
Targets Alcohol	1 / 25	0.380		2 / 64	0.460	*	0.080	0.73
Generic Program	4 / 797	0.248	*	5 / 245	0.011		-0.237	0.00
Alc-Generic		0.132			0.449	*		

For non-alcohol use measures, the Comprehensive Life Skills programs have larger effects than the Social Influences programs, but that the difference is neither clinically nor statistically significant, irrespective of whether the program targets cigarette use or is a generic substance abuse prevention program. For four of the five sets of alcohol use measures in Table 3, when alcohol is targeted, the difference between the Social Influences and Comprehensive Life Skills programs is not statistically significant, and in the one case where it is statistically significant, the results are skewed by the fact that only one Social Influences program targeted alcohol use and the effect size associated with that program is an outlier.

Among the generic substance abuse prevention programs, the Social Influences programs are clinically and statistically superior to the Comprehensive Life Skills programs for all five subsets of alcohol use measures. The possibility that our finding of Social Influences superiority is a consequence of the fact that the Social Influences programs targeted alcohol use while the Comprehensive Life Skills programs did not is thus ruled out.

Discussion

Our goal when we undertook this study was to identify those life skills that were primarily responsible for the superiority of Comprehensive Life Skills programs over Social Influences programs. We intended to develop a more refined classification scheme that would provide useful information to curriculum developers about which curriculum components are essential and which are extraneous. To that end, we initially collapsed our classification scheme and pooled the Comprehensive Life Skills programs and Social Influences programs so we could undertake a study of curriculum components common to interactive drug education programs for middle school students. We found that in most cases, when we compared the effects of programs that taught a life skill (e.g., communication, assertiveness, decision-making, coping, goal-setting) to the effects of programs that did not, programs that did not teach that life skill had larger weighted mean effect sizes than those that did. So rather than redefine our classification scheme, we hypothesized that Comprehensive Life Skills programs were not more effective at the middle school level than Social Influences programs. When we found support for that hypothesis for all drugs combined, we sought to determine whether the hypothesis held for each type of substance and level of use. We discovered that the hypothesis of no difference was generally

true, but in the case of alcohol, there was a difference and it favored Social Influences programs because Comprehensive Life Skills programs encouraged alcohol use. Furthermore, we found that Social Influences programs effectively reduced heavy alcohol use (which is the only proxy for abuse).

We began our study of curriculum components by examining the importance of refusal skills. There were only eight interventions in the combined set of Social Influences programs and Comprehensive Life Skills programs that did not teach refusal skills, but of all the possible skills these interactive middle school programs could teach, we expected refusal skills to account for the greatest amount of variance in estimates of program effect. To our surprise, interactive middle school drug education programs that did not teach refusal skills were just as effective as interactive middle school programs that taught refusal skills. The weighted mean effect size for both groups was 0.12. We were so surprised by this finding that we reread the evaluations of the eight interventions that did not teach refusal skills in search of an explanation. We concluded that either Hansen and Graham (1991) are correct when they argue that establishing conservative norms is more important than teaching peer pressure refusal skills or the evaluations of the small number of interventions that did not teach refusal skills are anomalies.

In his meta-analysis of juvenile delinquency programs, Lipsey (1992) found that 25% of the variance in effect sizes was attributable to different research methods employed by program evaluators. Given a sufficiently large number of cases, these methodological differences (and other factors that may impact the magnitude of effect sizes) can be ignored if the differences are equally distributed among the different types of interventions being compared. But when the number of interventions in one or more comparison groups is small, we ignore these methodological differences at our peril. Hansen and Graham, for example, used schools as the unit of assignment to treatment conditions but classrooms as the unit of analysis. Assigning subjects at one level and analyzing data at lower level is quite common in evaluations of school-based interventions. In fact, random assignment of students rarely occurs, because students generally are assigned to classes based on the scheduling algorithms used by school administrators, so different treatment conditions can only be applied to classes, not students. Yet most evaluations of school-based interventions use students as the unit of analysis, not classes. Hierarchical statistical techniques have been developed that take intra-class correlations into effect, but many evaluations do not employ these techniques because until recently they were not built into most statistical software packages used by evaluators. Methodologists believe that not statistically controlling for intra-class correlations results in a greater likelihood that statistically significant effects will be found (although one of the most famous evaluations of a drug education program found similar results when individuals and schools were used as the unit of analysis [Botvin, Baker, Dusenbury, Tortu, &

Botvin, 1990, p. 442, footnote 3]).

Given a sufficiently large number of evaluations and randomly distributed program characteristics that are unrelated to treatment (in this case, curriculum) characteristics, we can speak with confidence about the relative superiority of one treatment over another. If the number of evaluations is small or program characteristics unrelated to treatment characteristics are not randomly distributed, the provisional and contingent nature of the scientific findings must be stressed. Our findings that, at the middle school level, Comprehensive Life Skills programs are less effective than Social Influences programs and refusal skills are not an essential component of interactive drug education curricula fly in the face of the prevailing wisdom. They may be affirmed, refuted, or modified as a result of subsequent research. It may be, for example, that our findings are valid, but only for the first year. If we looked at program effects two or more years out, different findings might emerge. Different findings might also emerge if we used different formulas for computing effect sizes or different analytical strategies.

While meta-analysts everywhere use the probit transformation to compute effect sizes based on categorical data (and we have followed that convention in this study) we believe using probits generally overestimates program effects, especially when the effects are small, because most drug use is not normally distributed. We have developed an alternative transformation using the natural logarithm that assumes drug use is exponentially distributed because we believe an exponential distribution better reflects that pattern of substance use among school-aged children for most substances. Most youths do not use, most of those who use do not use heavily, and only a few are very heavy users. The normal distribution simply does not reflect this pattern of use very well. We hope in the future to examine the influence that this method of estimating program impact has on our assessment of the relative effectiveness of different types of drug education programs. We also hope to reanalyze our data using maximum likelihood estimation techniques, rather than simply assigning an effect size of zero when evaluators report non-significant findings without reporting numerical statistics. We suspect most non-significant effects are positive effects. Using maximum likelihood models will enable us to more properly account for those effects whereas assigning an effect size of zero to non-significant findings masks beneficial program effects. Finally, researchers have consistently found heterogeneity of effects due to reasons other than sampling fluctuations, e.g., due to programmatic and implementation variability. We can directly model this additional variance component and estimate so-called mixed effects models using likelihood-based methods.

The dilemma for policy makers is to make informed decisions about what types of programs to implement given the contingent nature of evidence regarding program effectiveness. Because the Social Influences programs have consistently

demonstrated their superiority as generic substance abuse prevention programs at the middle school level for all substances we examined, we feel fairly confident in recommending them. We are less certain of the finding that refusal skills are not an essential curriculum component, simply because so few interactive drug education programs did not teach refusal skills. But if the time and effort required to do role plays that teach refusal skills is great, interactive drug education programs that do not teach them may be more cost effective and easier to implement. Further research is needed to assess the importance of teaching refusal skills and the costs of teaching them relative to their benefits.

If prevention interventions in American schools are to take the form of classroom-based universal drug education curricula that focus primarily on middle schoolers (as the Safe and Drug Free Schools Program's Middle School Prevention Coordinator Initiative seems to indicate) and emphasize preventing tobacco, alcohol, and marijuana use (rather than "hard" drug use), perhaps the curricula should emphasize social influences rather than comprehensive life skills. Determining whether school-based prevention interventions should take the form of universal drug education curricula implemented in classrooms, however, is beyond the scope of this paper. It may well be the case that student assistance programs targeting the needs of "at risk" youth or school restructuring activities that create protective schools or some combination of the above is a better approach.

References

Botvin, G., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. (1990). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology,* 58(4), 437-446.

Hansen, W. & Graham, J. (1991). Preventing alcohol, marijuana, and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. *Preventive Medicine,* 20, 414-430.

Lipsey, M. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. Cook, H. Cooper, D. Cordray, H. Hartmann, L. Hedges, R. Light, T. Louis, F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook.* (pp. 83-127). New York: Russell Sage Foundation.

Sherman, L.W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1998). *Preventing crime: What works, what doesn't, and what's promising* [HTTP]. Available: www.ncjrs.org/works/wholedoc.htm <http://www.ncjrs.org/works/wholedoc.htm> [7/13/00].

Tobler, N., Roona, M., Ochshorn, P., Marshall, D., Streke, A., & Stackpole, K. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *The Journal of Primary Prevention,* 20(4), 275-336.

Tobler, N. & Stratton, H. (1997). Effectiveness of school-based drug prevention programs: A meta-analysis of the research. *The Journal of Primary Prevention,* 18(1), 71-128.